Universidad de Chile
Facultad de Ciencias Sociales
Carrera de Psicología

Percepción de rostros: Top Down vs. Bottom Up

Captura de la atención comparada de rostros vs. cuerpos humanos completos mediante registro de movimientos oculares durante la exploración libre de imágenes naturalistas.

Memoria para optar al Título Profesional de Psicólogo

Autor:
Diego Alonso Becerra Q.
Profesor Patrocinante:
PhD. Pedro Maldonado A.
Profesora Guía:
Margarita Bórquez Q.

2016

La última acción posible es la que define la propia percepción.

— 89 Puñaladas

Resumen

Los rostros son estímulos que capturan la mirada en humanos de manera involuntaria, sostenida, casi universal y desde temprano en la ontogenia, sin embargo, no presentan características físicas (o bottom-up) que expliquen su saliencia, ni calzan por completo dentro del conjunto de objetos cuya captura atencional es realizada de manera voluntaria, o altamente influida por factores cognitivos (usualmente referidos como top-down). Se diseñó un experimento para investigar si existe un efecto del cuerpo (como factor contextual) en el procesamiento extra-foveal temprano de escenas naturalistas, que influya en la velocidad del primer movimiento sacádico dirigido al rostro. Para esto, se registró la conducta ocular de 29 sujetos sanos con un oculómetro en una tarea de exploración libre, compuesto por versiones con y sin el cuerpo de un set de imágenes naturalistas. No se encontraron diferencias, atribuibles a la presencia del cuerpo, en la velocidad del primer movimiento sacádico dirigido al rostro.

Palabras clave: Rostros, captura atencional, saliencia, historia de selecciones, exploración libre.

Abstract

Faces are stimuli that captivates gaze in a sustained, involuntary, almost universal and ontogenetically early manner, nevertheless, they does not show physical (or bottom up) characteristics which could explain their high salience, nor typical characteristics of objects whose attentional capture is voluntary or highly influenced by cognitive factors such as expectations, learning history and subject's previous knowledge (usually referred as top-down). We conducted an experiment to test if there is in the extra-foveal early processing of naturalistic scenes, any effect attributable to the body (as a contextual factor), capable of affect the speed of the first saccadic movement directed towards the face. In order to do this, the ocular behavior of 29 healthy subjects in a free-viewing task was registered with an oculometer, the task compared a set of naturalistic images containing faces with and without the body. No differences had been found comparing the first two conditions.

Keywords: Faces, attentional capture, salience, selection history, free-viewing.

Índice de contenido

Índice de figuras

Introducción

We do sometimes have to live in the house while it is being rebuilt.

—William Wimsatt

Explorando cotidianamente el mundo con los ojos, ciertos estímulos llaman inevitablemente más nuestra atención que otros: luces intermitentes, colores intensos, palabras (ya sea en graffitis, marcas comerciales o anuncios) y entre ellos, los rostros (Cerf, Frady y Koch, 2009; Itti y Koch, 2000; Mack, Pappas, Silverman y Gay, 2002). Incluso si intentásemos evitar conscientemente el fijarnos en cualquiera de estos estímulos, nuestros primeros movimientos sacádicos (latencia alrededor de 100-500ms) irían dirigidos hacia ellos de todos modos (Cerf et al., 2009; Crouzet, Kirchner y Thorpe, 2010; Pinto, van der Leij, Sligte, Lamme y Scholte, 2013; Theeuwes y van der Stigchel, 2006), algunos investigadores se refieren a este fenómeno como saliencia, es decir, la propiedad que presenta un estímulo de captar atención por sobre el resto de la escena (Esber y Haselgrove, 2011), la cual se postula como expresión de mecanismos neurales centro-periferia (Itti y Koch, 2000), mientras que otros limitan el uso de dicho término al grado en el que un estímulo atrae la atención (i) desde sus características de bajo nivel (concepto que excluye las palabras y los rostros) y (ii) con independencia del estado mental del sujeto en el momento de la fijación de la mirada (Awh, Belopolsky y Theeuwes, 2012) en adelante, se utilizará saliencia con esta última acepción.

La captura involuntaria de la mirada se debe a factores filogenéticos de la especie (Chica, Bartolomeo y Lupiañez, 2013), así como también, a factores ontogenéticos de socialización (Kubota e Ito, 2007; Miyamoto, Nisbett y Masuda, 2006), dicho de otro modo, para comprender la percepción visual humana, esta requiere enmarcarse en la *EvoDevo* (*Evolution and Developmental Biology*) además de la neurociencia cognitiva y la psicología de la percepción. Se requiere un estudio transdisciplinario y que se proponga abordar el fenómeno en toda su complejidad. Pero históricamente, los programas de investigación que han abordado el estudio de la percepción visual se han enfocado en testeos experimentales controlados con muy poca validez ecológica (Birmingham, Bischof y Kingstone, 2009; Rolls, 2008; Wieser et al., 2014), que distan bastante del ambiente

complejo donde la percepción se lleva naturalmente a cabo (Gallant, Connor y Essen, 1998; Vinje y Gallant, 2000), y que raras veces combinan enfoques para aproximarse al objeto de estudio.

A grandes rasgos, se proponen dos tipos de mecanismos perceptivos: Bottom-up y Top-down (Corbetta y Shulman, 2002; Desimone y Duncan, 1995; Kastner y Ungerleider, 2000; Posner, 1980). Si la atención es atraída de manera automática por características sobresalientes del estímulo visual, independientes de la historia de aprendizajes, de la tarea que el sujeto está realizando en el momento, de sus expectativas, volición, cultura y conocimiento previo, estamos hablando de atención exógena o procesos bottom-up. Parece relativamente fácil distinguir que las propiedades físicas (de bajo nivel) de estímulos como el movimiento, los bordes, los ángulos o el contraste gatillan saliencia del tipo bottom-up; y que otros estímulos, que presentan propiedades semánticas, como las ilusiones ópticas (como las de Müller-Lyer, Poggendorff y Zöllner), cubos de Necker, textos o logos de marcas, llaman nuestra atención desde lo cognitivo, *i.e.* de manera endógena o top-down (Borji et al., 2013b; LeMeur y Liu, 2015; Navalpakkam e Itti, 2005; Pinto et al., 2013; Posner, 1980; Tatler, 2014). Sin embargo, afirmaciones tan taxativas como esta son cuestionadas hoy en día por estudios que destacan que la percepción de objetos ambiguos puede ser modulada por diversos factores exógenos y endógenos independientes; e incluso que la clasificación de un proceso como top-down o bottom-up dependerá en parte de cuáles entre las definiciones disponibles de dichos conceptos son seleccionadas por los autores (Awh et al., 2012; Barragan-Jason, Besson, Ceccaldi y Barbeau, 2013; Kornmeier, Hein y Bach, 2009). La división entre procesos bottom-up y top-down no es la única demarcación posible de procesos perceptivo-atencionales, y dista de ser prístina o evidente. La literatura científica plantea distintos modelos teóricos que presuponen mecanismos diferentes y diferenciables empíricamente (Chica et al., 2013; Corbetta y Shulman, 2002; Coul, Frith, Büchel y Nobre, 2000; Einhäuser, Spain y Perona, 2008; Hahn, Ross y Stein, 2006; Pinto et al., 2013), no obstante, existe una estrecha interacción entre mecanismos, tanto a nivel de redes neuronales (Borji et al., 2013b; Sarter, Givens y Bruno, 2001) y de mecanismos perceptuales (Henderickx, Maetens y Soetens, 2012; Henderson y Hollingworth, 1999; Kornmeier et al., 2009; Mannan, Kennard y Husain, 2009; Nyström y Holmqvist, 2008), como a nivel ontogenético (Hubel y Wiesel, 1963; Valberg, 2005). Todo esto lleva a que algunos autores rechacen de pleno la

distinción bottom-up y top-down por considerarla insuficientemente explicativa (Awh et al., 2012; Wang, Kristjansson y Nakayama, 2005).

Hace ya casi un siglo atrás, Buswell (1935) notó que en escenas que contienen figuras humanas, estas figuras eran desproporcionadamente más atendidas. Y dentro de las figuras humanas, los rostros humanos tienen una especial relevancia biológica y social para nuestra especie (Jingling, Lin, Tsai y Lin, 2015; Theeuwes y van der Stigchel, 2006), por lo que resultan muy "salientes" pese a que no presenten características físicas (*i.e*. de bajo nivel) que atraigan particularmente la atención bottom-up (por ejemplo, en cuanto contraste, intensidad o color) (Birmingham et al., 2009; Cerf et al., 2009). Los sujetos tienden a fijar muy rápidamente su vista en rostros, con independencia de la tarea que estén realizando o el objetivo explícito que reporten (Bindemann, Burton, Hooge, Jenkins y de Haan, 2005; Birmingham et al., 2009; Cerf et al., 2009; Fletcher-Watson, Findlay, Leekam y Benson, 2008; Theeuwes y van der Stigchel, 2006). Sumado a lo anterior, mirar algunos rostros o hacer contacto visual, activa áreas cerebrales relacionadas al placer y la recompensa, como la corteza órbitofrontal medial, el tálamo anterior o el estriatum ventral (Kampe, Frith, Dolan y Frith, 2001; O'Doherty, 2003). Dicho esto, se podría hipotetizar que son factores cognitivos top-down los que determinan su elevada captura atencional. Sin embargo, se ha mostrado que los rostros atraen la atención desde edades muy tempranas (Di Giorgio, Turati, Altoè y Simion, 2012) e independiente de la tarea que se esté realizando (Cerf et al., 2009), ambos fenómenos usualmente atribuidos a procesos bottom-up. Considerando lo anterior, no resulta para nada sencillo clasificar la percepción de rostros humanos dentro de uno u otro mecanismo (Nyström y Holmqvist, 2008).

La presente memoria propone un experimento que aporte a la comprensión del mecanismo a la base de la percepción de rostros. Es conocido que el contexto influye en la interpretación de los rostros (Wieser et al., 2015); la pregunta que queda abierta, es si el contexto influye en la percepción de los rostros; si dicha captura atencional tiene a su base mecanismos automáticos que dependen de la saliencia física de los rostros –de algún modo que escapa a la computación en mapas de saliencia–, o bien si se utiliza información contextual para orientar la vista hacia los rostros. Mediante el uso de una cámara infrarroja Eye-Tracker se estudiará cómo los sujetos exploran imágenes naturalistas para determinar si la atención se dirige inmediatamente (automáticamente) a los rostros, o si se explora rápidamente la imagen y se utilizan elementos como el cuerpo para orientar la atención eficazmente hacia los rostros.

Antecedentes Teóricos

Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte.

—Blaise Pascal

2.1. Percepción y sistema visual

La capacidad de reaccionar a ciertos aspectos del ambiente es un fenómeno invariablemente presente en todo lo viviente, desde organismos unicelulares (Peil, 2014; Yoshimura, 2011) hasta plantas (Mescher y de Moraes, 2014) y animales complejos. Esta capacidad viene dada por la existencia de proteínas receptores en la membrana de las células receptoras del organismo.

Aún existen controversias acerca de dónde en el árbol filogenético, dicha capacidad de detección o sensación, que caracteriza lo viviente; pasa a ser percepción –una colección de mecanismos más complejos para relacionarse con el entorno–, controversia que depende de cómo entendamos y definamos la percepción. En la filogenia de organismos con sistema nervioso (SN), es común el desarrollo de sentidos, esto es, sistemas (órganos, en varios *phyla*) conformados por agrupaciones de receptores altamente especializados que responden a estímulos de determinada modalidad. La percepción entonces, corresponde al proceso de organización e interpretación de las señales sensoriales multimodales producto de la actividad integrada de varios sub-sistemas neurales y sensorimotores dinámicamente acoplados de un organismo con su ambiente (Noë, 2004; Pomerantz, 2003; Uttal, 1981). Los estímulos provenientes de los distintos sistemas sensoriales se combinan generando una señal integrada (Zmigrod y Hommel, 2013), lo cual permitió aumentar y hacer más rápida la detección, localización y respuesta a estímulos biológicamente significativos (Cook, Carvalho, y Damasio, 2014; Moreno y Lasa, 2003; Stein, Stanford y Rowland, 2009).

La visión apareció en el árbol filogenético hace aproximadamente 350 millones de años, en artrópodos del periodo cámbrico temprano; de ahí a la fecha, las variedades de órganos visuales se han complejizado y diversificado enormemente (Land y Nilsson,

2012). En los vertebrados se fue desarrollando uno de los sistemas visuales más sofisticados, que consiste en ojos tipo cámara (en contraste con los ojos compuestos de insectos y crustáceos) que presentan fotorrecepción direccional y no direccional, a la vez que visión espacial tanto de alta como de baja resolución. A diferencia de varios mamíferos, cuya modalidad sensorial dominante es el olfato, cuando llegamos a los primates haplorrinos, la visión es un sentido que presenta dominancia y una altísima especialización morfológica (Ross y Kirk, 2007).

2.2 Movimiento ocular

Para percibir contamos con órganos sensoriales asociados a vías neuronales que integran los distintos estímulos externos (en el caso de la visión: forma, movimiento, profundidad, tamaño, longitud de onda de radiación electromagnética) captados por separado mediante la reacción de receptores específicos ante estos, además de la excitación/inhibición de circuitos neurales especializados. Percepción y acción no se dan de manera secuencial, sino simultánea y son condiciones necesarias una de la otra.

La porción del mundo que somos capaces de ver con detalle sólo es la que corresponde a la suma de los campos receptivos de las neuronas ganglionares que se disponen alrededor de la fóvea, “la pequeña región circular (de alrededor de 1.5mm de diámetro) en el centro de la retina que está densamente empaquetada con los fotorreceptores conos” (Purves et al., 2008, p. 499). Si uno quiere ver algo con claridad, debe dirigir las fóveas de los ojos (correspondiente al centro de nuestro campo visual) hacia ese algo. A ese proceso se le llama “foveación”, y es usualmente el objetivo de los movimientos oculares que realizamos.

Los tipos de movimientos oculares pueden clasificarse funcionalmente de tres maneras:

(a) De mantenimiento de la mirada (involuntarios o reflejos)

(b) De desplazamiento de la mirada

   a. Sacádicos: Cambios bruscos de fijación que demoran 10-30ms, tiempo en el cual el sujeto no ve. Son balísticos (una vez iniciados, no pueden ser detenidos ni corregidos en respuesta a un cambio de posición del objeto en un tiempo menor a la sacada) y estereotipados (cada vez que hacemos

una sacada de un tamaño particular, nuestros ojos siguen el mismo patrón de movimiento).

b. de Seguimiento suave [*smooth pursuit*].

c. de Vergencia.

(c) Micromovimientos de fijación binocular

a. Trémores (o *Drifts*).

b. Fluctuaciones.

c. Micro-sacádicos. (Martínez y Pons, 2004).

En condiciones normales, realizamos 3-4 sacadas por segundo (Snowden et al., 2012), estas se intercalan con periodos –usualmente de alrededor de los 200-400ms en ambientes reales– donde los ojos se mantienen casi estacionarios en un punto, a estos se les llama Fijaciones (Land y Tatler, 2009; Land, 2011).

Por diversos motivos, los seres vivos no podemos reaccionar instantáneamente a los cambios del ambiente, en el caso de la percepción visual, entre la presentación de un estímulo y el inicio de movimiento ocular hacia dicho estímulo (o alejándose de éste: antisacada) ocurre cierta demora, a la que se le llama **latencia sacádica** (o tiempo de reacción). En estudios experimentales en conducta ocular, esta corresponde al tiempo (ms) que demora un sujeto entre la presentación de una imagen y la primera sacada que realiza hacia algún estímulo en particular. En sujetos no entrenados esta va de 100 a 500ms. El hecho de que existan sujetos con un tiempo de respuesta ocular en promedio cinco veces más rápida que otros presenta problemas importantes a la hora de generalizar resultados (Sumner, 2011). Por ello resulta necesario determinar las causas de la amplia variación de la latencia. Esta varía según tipo de atención (endógena o exógena, ver apartado 4.1) y características del estímulo (más latencia a mayor excentricidad, frente a estímulos cromáticos en comparación con lumínicos, de alto contraste en comparación con de bajo contraste). Es posible afirmar que la atención es un factor determinante potente de la variación de la latencia (Sumner, 2011).

2.3 Modelos explicativos y predictivos de la percepción visual: Saliencia, atención y mecanismos perceptivos

Conocer las estructuras nerviosas involucradas en la conducta ocular, y conocer las clases de conductas oculares posibles en primates humanos no resulta suficiente para explicar los fenómenos de la percepción visual ni de la conducta ocular. Cuando realizamos una sacada desde una región de nuestro campo visual hacia otra, ¿qué nos lleva a hacerla? ¿son propiedades del objeto/región del mundo que capta nuestra atención de manera automática (i.e. innata en algún sentido a nuestro acervo genético), es nuestra historia de aprendizaje asociativo, es nuestra capacidad cognitiva de control voluntario de los músculos oculares o la combinación de algunas de las anteriores según sea el caso? ¿Es posible predecir la conducta ocular? ¿Cuántos parámetros (información empírica contextual) necesitamos para realizar predicciones significativas en un individuo particular? Y aún si respondiésemos a todas esas interrogantes, quedan aún las preguntas más básicas: ¿Existen estímulos o características universales (comunes a toda la especie) a las que miramos preferentemente? ¿De existir, cuántas de estas compartimos con otras especies? ¿Cuánto de nuestros patrones de exploración visual es condicionable/aprendible? ¿Cuánto de nuestra historia individual y cultural (*i.e.* biografía, personalidad, epigenética) se refleja en nuestros patrones de exploración visual? ¿Cuánto de nuestra historia evolutiva como especie? ¿Cuánto de nuestro estado emocional?

Queda muy por encima del objetivo puntual de este trabajo el abordar siquiera un tercio de estas interesantes interrogantes, sin embargo considero útil el dejarlas planteadas como contexto que permite el florecimiento de distintas hipótesis explicativas que buscan dar cuenta de la percepción visual humana.

2.3.1 Modelos atencionales

Una escena visual típica tiene muchos más elementos que los que podemos atender a la vez. La atención es la capacidad cognitiva de enfocar la percepción en uno o un grupo de estímulos relacionados entre sí, excluyendo o suprimiendo los estímulos irrelevantes (vanSwinderen, 2011). El constructo de "relevancia" juega un rol doble en la explicación y la definición de la atención, por un lado, se refiere a que la atención está vinculada con la historia de la especie y del animal: habrían estímulos filogenéticamente relevantes, junto con estímulos ontogenéticamente relevantes (historia de selecciones e historia de

refuerzos/castigos) –explicación distal–, y a su vez, estímulos u objetos cuya relevancia está sesgada por el estado mental del animal en tiempo presente: metas/tareas [*current goals*], emoción actual, selecciones recientes y refuerzos/castigos recientes (Awh et al., 2012) –explicación proximal–. Por esto último, ha sido propuesta como un prerrequisito de la conciencia (Chica et al., 2013; Posner, 1978), sin embargo, aún hoy en día la relación entre los procesos atencionales y la conciencia resulta controversial (cfr. Chica, Botta, Lupiañez y Bartolomeo, 2012; Pinto et al. 2013; Tallon-Baudry, 2012), sobre todo cuando entran en juego las distinciones entre tipos de atención que veremos a continuación. Posner (1980) propone que la atención visual presenta las funciones de alertar (seleccionar regiones importantes de nuestro campo visual) y de búsqueda (seleccionar un objetivo en escenas hacinadas [*cluttered*]). Del estudio de los mecanismos de dirección de la atención visual, se postularon dos tipos de atención distinguibles mediante el criterio principal de involuntariedad/voluntariedad de la fijación de la mirada:

(a) Exógena (Bottom-up o transitoria). Atención basada en la saliencia del estímulo visual, *i.e.* características físicas de estos, tales como color, luminancia, textura, contraste, orientación, y movimiento que "salta a la vista" [*"pop out"*] al diferir marcadamente de su fondo [*background*]; con independencia del estado mental del observador (Awh et al., 2012). Esta separación es primero realizada por Posner (1980) quien escribe sobre el control periférico o reflejo de la orientación (en contraste al control central de la orientación, luego llamado top-down). Está presente en especies simples (e.g. *Drosophila melanogaster*), por lo que se asume que es una forma filogenéticamente más primitiva de atención. Se caracteriza por ser automática (involuntaria). Desplegarla toma entre 100 y 120 ms, afecta la discriminación de orden temporal empeorándola. Usualmente, se relaciona el fenómeno de Inhibición del Retorno (IdR, *en inglés IOR*) – que consiste en la facilitación de la atención en una región señalizada, seguida por una disminución de la velocidad y precisión para volver a atender a dicha región – únicamente con este tipo de atención. Estos movimientos se asocian al colículo superior y al área lateral intraparietal. (Borji et al., 2013b; Carrasco, 2011; LeMeur y Liu, 2015; Pinto et al., 2013; Peelen, Heslenfeld y Theeuwes, 2004: Treisman y Gelade, 1980).

(b) Endógena (Top-down o sostenida). Atención dirigida voluntariamente, dependiente de fenómenos cognitivos como conocimiento, expectativas, experiencias previas, recompensas, memoria, estado de ánimo o emociones presentes, metas, contexto

global de la escena visual y requerimientos de tareas. Su función es aumentar el procesamiento de estímulos relevantes mediante la examinación de los elementos en el campo visual uno a uno, muchas veces en ausencia de diferencias sustanciales entre el *target* y el fondo de la imagen; facilitar la discriminación entre señal y ruido o distractores y predisponer al sujeto a atender a ciertas regiones donde el estímulo puede aparecer. Desplegarla toma entre 150 y 300ms si esta es dirigida hacia una región espacial (e.g. esquina superior izquierda de la pantalla), y entre 300 y 500ms si está dirigida a características específicas u objetos (e.g electrodomésticos, objetos rojos). Mejora la discriminación de orden temporal, aumenta el contraste e iluminación percibidos. El control de estos movimientos se asocia al lóbulo frontal, especialmente campo frontal ocular y corteza prefrontal dorsolateral (Borji et al., 2013b, Carrasco, 2011; Katsuki y Constantinidis, 2014; LeMeur y Liu, 2015; Liu, Stevens y Carrasco, 2007; Pinto et al., 2013; Posner, 1980; Rolls, 2008; Sarter et al., 2001).

Ya en 1980 Posner plantea la dificultad en la separación de ambos mecanismos: “Comparisons of exogenous (reflexive) and endogenous (central) control of orienting is made difficult because external signals do not operate completely reflexively but will only summon attention and eye movements if they are important to the subject” (p.19). Corbetta y Shulman (2002) plantean que factores que afectan la atención, como novedad y sorpresividad, implican la interacción dinámica entre mecanismos top-down y bottom-up.

De todos modos, es posible generar otra clasificación de la atención visual según si efectivamente involucra o no movimientos oculares:

(a) Orientación abierta o explícita (*overt*). Orientación selectiva de la atención hacia un objeto o región espacial que implica que los ojos apunten a dicha dirección. Puede ser gatillada exógena (Bottom-up) o endógenamente (Top-down). Las secuencias de fijaciones se denominan trayectorias de escaneo visual [*visual scanpaths*] (Posner, 1980; LeMeur y Liu, 2015).

(b) Orientación encubierta o implícita (*covert*). Cambio cognitivo del foco de la atención sin cambio de posición o movimiento de los ojos. Es estudiada mediante electrooculografía y eye-tracking en tareas de filtro y señalización. La atención implícita disminuye la latencia de reacción del sujeto. (Posner, 1980; LeMeur y Liu, 2015).

Para Itti y Koch (2000) se puede afirmar operacionalmente que un estímulo fue atendido sólo si este entra a la memoria de corto plazo y permanece ahí el suficiente tiempo para poder ser voluntariamente reportado. En animales no humanos, la única forma de estudiar atención y percepción visual es mediante la orientación explícita, al no poder acceder a autorreporte o a tareas explícitas de búsqueda o memoria, resulta necesario inferir los procesos atencionales desde patrones conductuales o similitudes fisiológicas (vanSwinderen, 2011).

Un esquema simplificado de la relación entre movimientos top-down, bottom-up (entendidas de manera atencional, como en el apartado 2.3.1) y activación de áreas cerebrales, puede observarse en la figura 1.

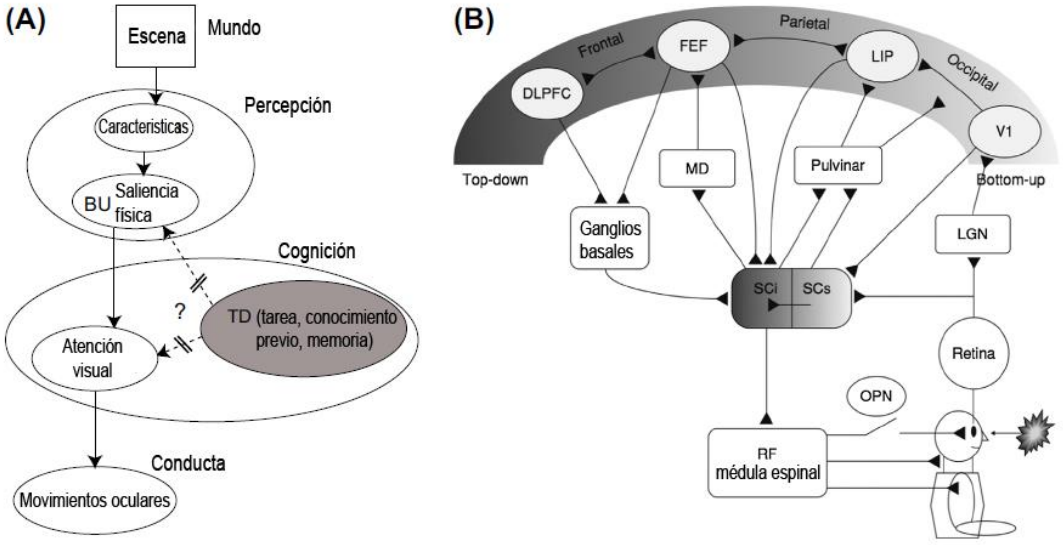


Figura 1. Esquema simplificado (A) del control de la mirada y (B) de las estructuras del SN que participan en la visión. BU= bottom-up; TD= top-down; las abreviaturas en inglés de (B) son DLPFC=corteza prefrontal dorsolateral; FEF= campo frontal ocular; LIP= área lateral intraparietal; V1= corteza visual primaria; MD= tálamo mediodorsal; LGN= núcleo geniculado lateral; SCi/s= colículo superior, capas intermedias/ capas superiores; OPN= neuronas omnipausa y RF= formación reticular. (modificado de Borji et al., 2013b).

A grandes rasgos, las dificultades de la clasificación atencional top-down vs. bottom-up pueden ser agrupadas en dos niveles: el ontogenético/diacrónico y el sincrónico. Sin embargo, para profundizar en los problemas de esta clasificación, es necesario primero ver otros modelos explicativos y predictivos relacionados a la percepción visual (apartados 2.3.1.1, 2.3.2 y 2.3.3), y realizar ciertas precisiones conceptuales, dada la ambigüedad de

estos términos. Las críticas a esta clasificación – junto con la proposición de una alternativa – serán elaboradas en el apartado 2.3.4.

2.3.1.1 Modelos predictivos de la atención

Con el objetivo de poder predecir la orientación implícita de la mirada en humanos, Koch y Ullman (1985) introducen la idea de Mapa de Saliencia, planteando un modelo explicativo de la atención involuntaria y automática, compuesto por dos mecanismos hipotéticos: el primero es la red neuronal Ganador-se-lleva-todo [*Winner-take-all network*], que detecta la región más activa en el mapa de saliencia y dirige la atención hacia esta, el segundo mecanismo es el mapa de saliencia mismo, que mide la visibilidad de una región en la escena visual, entregando un panorama sesgado de ésta, producido por la competición entre mapas de Color, Intensidad y Orientación. Posteriormente Itti, Koch y Niebur (1998) diseñan un modelo que busca predecir –de manera probabilística– la dirección y los cambios de la atención a estímulos simples en competencia. El mapa resulta en la predicción de la región de la imagen correspondiente al objeto más saliente, que luego es suprimida transitoriamente (por 500-900ms) mediante un mecanismo de IdR, para dar paso a la segunda región más saliente (en 30-70ms) y así sucesivamente. Se han formulado diversos modelos computacionales basados en esta idea, y algunos han sido capaces de predecir movimientos oculares explícitos y orientación implícita también en imágenes complejas o naturales, teniendo mejores resultados en estas últimas cuando se le agrega la información semántica de alto nivel al mapa de saliencia (Cerf et al., 2009; Itti y Koch, 2000; Wei y Luo, 2015).

Este modelo ha sido criticado en distintos aspectos por diversos autores, siendo posible resumir las críticas en los siguientes puntos: (1) Incluso si las características del mapa de saliencia son buenos correlatos probabilísticos de las fijaciones, esto no implica que necesariamente desempeñen un papel causal en la atención (Carmi e Itti, 2006; Tatler, 2007; Einhäuser et al., 2008), (2) el contenido semántico resulta más significativo y atrae la atención por encima de las predicciones de los mapas de saliencia en primates, tampoco predice mejor que el azar las fijaciones en ojos de otras personas en contextos sociales (Birmingham et al., 2009; Kayser, Nielsen y Logothetis, 2006), (3) los patrones y cantidad de sacadas y fijaciones realizadas son afectados por las instrucciones que se le den a los sujetos. E incluso cuando se les pide que exploren libremente, los movimientos oculares están fuertemente influenciados por propiedades de orden superior de la escena

y por factores idiosincráticos, como estado de ánimo, cultura, género e intereses personales (Borji et al., 2013b, Buswel, 1935, Henderson, 2003; Land y Hayhoe, 2001; Tatler, Hayhoe, Land y Ballard, 2011; Yarbus, 1967), (4) además, estos modelos presuponen que la saliencia es computada en áreas tempranas de la corteza visual primaria (V1) (Itti y Koch, 2000; Li, 2002), sin embargo, la evidencia favorece más bien la hipótesis de que la saliencia es computada en áreas frontales de jerarquía visual, como los campos frontales oculares, donde se representaría la saliencia y en las conexiones que tiene con V4 (Einhäuser et al. 2008).

En respuesta a algunas de las críticas antes presentadas, han surgido otros modelos predictivos, todos ellos integrando atención top-down y bottom-up (Kollmorgen, Nortmann, Schröder y König, 2010; Navalpakkam et al., 2005; Wei y Luo, 2015) sin embargo, existen autores que consideran necesario ir más allá de los factores atencionales y generar un modelo de la percepción visual que se base en las características de estímulos visuales. Estas propuestas serán abordadas a continuación.

2.3.2 Modelo centrado en características de los estímulos visuales: Alto nivel vs Bajo nivel

Henderson y Hollingworth (1999) plantearon que las fijaciones más tempranas en una escena están determinadas por características de bajo nivel [*low-level features*] tales como color, contraste, textura, movimiento, titileo, orientación (Le Meur y Liu, 2015; Wei y Luo, 2015) y la generación de representaciones de superficies y bordes, que no están influenciados por factores semánticos de la escena. Previo a realizar una sacada para dirigir la mirada a una región específica, algunas características de dicha región deben ser analizadas como estímulos en la visión periférica para permitir que los procesos de selección operen; los bordes de alto contraste como aristas, alto contraste de luminancia o características de 2do orden como esquinas son detectados rápidamente en visión periférica (Fletcher-Watson et al., 2008; Foulsham, Barton, Kingstone, Dewhurst & Underwood, 2011). Luego, la visión de nivel intermedio tiene que ver con la extracción de la forma y relaciones especiales mediante procesos selectivos en serie sin recurrir a elementos semánticos. Por último, las características de alto nivel emergen de formas específicas de organizar características de bajo nivel (Rouw, Kosslyn y Hamel, 1997), corresponden al mapeo de elementos semánticos y las representaciones visuales: identificación de objetos (mediante por ejemplo, leyes gestálticas como clausura y

simetría de la forma) y escenas, estas últimas entendidas como *"a semantically coherent (and often nameable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner"* (Henderson y Hollingworth, 1999, p. 244). El *gist* de la escena sería una propiedad de alto nivel. Desde la teoría de integración de características, se explica que la búsqueda de objetos requiera un escaneo visual lento, dado que se necesita tiempo para integrar las características de bajo nivel de la escena en objetos definidos (Torralba et al., 2006).

Las características de alto nivel no son reducibles a la atención endógena ni viceversa; las primeras dependen de la escena en su conjunto y no de la velocidad con la que son atendidas, en ese sentido, una propiedad de alto nivel puede ser o no ser procesada pre-atentivamente (Rouw et al., 1997). Autores como Borji e Itti (2013) afirman que los rostros humanos pertenecen a esta clasificación. A la hora de estudiar IdR también pueden apreciarse diferencias entre ambos marcos conceptuales (i.e. alto-bajo nivel vs. atención endógena-exógena), la IdR ocurre cuando se orienta la atención mediante saliencia física de estímulo, tradicionalmente se creía que la atención exógena y el procesamiento de estímulos de bajo nivel eran indisociables, estudios posteriores muestran que la ocurrencia de IdR en atención endógena es posible si el procesamiento de estímulos de bajo nivel está implicado (Henderickx et al., 2012). Sumado a lo anterior, un estudio encontró que la atención top-down (control voluntario) puede cambiar su acoplamiento de características de alto nivel a características de bajo nivel y viceversa (Al-Aidroos, Said y Turk-Browne, 2012). Además de las características de estímulos, los movimientos oculares se ven influenciados por constreñimientos espaciales y propiedades del sistema óculomotor, ello llevó a recuperar el interés por el estudio de las sacadas de baja amplitud, pasadas por alto en gran parte de los estudios clásicos (Kollmorgen et al., 2010; Pastukhov, Vonau, Stonkute y Braun, 2012).

Einhäuser et al. (2008) concluyen de su estudio que existen efectos inmediatos de las características de alto nivel en la orientación de la mirada, y esto no puede ser pasado por alto si se quiere predecir la conducta ocular en contextos complejos. Por lo tanto, las regiones hacia las que los sujetos dirigen su primera sacada en escenas complejas semánticamente ricas son descriptas mejor por las ubicaciones de objetos que por *peaks* en un mapa de saliencia bottom-up. Los objetos serían mejores predictores de la saliencia temprana que las propiedades simples de los estímulos [*stimulus-driven*]. No obstante, Borji, Sihite e Itti (2013a) re-analizaron los datos de Einhäuser et al. y concluyen que la

interpretación de los datos de Einhäuser et al. se debilita bastante al considerar el sesgo de centro [*center bias*], esto es, la tendencia de los sujetos a realizar más sacadas y fijaciones hacia un radio alrededor del centro de imágenes naturales estáticas (Buswell, 1935; Tatler, 2007; Tseng, Carmi, Cameron, Munoz e Itti, 2009). Al usar otros modelos predictivos encontraron que la conclusión de que los objetos vencen a la saliencia no se sostiene (Borji et al., 2013a).

Como Egaña et al. (2013) señalan, aun cuando la dicotomía alto nivel vs bajo nivel no debe confundirse con Top-down vs Bottom-up, ambas definiciones comparten la idea de que tras los movimientos oculares, existen dos mecanismos que difieren significativamente en complejidad, modo y contexto en el que son activados (según la tarea a la que es expuesta el sistema).

2.3.3 Modulación de la percepción por procesos Bottom-up y Top-down

La contraposición entre "Bottom-up" y "Top-down" se utiliza ampliamente en neurociencias, sin embargo, los términos toman significados diferentes según el tópico que se esté tratando: resultan polisémicos. La caracterización de un proceso como top-down o bottom-up dependerá en gran medida de cómo los autores estén entendiendo esta dicotomía (Kornmeier et al., 2009), sin embargo, muchas veces ocurre que en los artículos, las definiciones de sus conceptos básicos permanecen implícitas.

Generalmente se entiende lo bottom-up como referido a sistemas anatómicos: proyecciones ascendientes, del SN periférico al central, o de lo subcortical a lo cortical. A su vez, se entiende lo Top-down como proyecciones descendentes, desde la corteza a áreas subcorticales, la médula, SN periférico, sistema inmune y/o endocrino. Sarter et al. (2001) argumentan que ese no es el caso en la percepción visual, en cambio, la regulación atencional Top-down o Bottom-up representa principios conceptuales. Para estos autores, los procesos Top-down describen mecanismos guiados por el conocimiento del sujeto, cuya función es realzar el procesamiento neuronal de estímulos sensoriales relevantes, para sesgar al sujeto hacia regiones particulares donde las señales (como elementos semánticos opuestos al ruido) podrían aparecer. Los procesos top-down no consisten únicamente en proyecciones descendentes, aunque sí son predominantes; la activación de procesos Top-down ha sido tradicionalmente identificada con la mediación frontal cortical de funciones ejecutivas; no obstante, Sarter et al. (2001) destacan el rol de

las proyecciones colinérgicas ascendentes del proscencéfalo basal hacia la corteza en los procesos atencionales top-down de atención sostenida. Los procesos Bottom-up describen los mecanismos neuronales a la base de la captura atencional guiada por las características del estímulo *target* y su contexto sensorial, buscando explicar la capacidad de los individuos de detectar *targets* y desencadenar procesos atencionales gatillados por la saliencia sensorial de los *targets*. Acá sí son dominantes las vías ascendentes hacia las áreas corticales superiores (de V1 a regiones temporales para identificación de objetos y de V1 a regiones parietales para localización). Es importante que en esta caracterización, los procesos Top-down y Bottom-up ya no serían constructos dicotómicos de atención, sino, procesos en sobreposición [*overlap*] que en la mayoría de las situaciones interactúan para optimizar el desempeño atencional.

2.3.4 Polisemia y la dicotomía conceptual top-down vs. bottom-up en visión ¿un caso perdido?

Sin embargo, dicha postura dista de ser hegemónica, Valberg (2005), Maldonado (2008) y McDowell, Dyckman, Austin y Clementz (2008) sí privilegian el criterio anatómico sobre el cognitivo para hablar de la distinción Top-down – Bottom-up. Por otro lado, Rolls et al. (2008) entienden, al igual que Sarter et al., dicha distinción como principios conceptuales, pero no necesariamente cognitivos (por ejemplo, olores pueden influenciar sabores). Lo Top-down abarcaría lo atencional (e.g. sesgar el procesamiento espacial o de objetos), además de lo que Richard Gregory bautizó “Percepción como inferencia o testeo de hipótesis” (e.g. percibir un cubo de Necker como tridimensional). Resaltando el rol de la corteza prefrontal en enviar señales que permiten sesgar de manera top down el procesamiento visual en la Corteza temporal inferior y la modulación de representaciones en la corteza órbitofrontal; empero, es importante que los *inputs* bottom-up dominen el proceso, de lo contrario lo que ocurriría serían alucinaciones. Finalmente, otros como Heinderickx et al. (2012), Pinto et al. (2013), Chica et al. (2013) y Katsuki y Constantinindis (2014), están usando la definición atencional de top-down y bottom-up al habar de los mecanismos neurobiológicos que están a su base.

Los problemas para separar bottom-up de top-down como criterios neuroanatómicos en el nivel diacrónico (a lo largo de la ontogenia de los organismos) tienen que ver (1) con el rol del aprendizaje en la atención exógena: si bien los estímulos visuales que captan nuestra

atención de modo exógeno son relativamente universales a la especie y dependen de vías biológicas bastante directas, ¿qué tan condicionable –adquirible o extinguible– es su saliencia? Un antecedente importante es el estudio realizado por Hubel y Wiesel (1963) donde concluyen que gatos adultos sólo pueden distinguir tipos de estímulos visuales a los que hayan sido expuestos en una ventana temporal correspondiente a las primeras semanas desde que abren los ojos. Para que se desarrollen las conexiones nerviosas que subyacen a la percepción y saliencia de determinadas características, es necesaria la estimulación temprana en ciertas "ventanas de desarrollo". A su vez, (2) la percepción bottom up es condición necesaria para que se desarrollen mecanismos top-down. Valberg (2005, p.11) señala que el condicionamiento o la adaptación de la percepción de estímulos, en este caso visuales, parecen corresponder a un proceso bottom-up de ajuste de la sensibilidad del SN a ciertas características del mundo externo (en este caso, tipos de estímulos visuales). Mediante dicho proceso el SN adquiere una estructura y por lo tanto conocimiento del mundo que será usado posteriormente para interpretar las impresiones sensoriales de manera top-down: la nueva información proveniente de los sentidos es comparada con la información previa.

> *This combination of 'bottom-up' and 'top-down' processes may appear as a closed loop with no beginning or end, except that it is limited to a particularly plastic period of the animal's life. It may be difficult to decide where one process begins and the other one ends during the plastic learning period* (Valberg, 2005, p.11).

Valberg se está refiriendo a top-down y bottom-up como procesos neuronales plásticos, no atencionales, que se desarrollan a través de la ontogenia en función de la relación con un ambiente complejo. No obstante, los sistemas neurales que controlan la atención visual se superponen con aquellos que controlan movimiento ocular, por lo que podría no ser posible atender a una región u objeto sin por lo menos preparar el sistema óculomotor para realizar una sacada hacia dicha región o alejándose de esta (Sumner, 2011).

Un último (y quizá más devastador) problema a la caracterización atencional bottom-up / top-down fue señalado por Awh et al (2012), que apunta justamente a la ambigüedad y polisemia con la que los autores utilizan estos conceptos en el área de percepción visual y atención, lo cual lleva a dicha dicotomía a ser teóricamente insostenible. A grandes rasgos, quienes utilizan la definición atencional buscan separar el control atencional automático vs voluntario, para lo cual equiparan bottom-up a control exógeno o guiado por

factores externos al sujeto; y a su vez, equiparan top-down a control endógeno; constructo que amalgama la atención guiada por (i) la meta o tarea actual del sujeto [*goal-driven attention*], con (ii) la historia de selección y (iii) la historia de recompensa. Por ejemplo, el *priming* –fenómeno que consiste en que la presentación de un estímulo, influencia la respuesta hacia un estímulo posterior– es caracterizado tradicionalmente como un efecto bottom-up en origen, ya que no puede ser contrarrestado por modulación top-down voluntaria (Maljkovic y Nakayama, 1994; Olivers y Hickey, 2010); sin embargo, otros autores afirman que el efecto de *priming* es top-down debido a que depende de lo que el sujeto aprendió en ensayos [*trials*] previos y no solamente del estado del estímulo (Wolfe, Butcher, Lee y Hyle, 2003). Theeuwes y Van der Burg (2011) midieron el tiempo de respuesta de sujetos cuando se avisa con 1.5 s de anticipación que el estímulo objetivo (una línea horizontal o vertical) aparecerá dentro de un estímulo saliente particular (e.g. "círculo rojo") de un arreglo de estímulos compuesto por círculos rojos y verdes; en los siguientes ensayos, si se repite la asociación, la saliencia aumenta su fuerza, y el sujeto dirige su mirada más rápido hacia los círculos rojos que aparecen (efecto acumulativo de *priming of pop-out*); pero si luego, se le avisa al sujeto que el estímulo aparecerá dentro de uno de los círculos verdes, estímulos que antes eran distractores, de todas maneras el sujeto no puede evitar estar predispuesto a ver los círculos rojos primero; aún si su objetivo volitivo explícito es no hacerlo. Lo mismo ocurre en los casos de historia de recompensa, cuyo efecto consistente en gatillar un sesgo de selección hacia la característica recompensada, que puede ir en contra de los objetivos de selección voluntarios en ensayos posteriores. En el caso de la saliencia bottom-up, Awh et al. (2012) resaltan la evidencia que señala que los estímulos con alta valencia emocional capturan la atención (independiente de la voluntad, historia de selección u objetivos del sujeto), este fenómeno puede ser incluso observado con estímulos compuestos por palabras; ambas clases de estímulos de alto nivel, pero cuya saliencia es descrita como bottom up por el hecho de ser automática/involuntaria.

La solución propuesta por Awh et al. (2012) es agregar la variable "historia de selección" (que incluye la historia de recompensa y asociaciones negativas) a un mapa de prioridad integrado por (i) Metas/objetivos actuales, (ii) Historia de selección y (iii) Saliencia física, como fuentes distintas de sesgo de selección, que se encuentran en competencia, pero también pueden operar de manera coordinada. A mi juicio, dicha conceptualización podría

ser aun insuficientemente clara. Con el objetivo de precisar la referencia de los conceptos a ocupar, ocuparé las combinaciones entre los siguientes criterios dicotómicos:

(a) *Goal-driven* vs*. goal-independent* (saliencia física e historia de selección)

(b) Bajo nivel (captura por saliencia física) vs Alto nivel (captura semántica de la mirada)

La utilidad de dicha clasificación reside en que pueden existir fenómenos en las 4 combinaciones, y dentro de los fenómenos *goal-independent*, estos pueden ser gatillados por la saliencia física o por la historia de selección (que puede responder al *priming* o a la asociación positiva / negativa). Fenómenos que muchos autores llaman top-down a secas, pueden ser "goal-driven – alto nivel", "*goal-independent* – alto nivel" o "*goal-independent* – bajo nivel", lo que permite una clasificación mucho más fina, y a la vez independiente de la idea de "predominio" de proyecciones corticales descendentes o que ascienden desde las terminaciones nerviosas.

2.4 Percepción de rostros

Los rostros son una fuente relevante de información en las especies sociales, dan a conocer sexo, edad, emoción y dirección de la atención; se ha estudiado bastante la preferencia que le dan los primates y sobretodo los grandes simios (humanos, chimpancés, gorilas y orangutanes) frente a otros estímulos (Tomonaga, 2010).

En humanos, estudios con resonancia magnética funcional (fMRI) y tomografía por emisión de positrones (TEP) arrojan que los rostros son preferentemente procesados en V4 y en la Circunvolución fusiforme localizada en la corteza extraestriada de la región occipito-temporal inferior, conformando el área facial fusiforme (AFF) (Kanwisher, MnDermott y Chun, 1997) además del área facial occipital, localizada en el lóbulo occipital. El AFF sin embargo también se encuentra relacionado con la memoria y responde a estímulos que no son rostros, por esto se ha propuesto que podría procesar todo conjunto de estímulos que compartan una forma común y para los cuales el sujeto tenga bastante experiencia; otra hipótesis plantea que la codificación de los rostros es distribuida por la vía ventral (de V1 a la corteza temporal inferior) y que ciertos objetos que no son rostros comparten vías comunes con los rostros (Golarai et al., 2007; Tsao y Livingstone, 2008). Una característica de la percepción de rostros es que estos se

procesan de forma holística, como totalidades no-descomponibles, sin embargo, esta característica emerge en la ontogenia, alrededor de los 10 años; anterior a ello, los niños muestran patrones de reconocimiento más analíticos (descomposición del objeto en partes-estímulos) (Joseph, DiBartolo y Bhatt, 2015). Además, resulta relevante que solemos percibir rostros ante la mínima configuración, a este fenómeno ilusorio se le conoce como pareidolia y es lo que está a la base de nuestra tendencia a crear representaciones de rostros, cuya expresión más minimalista puede encontrarse en los emoticones (e.g. :D , >:C). Cabe destacar que a nivel anatómico, el AFF también se activa ante la ilusión del vaso de Rubin, cuando los sujetos están percibiendo los dos rostros de perfil (Andrews, Schluppeck, Homfray, Matthews y Blakemore, 2002) y en casos de pareidolias, donde también está presente un fuerte componente pre-frontal (Liu et al., 2014). Por último, estudios en imaginería de resonancia magnética (MRI) en sujetos con prosopagnosia (en este caso, lesiones en área occipital facial, partes de los giros inferiores occipital y temporal, fusiforme, temporal y superior medial) muestran que estos no tienen conciencia [*awareness*] de ver rostros, pero que sí los procesan de manera distinta a otros objetos, lo cual destaca el rol de la vía occípito-lateral izquierda y V1 en la percepción básica de rostros, además del uso de elementos contextuales para suplir los déficits, que se evidencia en la exploración topográfica en vez de directamente dirigida hacia la mirada, boca o nariz de los sujetos control (Lê, Raufaste, Roussel, Puel y Démonet, 2003; Mannan et al., 2009).

La preferencia por los rostros en humanos no se ha escapado del clásico debate naturaleza vs. crianza, siendo aún hoy debatido si es parte del *bauplan* del desarrollo del sistema visual en neonatos (dada la enorme relevancia de los rostros en nuestra filogenia, no resulta descabellado especular que haya cierta especie de patrón de acción modal que nos lleve inmediatamente hacia ellos apenas tengamos la capacidad de verlos; especialmente, el rostro del cuidador principal, que suele ser la madre) o bien que se aprende rápidamente (en esta explicación el neonato debería poder improntarse –al más puro estilo de los patitos de Lorenz– hacia cualquier tipo de "cuidador" al que esté más expuesto en sus primeros días, sea humano, otra especie animal o aún un objeto inanimado, un experimento así no se ha ni se podría realizar, por obvias razones deónticas).

La evidencia empírica más temprana al respecto señala que los rostros humanos son visualmente atractivos ya a las 6 semanas de vida (Cashon y Cohen, 2003), pero que no

son preferidos frente a otros estímulos complejos hasta al menos los 6 meses. A los 3 meses los bebés muestran preferencia por el rostro de su cuidador por sobre los demás rostros y por rostros de su mismo grupo étnico sobre otros rostros. Las representaciones de rostros (en este caso, humanos animados en la serie Charlie Brown) no son preferidas hasta los 6 meses aproximadamente (Di Giorgio et al., 2012). Golarai et al. (2007) señalan que si bien hay estudios que muestran que los rostros resultan atractivos desde el mismo nacimiento, hay una maduración neural importante de la vía ventral y que el tamaño del AFF derecho es sustancialmente más grande en adultos que en niños de 7 años, la memoria de reconocimiento de rostros alcanza su nivel adulto cerca de los 16 años. Una hipótesis explicativa de la conducta ocular frente a imágenes o videos complejos en niños de 3 meses, es que es la saliencia bottom-up, o las características de bajo nivel lo que guía su atención principalmente (Di Giorgio et al., 2012). Estudios en esquizofrénicos adultos, sin embargo, muestran que estos presentan una conducta ocular hacia rostros distinta de los sujetos control (Benson et al., 2012; Chen, Norton, Ongur y Heckers, 2008; Chen y Ekstrom, 2016; Egaña et al., 2013), presentando un pobre desempeño en la atribución de estados emocionales, captura atencional disminuida, menos exploración en general, mayor cantidad de *drifts* y capacidad de inhibición de distractores disminuida. Ver qué áreas cerebrales, conductas y elementos relacionados al conocimiento previo e historia de selección del sujeto puede ayudarnos a entender los componentes y mecanismos de la percepción de rostros.

Con independencia de la tarea o de la complejidad visual de las imágenes, los rostros atraen consistentemente nuestra mirada en detrimento de otros estímulos salientes y aún si no hay tarea alguna que los involucre (Cerf et al., 2009; Fletcher-Watson et al., 2008), la mayoría de veces en estudios experimentales la primera fijación que realiza el sujeto en la imagen está dirigida hacia características internas del rostro, principalmente los ojos (Birmingham et al., 2009), se hipotetiza que esto puede deberse a factores cognitivos top-down como la información sobre el estado emocional y dirección de la atención de la otra persona que nos entregan sus ojos (Emery, 2000); sin embargo, también puede deberse a que los humanos tenemos esclerótica blancas excepcionalmente largas (en comparación con otros mamíferos y sobretodo los demás primates sociales) que contrastan manifiestamente con los múltiples colores que puede tener el iris y la piel (Kano y Call, 2012), lo que corresponde a características que atraen la mirada de manera bottom-up, aunque el patrón de exploración de rostros humanos por humanos es

particularmente idiosincrático (a diferencia de otros estímulos como paisajes o fractales) (Leonards y Scott-Samuel, 2005). No bastando con aquello, el panorama se complejiza aún más al considerar las diferencias específicas producidas por el contexto de socialización intensiva donde se enmarca el desarrollo de los organismos humanos (Miyamoto et al., 2006). Resulta interesante entonces que los rostros borrosos igual atraen fijaciones aun cuando tengan baja *signature* en un mapa de saliencia (Nyström y Holmqvist, 2008). Y es más intrigante aún, que la saliencia de los ojos y las cabezas, al ser computada en un mapa de Itti y Koch, es muy baja; son regiones en absoluto no-salientes (Birmingham et al., 2009). Sumado a todo esto, mediante estudios de señalamiento [*cueing*] de objetos en imágenes simplificadas, se encontró que los rostros presentan IdR, fenómeno asociado tradicionalmente a captura de atención involuntaria y gatillada por características de bajo nivel, pero que sin embargo "the IOR [IdR] effect for faces is not due to low-level feature differences but is due to the semantic processing of faces." (Theeuwes y Van der Stitchel, 2006, p.606; Silvert y Funes, 2016). Sea como fuere, en tareas donde se le presenta a los sujetos dos escenas distintas, a la derecha e izquierda en una pantalla gris; y se les da la instrucción de realizar una sacada lo más rápido posible hacia el objeto que pertenezca a la clase objetivo que se indica; los sujetos fueron capaces de iniciar una sacada hacia la escena que contiene el rostro a 100-110ms desde la presentación de la imagen, siendo el promedio 140ms (Crouzet, Kirchner y Thorpe, 2010) y en estudios de visualización de escenas naturales complejas donde se les pide a los sujetos presionar un botón apenas vean un rostro, la velocidad de respuesta (correspondiente a pulsar el botón) puede ser tan rápida como 250-300ms desde la presentación de la imagen (Rousselet, Mace, y Fabre-Thorpe, 2003). También, los rostros retienen la atención más que otros objetos (Bindemann et al., 2005) Un estudio con electroencefalograma (EEG) identifica en humanos dos potenciales relacionados a eventos (ERP) que son claves para la percepción de rostros: uno occipital (P1) que responde preferentemente a características de bajo nivel de los rostros como espectro de amplitud y color, siendo responsable de las sacadas más tempranas (100-110ms); y otro occipito-temporal (N170) que responde a la configuración holística de los rostros, pero más intensamente a los ojos (Rossion y Caharel, 2011; Nemrodov, Anderson, Preston e Itier, 2014).

Es sabido el rol del contexto en la modificación de la conducta ocular, tanto en su patrón como su velocidad (Torralba et al., 2006), sin embargo, hasta la fecha no se ha estudiado

si existe algún efecto atribuible al contexto, que guíe cotidianamente las sacadas hacia los rostros; así como tampoco se ha estudiado en extenso el rol que tienen los cuerpos en la latencia y dirección sacádica en imágenes naturalistas; se sabe, sin embargo, que las sacadas guiadas hacia cuerpos activan áreas cerebrales distintas que las sacadas guiadas hacia rostros (Morris, Green, Marion y McCarthy, 2008).

Fletcher-Watson et al. (2008) encontraron experimentalmente procesamiento extrafoveal de rostros (que se refleja en la precisión con los que los sujetos hacen su primera sacada en una imagen que contenga humanos, aun cuando las características de alto nivel sólo fueron advertidas con la periferia visual), además, utilizaron una condición de exploración espontánea y otra de tarea de discriminación (determinar el género de los sujetos presentes en las imágenes naturales), no encontrando diferencias significativas entre ambas: los rostros son fijados rápidamente y con precisión, la mirada de los sujetos fotografiados es seguida de igual manera en las dos condiciones y tendían a enfocar también los cuerpos y permanecer ahí en promedio el 40% del ensayo. Finalmente concluyen que "this bias towards fixating the social scene, and specifically the person therein, argues strongly for a stimulus-driven visual system that is tuned to high-level properties" (p. 582), postura que si bien queda caracterizada de manera más precisa como atención exógena, *goal-independent*, hacia aspectos semánticos (o de alto nivel), suele equipararse descuidadamente a la idea de mecanismo bottom-up.

Desde una perspectiva opuesta, Birmingham et al. (2009) señalan que los estudios que presentan el rostro junto con otros objetos altamente salientes (por características de bajo y alto nivel) en escenas naturales complejas permiten sugerir que los observadores seleccionan como primera fijación los ojos por la información social que proveen más que porque sean los ítems más salientes en la escena. Esto ocurre en diversas tareas, incluso cuando se pide al sujeto explícitamente extraer información social de la escena; la hipótesis que los autores sostienen para explicar esto, es la siguiente

> *Currently the working hypothesis is that observers have an early default bias to inspect the eyes of others, not because they are visually salient within the scene, but because they understand them to be socially communicative stimuli that provide important information about a social scene* (p. 2992).

Pareciera entonces que estuviésemos frente a un caso de *infradeterminación empírica de la teoría*: existen dos hipótesis explicativas que presentan adecuación empírica con un

conjunto de evidencia, que puede ser interpretado como caso confirmatorio de ambos marcos teóricos (cfr. Quine, 1975). Argumentaremos que esto se debe más a un problema de imprecisión conceptual que a un desacuerdo fundamental acerca de cómo interpretar los datos; en términos gruesos, uno podría afirmar que para Fletcher-Watson et al. la atención a rostros es bottom-up y para Birmingham et al. es top-down; si se adopta la terminología arriba propuesta, ambos afirmarían que la captura atencional es *goal-independent* y a su vez, responde a propiedades de alto nivel. Sin embargo, aún existe una diferencia entre ambos enfoques, que debe ser resuelta empíricamente: Fletcher-Watson et al. afirman que el mecanismo a la base de la captura atencional automática (*goal-independent*) que presentan los rostros es la saliencia física, en cambio, Birmingham et al. concluyen que es la historia de recompensa (o selección) desde la temprana estimulación social en la infancia.

Como Birmingham et al. plantean, “if saliency does contribute to placement of the first fixation within complex social scenes, it is most likely to do so for fixations resulting from early (faster) saccades than for fixations resulting from later (slower) saccades” (2009, p. 2993); para lo cual, realiza experimentos de observación libre (n=20) utilizando como estímulos, escenas sociales complejas que contenían una o tres personas haciendo algo (escenas activas: e.g. leer un libro) vs haciendo nada (escenas inactivas: e.g. sólo sentarse en una banca); en otro experimento dividieron a los participantes (n=39) en tres grupos, cada uno con una tarea distinta (observación libre, descripción de la escena y atención a la atención de las personas en las escenas); finalmente utilizaron dos grupos (n=18) para ver si los ojos eran percibidos como informativos; a uno le dijeron que después de la observación de imágenes, les harían un test de memoria, mientras que omitieron esta información al otro grupo. Concluyendo que la saliencia no explica los patrones de fijación de la mirada.

Nuestro experimento comparte esa hipótesis, y la metodología de exploración libre, además del uso de escenas del mundo real (que aquí llamamos escenas naturalistas), pero nos parece que para comprender los mecanismos tras la captura en la percepción visual de humanos sanos socializados, es necesario conocer primero a profundidad los patrones de exploración visual hacia rostros del modo más cercano posible a como ocurre en el mundo real, y un paso interesante hacia ello es el conocer si el cuerpo juega algún rol en la captura atencional provocada por los rostros.

Pregunta de investigación y Objetivos

Ein Glaube wie ein Fallbeil, so schwer, so leicht.

— Franz Kafka

3.1 Pregunta de investigación

¿La presencia o ausencia de un cuerpo acompañando a un rostro en una imagen naturalista causa una diferencia significativa en el direccionamiento atencional hacia los rostros?

3.2 Hipótesis de investigación

Al contrastar empíricamente la latencia de la primera fijación dirigida hacia un rostro, se espera que si existe influencia semántica de elementos asociados al rostro en el proceso atencional, las miradas se dirigirán más rápido al rostro cuando este está acompañado por un cuerpo.

3.3 Objetivo General

Determinar si la percepción visual de rostros está influida por factores contextuales (en este caso, el cuerpo) que faciliten la dirección de la mirada hacia estos.

3.4 Objetivos Específicos

3.4.1 Identificar, en adultos humanos sanos, patrones sacádicos de exploración libre de escenas naturalistas donde aparece el rostro y el cuerpo completo de un individuo humano.

3.4.2 Identificar en adultos humanos sanos, patrones sacádicos de exploración libre de escenas naturalistas donde aparece sólo el rostro como si estuviese detrás de un objeto.

3.4.3 Contrastar experimentalmente si la presencia de un cuerpo asociado a un rostro tiene algún efecto significativo en la latencia (en milisegundos) del primer movimiento sacádico dirigido hacia dicho rostro.

3.4.4 Comparar el tiempo de la primera fijación que cae dentro de un rostro, el tiempo de permanencia de la mirada dentro de un rostro y la cantidad de fijaciones que caen dentro de un rostro, entre escenas naturalistas con y sin presencia de un cuerpo adyacente al rostro, para ver si se presentan diferencias significativas.

Materiales y Métodos

Ein Bild hielt uns gefangen.

— Ludwig Wittgenstein

i. Sujetos

La muestra consiste en 29 sujetos mayores de 18 años (22 mujeres, 7 hombres), edad promedio 22 años (Desviación estándar: 3.77 años). Reclutados por conveniencia entre agosto y septiembre del 2015. Todos tenían visión normal o corregida a normal, sin interacción previa con experimentos de *eye-tracking*, desconocían el propósito del estudio y firmaron un consentimiento informado antes de participar en el experimento. El proyecto fue aprobado por el Comité Ética de Investigación en Seres Humanos (CEISH) de la Facultad de Medicina (Universidad de Chile).

ii. Instrumentos

Se utilizó un oculómetro modelo EyeLink 1000 (SR Research, Ltd., Canadá) a 500Hz de tasa de muestreo (una muestra cada 2 ms), para registrar el movimiento binocular de los sujetos.

En 18 sujetos, se realizó la calibración con una matriz de fijación de 9 puntos (3x3), en los casos donde el software no validó dicha calibración después de tres intentos, se calibró con 5 puntos. Se aplicó corrección de deriva [*drift*] cada tres imágenes.

Para la construcción del experimento, se utilizó el software Experiment Builder (SR Research, Ltd., Canadá).

iii. Procedimiento

Se presentó un set de 54 imágenes (1920x1080 pixeles) en una pantalla plana Viewsonic modelo VX2753mh-LED de 27 pulgadas, a una tasa de refresco [*refresh rate*] de 60 Hz y

una resolución de 1920x1080. Los sujetos apoyaban su mentón y frente en las almohadillas del soporte del EyeLink, a 55cm de la cámara y a 79 cm del monitor.

Figura 2. Oculómetro Eyelink1000, joystick y soporte. Imagen de referencia tomada de http://www.sr-research.com.

Las imágenes fueron presentadas en orden aleatorio por aproximadamente 3s cada una, siendo siempre antecedidas por cruces negras de fijación ubicadas en el centro de la pantalla con un fondo gris cuya duración fue de 2s sumado a un tiempo aleatorio para que los sujetos no aprendan a predecir la aparición de las imágenes y por lo tanto, no presenten conducta de anticipación frente al estímulo (ver Figura 3). Al presentar las imágenes por un tiempo inferior a 5s y enfocarse en las primeras sacadas se minimiza el efecto que puedan ejercer los factores top-down idiosincráticos (*goal-driven*, o puramente "voluntarios"). Sin embargo, esto no interfiere con la pregunta de investigación aquí planteada, dado que esto no elimina factores relevantes para el experimento, tanto el procesamiento semántico de los objetos como el contexto de escena global resultan accesibles a los sujetos en el tiempo que tienen para explorar las imágenes (Borji et al., 2013b; Hwang, Wang y Pomplum, 2011; Torralba et al. 2006).

Cada tres imágenes, se presentaba un punto de calibración frente al cual los participantes tenían que presionar cualquier botón de un joystick para pasar a la siguiente cruz de fijación que antecede a una imagen, esto se realizó para que los sujetos pudiesen descansar la vista el tiempo que estimen necesario y luego proseguir con el experimento. Se le pidió a los sujetos que explorasen libremente las imágenes que aparecerían en el monitor. Se utilizó un diseño de medidas repetidas, cada sujeto observó todas las imágenes pertenecientes a las tres condiciones que serán presentadas a continuación.

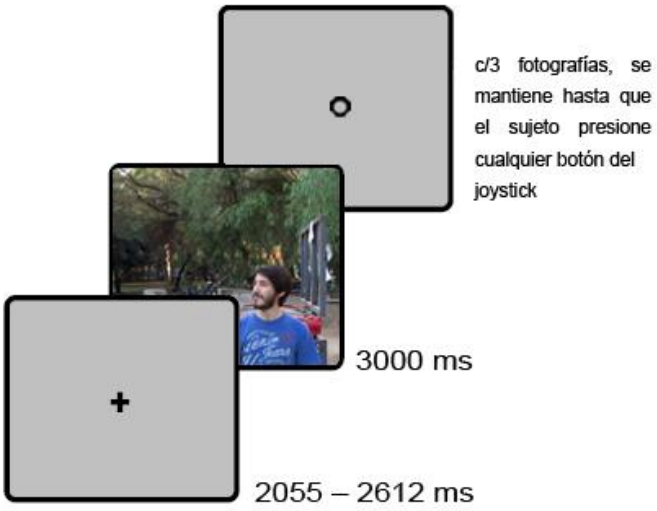


Figura 3. Protocolo experimental. La pantalla estaba configurada para presentar 1920x1080 pixeles sin dejar márgenes.

iv. Imágenes

El set de imágenes consiste en 3 versiones distintas de 18 fotografías de escenas naturalistas semánticamente ricas. Las fotografías fueron editadas en Photoshop CS4, generando finalmente 3 versiones para cada fotografía, la mayoría de la imagen se mantiene constante para las 3 condiciones, variando únicamente los pixeles que corresponden a la característica que se desea evaluar. Las versiones en cuestión son:

[i] Imágenes donde aparece una persona de cuerpo entero y de rostro identificable (**condición B**).

[ii] imágenes donde aparece el rostro de esa misma persona detrás de un objeto del mismo paisaje (**condición O**).

[iii] imágenes que sólo consisten en ese mismo paisaje urbano o natural (**condición L**, que sirve como control). Para ver la totalidad de las imágenes, dirigirse al **Anexo 1**.

Las fotografías buscan ser lo más ecológicas posible, *i.e.* cercanas a lo que los sujetos observan cotidianamente: constan con información semántica que compite con los rostros en cuanto orientación top-down de la atención o características de alto nivel (*e.g.* graffitis, libros, menús de almuerzo, afiches, letras, objetos tecnológicos vistosos, patentes de auto, personas de espalda o demasiado pequeñas como para que su rostro fuese

distinguible; elementos presentes en todas las imágenes excepto F004, F005, F010, F012, F013 y F014), también existen en las fotografías objetos que muy probablemente atraen la mirada de manera bottom-up (objetos de color rojo intenso: sillas, un corazón graffiteado de aprox.1400 pixeles –ver figura 11-, luces de automóvil, basureros o libros en las imágenes F018, F016, F011, F006 y F001 respectivamente; un árbol oscuro que contrasta con el fondo claro en F011). Sin embargo, se excluyó tajantemente la presencia de otros rostros distinguibles en las fotografías, con el objetivo de que no haya competencia entre rostros. El único rostro distinguible en cada fotografía de las condiciones B y O se encuentra en un radio de 6º de excentricidad, donde la región de características internas del rostro (complejo ojos-nariz-boca) presenta un tamaño de 5º en promedio, jamás excediendo los 6º (ver **figura 3**). La excentricidad de las áreas de interés (AI) generadas responde a que los sujetos suelen explorar más y explorar más rápido, objetos que se encuentran cerca de su centro visual, a esto se le conoce como "sesgo de centro" (Borji et al. 2013b; Tseng et al., 2009) es importante señalar que dicho sesgo está más presentes en condiciones de laboratorio (con escenas naturales) que en ambiente natural propiamente tal (Kollmorgen et al., 2010).

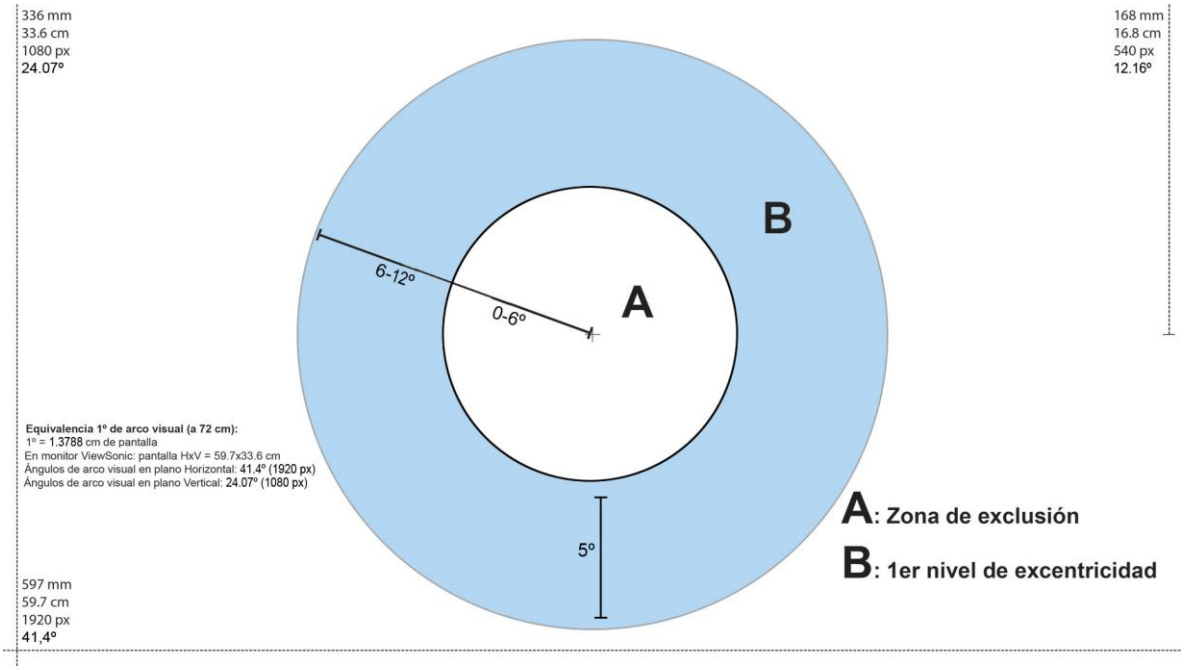


Figura 4. Protocolo utilizado para componer las imágenes naturales. Los rostros se encontraban distribuidos únicamente en la zona B. Para calcular ángulo visual se utilizó $S = 2 \cdot d \cdot \tan(\alpha/2)$.

Al editar las fotografías de la condición O, se buscó que el cuerpo pareciese estar oculto de manera natural por algo que estuviese delante de este (y que en la fotografía original, y en las condiciones B y L, está presente), esto para evitar el efecto de implausibilidad semántica o "*pop-out*" semántico (Loftus y Mackworth, 1978; Einhöuser et al., 2008). Ver apartado 6.6 y para un ejemplo, ver **figura 4**.

B

O

L

Figura 5. Imagen F014B, F014O y F014L (de arriba hacia abajo).

v. Extracción y análisis

Luego de completar los registros de los 29 sujetos, se obtuvieron 1566 registros de ensayos (522 de fotografías que pertenecen a cada condición –B, O y L–). Se utilizó EyeLink® Data Viewer 2.3.1 para generar manualmente AI's, que consistieron en elipsoides con área de 10221.96 pixeles alrededor de cada rostro (en caso de las condiciones B y O) y en el mismo lugar en la condición L donde no está presente el rostro (ver figura 6). Desde el Data Viewer se generó un reporte de AI que fue exportado a Excel, y que consistía en las columnas que se muestran en la siguiente tabla.

Tipo de dato (Columnas)	Descripción
Latencia primera Sacada hacia AI (ms)	Tiempo que transcurre entre el inicio de la imagen y el primer movimiento ocular tipo sacada del sujeto
Primera Fijación dentro del AI (ms)	Tiempo que transcurre entre el inicio de la imagen y la primera fijación ocular que cae dentro del AI.
Cantidad total de Fijaciones en el AI	Número de veces que el sujeto fija la mirada dentro de la región de interés definida para la imagen.
Tiempo de Permanencia en el AI (ms)	Milisegundos que el sujeto permanece con la mirada dentro de la región de interés.
% del ensayo Permaneciendo en el AI	Razón entre el tiempo total del ensayo (5s aprox.) y el tiempo de permanencia en el AI multiplicado por 100.
% total de Fijaciones en el AI	Razón entre las fijaciones realizadas dentro del AI en un ensayo y la totalidad de las fijaciones realizadas en ese ensayo, multiplicado por 100.

Tabla 1. Valores de cada columna generada mediante EyeLink.

Figura 6. Visualización de los ensayos 48, 45 y 35 (de izq. a der.) del sujeto NO240815. En azul se muestran las sacadas, en celeste las fijaciones (donde el número indica duración [ms]), en rojo los pestañeos y en naranja el AI.

Dado que los 5 segundos que se grababan en cada ensayo incluyen los 2 segundos de cruz de fijación y el tiempo aleatorio que se le agregó para evitar que los sujetos predijesen cuando aparecerán las imágenes (lo cual podría causar un efecto de disminución de la latencia por aprendizaje asociativo intra-experimento), se le restó a las columnas “Latencia primera Sacada hacia AI” y “Primera Fijación dentro del AI” los milisegundos exactos que duró en cada ensayo la cruz de fijación (que oscilan entre 2055ms y 2612ms; siendo el promedio 2072ms, DE:23.97ms; habiendo sólo 32 valores que se desvían más de una DE del promedio).

Posteriormente, se realizaron pruebas estadísticas con SPSS 17.0; se agruparon los datos según las condiciones (B, O y L), restando los datos perdidos. En caso de L, como se miden sacadas y fijaciones que caigan dentro del AI, se esperaría que la cantidad de datos no-perdidos sea igual a la probabilidad aleatoria de que los sujetos emitan sacadas hacia una región no saliente de 10221.96 pixeles; siendo más precisos, el n de datos no-perdidos para esa región debería corresponder a la probabilidad de fijaciones de la representación de esa región en un mapa de saliencia al aplicarle algún algoritmo (ya sea Itti y Koch, 2000; Navalpakkam e Itti, 2005 o Wei y Luo, 2015), cosa que no se hizo ya que escapa a los objetivos de este trabajo. Mediante SPSS, se realizó un análisis descriptivo para obtener medidas de tendencia central (media aritmética) y de dispersión (DE y varianza). Luego, para saber si la variable de estudio proviene de una población cuya distribución es normal, se aplicó el test Kolmogorov-Smirnov con corrección de significancia Lilliefors. Al no cumplirse el supuesto de normalidad, se procedió a trabajar con estadística no-paramétrica.

Resultados

Tu mirada me atrapa, algo tiene que me captura.

— Ñengo Flow

5.1 Estadísticos descriptivos

La cantidad total de ensayos por condición son 522, habiendo un 97.1% de ensayos válidos en la condición B, 95.8% en O y 21.5% en L para la variable “Latencia primera sacada hacia AI”. Un ensayo cuenta como válido si y solo si hay al menos una fijación dentro del AI desde que empieza a ser presentada la imagen (o sea, se excluyen todos los ensayos donde el sujeto mira por casualidad el AI antes de que haya un rostro o un fragmento de fotografía ahí). Entonces, es necesario separar los datos perdidos en dos categorías, (i) los que se perdieron por un falso positivo, o sea, hubo una fijación en el AI antes de que la fotografía apareciera, por lo que los valores en ms de la latencia de sacada y primera fijación eran negativos (B=14; O=18; L=29) y (ii) los ensayos donde simplemente no hubo ninguna fijación en el AI (B=1; O=4; L=381), siendo estos ensayos informativos para el experimento. Se privilegia la distribución de válidos/ perdidos/ no-fijación en la variable “Primera fijación dentro del AI” y no en “Latencia” porque las 6 sacadas que fueron iniciadas durante la cruz de fijación, al no haber aterrizado en el AI antes de que empezase la presentación de la imagen, no afectaron a lo que se busca medir, que es la captura atencional de los rostros una vez iniciada la imagen.

Condición (tipo de fotografía)			Latencia primera Sacada hacia AI (ms)	Primera Fijación dentro del AI (ms)	Cantidad total de Fijaciones en el AI	Tiempo de Permanencia en el AI (ms)	% del ensayo Permaneci-endo en el AI	% total de Fijaciones en el AI
B	N	Válido	503	507	522	522	522	522
		Perdidos	19	15	0	0	0	0
	Media		297.81	346.43	4.03	1479.59	32.15	35.05
	Desviación estándar		202.850	208.044	1.765	639.674	13.065	15.937
	Varianza		41148.325	43282.360	3.116	409182.642	17069.080	25398.188
	Mínimo		115	204	0	0	0	0
	Máximo		2377	2463	12	2772	62.66	88.89
O	N	Válido	498	500	522	522	522	522
		Perdidos	24	22	0	0	0	0
	Media		291.06	336.32	4.18	1528.68	33.13	36.47
	Desviación estándar		159.585	158.347	1.723	641.483	13.361	15.090
	Varianza		25467.493	25073.836	2.968	411500.878	17852.936	22769.591
	Mínimo		76	212	0	0	0	0
	Máximo		2301	2367	10	2810	67.38	85.71
L	N	Válido	112	112	522	522	522	522
		Perdidos	410	410	0	0	0	0
	Media		1352.50	1398.63	.41	103.00	2.29	3.12
	Desviación estándar		771.464	771.205	.798	224.374	4.909	6.230
	Varianza		595156.865	594757.732	.637	50343.697	2410.358	3881.262
	Mínimo		238	288	0	0	0	0
	Máximo		2995	3033	5	2000	40.29	40.00

Tabla 2. Estadísticos descriptivos de las tres condiciones.

Podemos observar la distribución de la variable “Latencia primera Sacada hacia AI” con más claridad en las figuras 7 y 8. Las condiciones B y O se parecen bastante a simple vista, presentando ambas una asimetría positiva (B= 5.925, error estándar= 0.109; O= 6.267, error estándar= 0.109), la condición L también tiene una asimetría positiva, pero mucho menor (0.555, error estándar= 0.228). La curtosis de B y O es positiva (B= 44.621, error std= 0.217; O= 62.206, error std.= 0.218), en el caso de L, la curtosis es negativa (-0.846, error std.= 0.453), lo cual señala que B y O presentan picos altos, y L tiende más bien a ser aplanada.

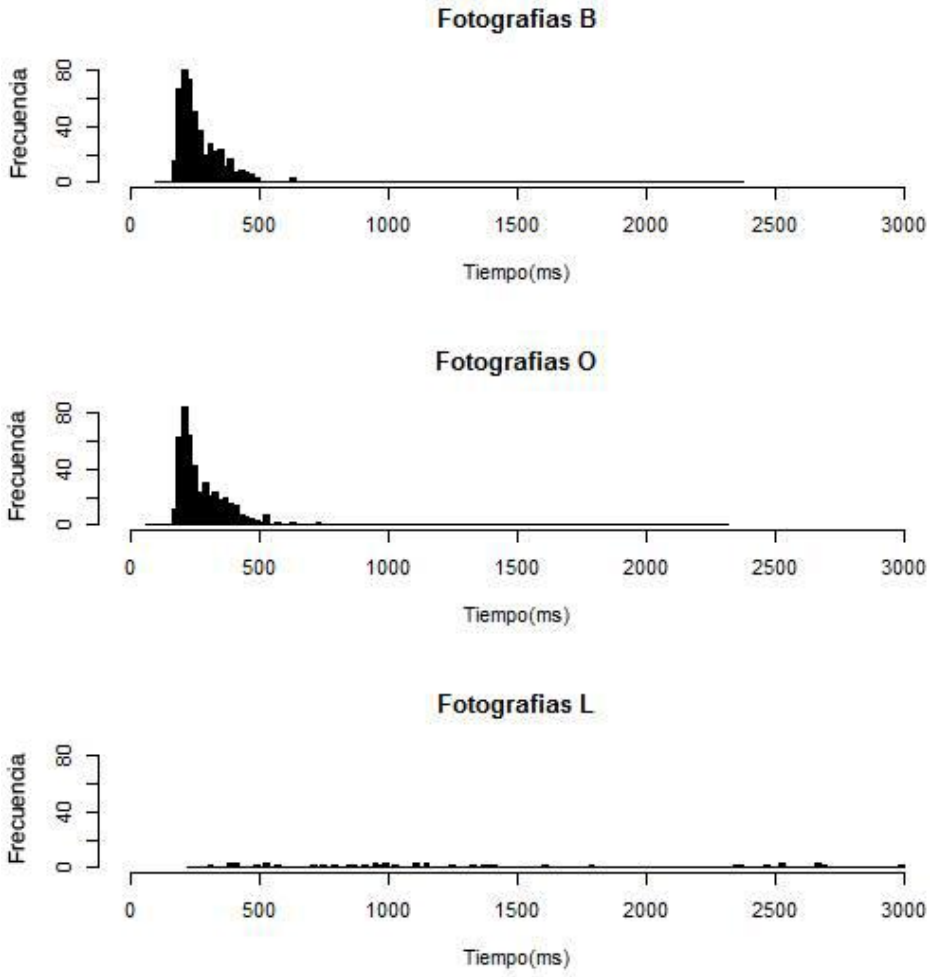


Figura 7. Histogramas que muestran la distribución de los tiempos de latencia de la primera sacada hacia la AI en las tres condiciones.

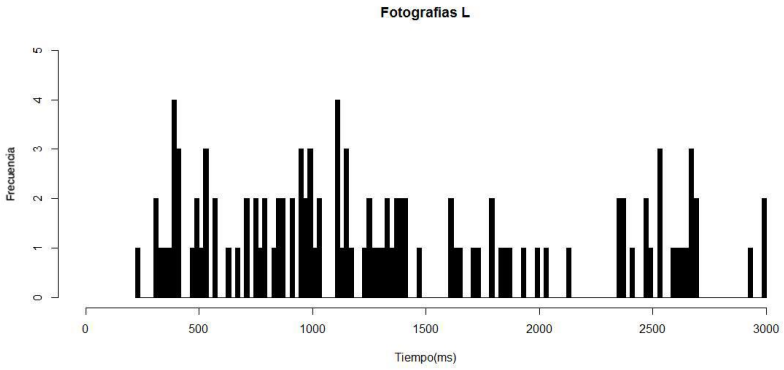


Figura 8. Histograma de la latencia de la primera sacada hacia la AI, condición L amplificada.

5.2 Prueba de normalidad

Para poder comparar si existen diferencias significativas entre las latencias de las condiciones B y O, primero se debe averiguar si la distribución de la variable en la población de donde provienen las muestras es normal, para ello, se corrió un Kolmogorov-Smirnov (debido a que el número de datos era mayor a 50), dando como resultado que se rechaza la hipótesis nula (que afirma que la distribución es normal) para todas las variables testeadas, los resultados se pueden ver en la tabla 3 a continuación.

	Tipo de Foto	Kolmogorov-Smirnov[a]		
		Estadístico	gl	Sig.
Latencia primera Sacada hacia AI (ms)	B	.251	503	.000
	O	.223	498	.000
	L	.118	112	.001
Primera Fijación dentro del AI	B	.254	503	.000
	O	.225	498	.000
	L	.117	112	.001
Cantidad total de Fijaciones en el AI	B	.147	503	.000
	O	.142	498	.000
	L	.420	112	.000
Tiempo de Permanencia en el AI (ms)	B	.054	503	.001
	O	.057	498	.001
	L	.168	112	.000
% del ensayo Permaneciendo en el AI	B	.048	503	.009
	O	.055	498	.001
	L	.179	112	.000
% total de fijaciones en el AI	B	.074	503	.000
	O	.069	498	.000
	L	.250	112	.000

a. Corrección de significación de Lilliefors

Tabla 3. Pruebas de normalidad para las variables del experimento.

5.3 Prueba U de Mann-Whitney

Debido a que no se cumplen los supuestos para trabajar con estadística paramétrica, se procedió a utilizar una prueba no paramétrica (Krzywinski y Altman, 2014). Para contrastar la hipótesis del experimento, se comparó mediante la prueba U de Mann-Whitney si las condiciones B y O, B y L, y O y L, fueron extraídas de la misma población o no. Respecto a la variable "Latencia primera sacada hacia AI", no se encontraron diferencias significativas entre los tiempos de latencia de las condiciones B y O (U= 123993; p= 0.784); aunque tanto B como O son significativamente diferentes a L: comparación B-L (U= 1903; p=0.000), comparación O-L (U=1714; p=0.000). Lo mismo ocurrió con la variable "Primera Fijación dentro del AI": comparación entre B y O (U= 126487.5; p= 0.955), entre B y L (U= 1955; p= 0.000), y entre O y L (U= 1617; p= 0.000). Se aplicó corrección de Bonferroni. Si bien no se encontraron diferencias en la velocidad con la que los sujetos dirigen su atención hacia los rostros en imágenes naturales cuando hay o no hay cuerpo, surgió la duda si existe alguna diferencia entre las condiciones B y O, por lo que se aplicó el test en las seis variables, no encontrándose, sin embargo, diferencias significativas en ningún caso (ver Anexo 2).

Discusión

Even if we knew every rule, however, we might not be able to understand why a particular move is made in the game.

— Richard Feynman

6.1 Captura atencional de los rostros, comparación dentro de la misma imagen

Como aparece en la tabla 2, existe una diferencia notoria a simple vista entre la condición L y las otras dos. Las AI que contienen un rostro son fijadas al menos 4.37 veces más que las AI que no lo contienen, lo cual se condice con la amplia literatura que señala la alta captura atencional que los rostros humanos presentan (Birmingham et al., 2009; Cerf et al., 2009; Fletcher-Watson et al., 2008; Theeuwes y van der Stigchel, 2006); aun así, resulta mucho menor que las 16.6 veces que Cerf et al. (2009) encontraron en su experimento realizado en imágenes naturales, sospechamos que dicha diferencia puede deberse al cálculo de línea basal de fijaciones que hicieron los autores para contrastarlo con las fijaciones dirigidas a rostros; ellos consideraron todas las fijaciones realizadas por cada sujeto, excepto las grabadas en el ensayo particular a computar, posteriormente computaron la fracción de fijaciones de todos los demás ensayos que caen dentro del AI del ensayo objetivo dividido por el total de fijaciones; ese número se comparó con las fijaciones que caen dentro del AI del ensayo objetivo durante ese mismo ensayo. El promedio de ese cálculo para cada ensayo de cada sujeto arrojó 16.1:1 para los rostros. En otras palabras, se estaba comparando la captura atencional de una región de X imagen, contra la posibilidad de que una fijación de ese mismo sujeto caiga por azar en esa región, superponiendo todas las fijaciones de demás imágenes que el sujeto observó. Acá, en cambio, comparamos las fijaciones de 29 sujetos en una región de X imagen cuando hay rostro, con las fijaciones de esos mismos 29 sujetos en esa misma región de la misma imagen, sólo que sin el rostro (ni la persona). Dicho en síntesis, la comparación que Cerf et al. efectuaron fue intrasujeto e inter-imágenes, y acá se realizó una comparación intersujeto e intra-imagen.

6.2 Velocidad de la captura atencional de los rostros

La latencia de la primera sacada hacia el AI se distribuye en los ensayos de la condición B principalmente (en el 68% de los casos) entre los 115 ms (valor empírico mínimo) y los 500.66 ms de proyectada la fotografía. Para la condición O, la latencia se distribuye en el 68% de los casos entre los 131.475ms y los 450.645 ms (ver DE y "mínimo" en tabla 3); ambas condiciones muestran velocidades rápidas de inicio de sacada, contrastando de manera significativamente con la condición control L; hasta el momento, es posible afirmar que los rostros son estímulos de alto-nivel, cuya captura atencional en sujetos sanos, es involuntaria; y los tiempos de inicio de sacada encontrados en este experimento son ligeramente más tardíos que los encontrados por Crouzet et al. (2010), quienes señalan que las sacadas más rápidas hacia rostros se encuentran entre los 100-110 ms; el ligero retardo puede ser explicado porque ellos usaron un paradigma de competición de imágenes artificiales, y acá se presentaron escenas naturalistas; Rousselet et al. (2003) ejecutaron un estudio donde utilizando escenas naturalistas con rostros en ellas, pedían que los sujetos presionasen un botón apenas viesen un rostro, obteniendo velocidades tempranas de reacción entre los 250-300ms; esta tarea puede ser considerada una conjunción de atención *goal-independent* con *goal-driven*, lo cual presumiblemente aceleraría la velocidad de sacada y por tanto de reacción motora en los sujetos. La pregunta que sostenemos es si esta captura, despojada de instrucciones y factores *goal-driven* explícitos, se debe a un tipo especial de saliencia física (no representada en los algoritmos tradicionales como el Itti y Koch), o si se debe a la historia de selección temprana de los sujetos; sostenemos la hipótesis de que la respuesta a esta pregunta puede inferirse en parte de la velocidad con la que los sujetos inician la primera sacada hacia los rostros. Liu et al. (2007) indican que dentro del constructo atencional "top-down", toma 150-300 ms para desplegar atención espacial; y 300-500 ms para atención hacia rasgos [*featural attention*] (Pinto et al., 2013); dadas las críticas potentes hacia la confusión conceptual que encierra lo top-down (Awh et al., 2012), es importante no dar por sentado el significado de este constructo, y revisar la metodología del experimento. Este consistía en una tarea de señalamiento Posneriano [*Posner cueing*] con estímulos sencillos (flechas que señalizaban patrones de puntos móviles), lo cual implica que es Goal-driven y de bajo nivel. Cabe preguntarse si esas velocidades son únicamente características de la atención dependiente de metas [*goal-driven*], o si también dan cuenta

de los fenómenos involuntarios (historia de selección y recompensa) que seleccionan rasgos o regiones espaciales.

Típicamente, los procesos atencionales bottom up presentan una latencia de sacada que va entre 100-120ms, y los procesos top-down entre 150-500 ms (Pinto et al., 2013), sin embargo, como lo indica la afirmación que Pinto et al. realiza enfáticamente: “Top-down attention is called endogenous because, unlike bottom-up attention (which is automatic/involuntary), it is under clear voluntary control” (2013, p. 2) esa distinción no podría aplicar a lo que en nuestro experimento se entiende como top-down (*goal independent* – historia de selección – alto nivel), es debatible que algo *goal-independent* esté *claramente* bajo control voluntario. Anteriormente otros autores señalaron que la latencia propia de sacada top-down es de 200-250 ms (Mayfrank, Kimmig, y Fischer, 1987); no obstante, las tareas utilizadas para estimar dicha latencia repiten el mismo problema que las de margen más generoso a las que llegan Pinto. et al.; las tareas donde empieza a aparecer la distinción conceptual entre historia de selección y sacadas goal-driven, en su mayoría no miden el tiempo de respuesta como una distinción relevante (Theeuwes y Van der Burg, 2011; Wang et al., 2005) o cuando lo hacen, es más bien el tiempo causado por la interferencia entre ambos efectos, donde la atención voluntaria resulta retardada en 80ms por factores de historia de selecciones (Wolfe et al., 2003). Un problema adicional en la extrapolación de estos resultados, es que las áreas visuales corticales responden de manera distinta al mismo estímulo presentado en una escena simple versus una compleja (Vinje y Gallant, 2000); y todos los experimentos que resaltan la influencia de la historia de selección fueron realizados en contextos simplificados, presentando imágenes con pocos elementos, generalmente figuras geométricas monocromáticas en fondo gris; es posible que los tiempos de reacción cuando se necesitan procesar tan pocos elementos versus las fotografías naturalistas varíen en algún grado, aumentando en ensayos que utilizan estas últimas. Es necesario también recalcar que considerando la alta variación intersujeto en las latencias de la primera sacada, siempre resulta complicado interpretar generalizando en estas investigaciones. Esto sumado a la metodología de exploración libre, que desdibuja más la línea entre cuáles respuestas de exploración son top-down y cuáles bottom up.

En los casos que la mirada caía en el AI cuando la fotografía no contenía a la persona, la latencia se distribuye (en el 68% de los casos) entre los 581.036 ms y los 2123.964 ms (ver DE en tabla 3), resulta importante destacar que la desviación estándar implica que

incluso los valores típicos más bajos son superiores a la velocidades típicas más altas con las que los sujetos iniciaban sacadas hacia el AI espontáneamente. Esto vendría a confirmar lo ampliamente explorado en la literatura, que los rostros tienen captura atencional espontanea en escenas naturalistas (Birmingham et al., 2009; Cerf et al., 2009; Fletcher-Watson et al., 2008).

Respecto a los valores de latencia inferiores a una DE de la media (Latencia: B= 0 valores < 95ms, O= 2 valores < 132ms); el menor (76ms) presenta una latencia incluso inferior a la mínima requerida para realizar una sacada guiada por la saliencia física del estímulo, sumado a que la primera fijación en ese ensayo ocurrió a los 270ms, podemos aseverar justificadamente que dicha sacada (i) no aterrizó en la IA y (ii) probablemente se debe a un error de medición del eyetracker, como un pestañeo. Cabe destacar que ningún valor de la variable "primera fijación" es inferior a una DE de la media (Primera fijación: B= 0 valores < 138ms, O= 0 valores < 178ms). Los valores de latencia de sacada inferiores a los 150 ms fueron cinco (B= 3, O= 2), lo cual señala que son extremadamente atípicos. Sin embargo, no es posible dilucidar con la información disponible si hubo estímulos con saliencia física que resultaron consistente o significativamente más salientes que los rostros en alguna de las fotografías. Una pregunta interesante de abordar en el futuro sería respecto a la competición entre estímulos con alta saliencia física según el mapa de saliencia de la imagen naturalista y rostros de humanos que se encuentren en dicha fotografía.

6.3 Secuencia ordinal de la primera fijación en el AI

Siguiendo el hilo anterior, queda abierta también la pregunta sobre si existe algún efecto del procesamiento extrafoveal mediante el cual el cuerpo optimiza (disminuye) la latencia de sacada hacia el rostro. El experimento aquí presentado apunta a que no, empero, una variable que no se exploró es la "secuencia ordinal de la primera fijación en el AI" (o sea, qué número de fijación es la primera que da en el rostro). Este dato sólo es posible obtenerlo si se restan las sacadas y fijaciones que ocurren previas a la presentación de la escena, durante la cruz de fijación. Resulta importante obtener datos de esta variable, porque ahí podría existir cierta diferencia entre las condiciones B y O, la cual dependería del procesamiento extrafoveal de características, como los cuerpos, que indiquen la presencia del estímulo relevante (rostro), y que permitan sacadas más precisas hacia

estos. Resultados que se condecirían con los reportados por Fletcher-Watson et al. (2008).

6.4 Captura atencional *goal-independent* en características de alto nivel

En primer lugar, como Henderson y Hollingworth (1999) afirman, los estímulos que consisten en características de alto nivel capturados más lento la mirada que los estímulos simples o características de bajo nivel, Borji e Itti (2013) clasifican a los rostros como factores de alto nivel. Como desarrollamos anteriormente, es posible separar la distinción endógeno/exógeno, goal-independent/goal-driven y alto nivel/ bajo nivel. En este caso, la velocidad de captura de rostros debería ser un poco más lenta que la que se podría llamar "bottom-up pura" (saliencia física *goal-independent*), pero no tan lenta como la velocidad top-down implicada en tareas de atención hacia rasgos; aun cuando los rostros son definitivamente rasgos y no regiones espaciales. En esa línea, sería interesante comparar la IdR entre las condiciones B y O en experimentos futuros, para ver si existen diferencias atribuibles al cuerpo como elemento contextual; ya que la IdR hasta el momento se estudia en paradigmas simples de *priming* (e.g. SOA) y señalamiento posneriano; de haber un efecto, la latencia de retorno debería disminuir ante la presencia de características contextuales (i.e. cuerpo) que señalicen los rostros.

6.5 Permanencia de la mirada en rostros

Una característica ampliamente reportada de la captura por saliencia física de la mirada, es su transitoriedad, esto es, que se gatilla y decae rápidamente, lo cual contrasta con la capacidad de los sujetos de sostener la mirada indefinidamente cuando esta es voluntaria o impuesta por una tarea que se le indique a los sujetos (Carrasco, 2011). En nuestro experimento, tanto la condición B como O presentaron tiempos elevados de atención a los rostros (ver Tabla 2 y Figura 9), que superaron ampliamente los 200-400 ms promedio que Land (2011) le asigna a las fijaciones en escenas naturalistas, lo que se condice con lo encontrado por una investigación anterior, que señala que los sujetos permanecen un 40% en promedio del ensayo en el cuerpo y/o el rostro de los humanos que aparecen en las escenas naturalistas (Fletcher-Watson et al., 2008) y en investigaciones respecto a retención de la atención en rostros en contextos simplificados (Bindemann et al., 2005).

Es posible observar también, que los rostros le ganan ampliamente a otros estímulos salientes (tanto semánticos como por características físicas) en el tiempo de captura atencional (para un ejemplo, ver Figura 11) Sería interesante estudiar si hay diferencias entre las condiciones B y O en cuanto a la conducta ocular orientada al rostro (transitoria vs sostenida). Más interesante aún sería hacer competir en una escena naturalista bajo la premisa de exploración libre, rostros con otros estímulos que presentan típicamente saliencia física, como luces rojas o luces que parpadeen (en caso de videos), y a su vez, comparar transitoriedad (tiempo que se permanece en el estímulo que captura la mirada) entre los rostros y los estímulos físicamente salientes.

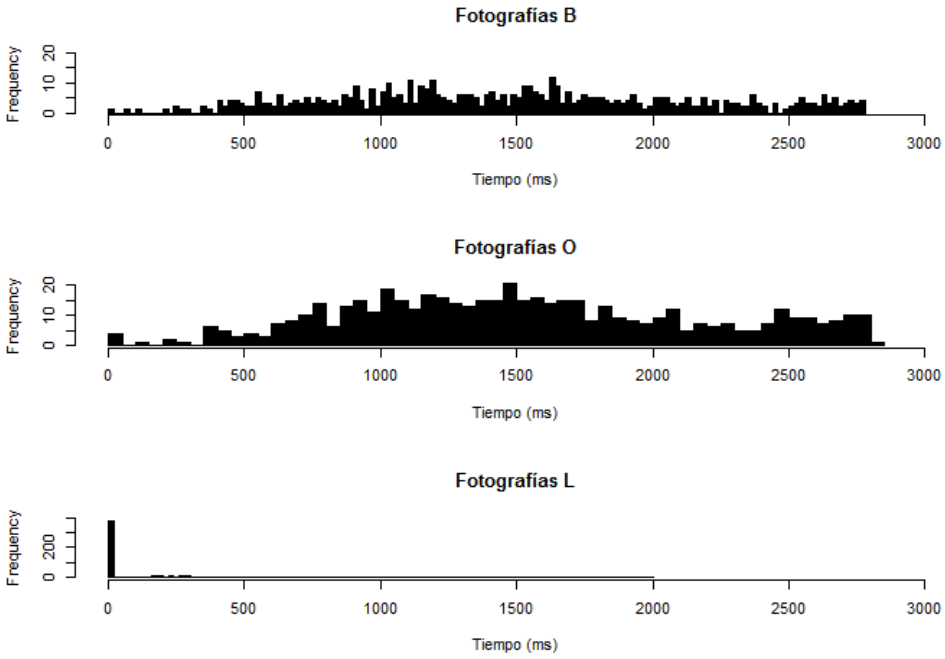


Figura 9. Histogramas que muestran la distribución de los tiempos de permanencia de la mirada [*dwell time*] dentro del AI en las tres condiciones. La condición L, al presentar demasiados datos iguales a cero, se encuentra en otra escala.

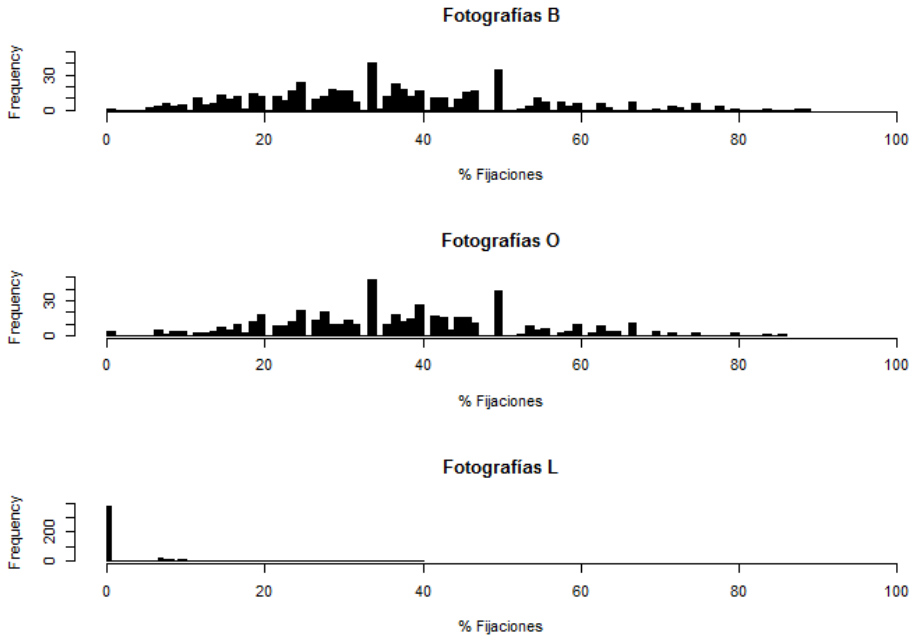


Figura 10. Histogramas que muestran la distribución de los porcentajes de fijaciones dentro del AI (nº de fijaciones dentro del AI / total de fijaciones en los 5s aprox de ensayo, incluyendo cruz de fijación) en las tres condiciones. La condición L, al presentar demasiados datos iguales a cero, se encuentra en otra escala.

Figura 11. Mapa de calor que representa la duración de las fijaciones del ensayo 51 del sujeto NO240815. El tiempo de permanencia en áreas de la escena está representado por las manchas, donde el color verde señala fijaciones cercanas a los 0s y el rojo fijaciones cercanas a la fijación más larga del ensayo (965.22ms en este caso). El mapa fue generado mediante el programa EyeLink® Data Viewer.

6.6 Implausibilidad semántica como posible efecto perturbador

El aprendizaje previo (de objetos, contextos y relaciones esperables entre estos) que presentan los sujetos, modifica sus patrones exploratorios oculares, aun en estudios donde no se dan instrucciones explícitas (Torralba et al., 2006; Tatler, 2014). Se le llama Objetos Implausibles o Inconsistentes (OI) a aquellos objetos que tienden a entrar en conflicto con el *gist* de la escena; o bien, que dado el contexto de la imagen, los sujetos no esperarían ni les parece coherente que estén ahí, como un pez encima de una nube o un rostro de bebé en el sol. Estos tienden a llamar más temprano y por más tiempo la atención que los objetos plausibles (*i.e.* menor latencia de la primera fijación hacia el objeto), son recordados más fácilmente, y además, la mayor amplitud de sacadas registradas hacia los OI puede indicar un procesamiento extra-foveal de las inconsistencias en la periferia visual. Otros estudios sin embargo, no encontraron una preferencia temprana en los sujetos hacia fijar la mirada en OI's, en cambio, estos son fijados por más tiempo y es más probable que sean fijados de nuevo (Henderson, Weeks y Hollingwood, 1999). Esta inconsistencia entre resultados a la hora de replicar el experimento, se ha explicado como efecto del tamaño y la complejidad de las imágenes o del hacinamiento visual; no obstante, resulta importante consignar que en muchos estudios se confunden las características semánticas (top-down) y las visuales (i.e. de bajo nivel o bottom-up, en este caso coincidiendo). Entonces, tomando como ejemplo el estudio emblemático de Loftus y Mackworth (1978), cuando un tractor es reemplazado por un pulpo en una fotografía de una granja, no sólo cambia el contexto semántico, sino que también las características de contraste, forma, color y luminancia del objeto cambian, además de su integración con las características de su entorno (esto sería un efecto de alto nivel, pero involuntario); partiendo por que la forma de un tractor es rectangular y la de un pulpo es relativamente redonda, se esperaría que los efectos de saliencia bottom-up fuesen distintos (Bonitz y Gordon, 2008) Esto no ocurre en el experimento de Henderson et al., y suponemos que tampoco en el nuestro (las condiciones B y O son idénticas en todas las variables analizadas). Es posible sin embargo, que un efecto de

implausibilidad semántica en la condición O esté enmascarando en algún sentido una diferencia de velocidad en el inicio de la primera sacada hacia el rostro (en el caso de que B provoque sacadas ligeramente más tempranas; pero que los OI en la condición O también disminuyan el tiempo de latencia de sacada, haciendo ver a ambas condiciones iguales). Para descartar esa opción, se tendría que validar el set de imágenes; por ejemplo, realizar una encuesta con un nº significativo, donde se mostrase todas las fotografías O a personas que no tienen conocimiento del experimento y se les pidiese señalar del modo menos inducidor-de-sesgos posible, ¿Cuáles fotografías les parecen menos verosímiles/naturales/inalteradas? luego, correr una prueba U de Mann-Whitney entre las subcondiciones O-plausible versus O-implausible, en caso de que alguna de las fotografías utilizadas fuese considerada poco verosímil para observar si existe diferencia entre ambas subcondiciones. Sin embargo, nos parece que esta posibilidad es poco robusta, (i) porque las fotografías fueron editadas en Photoshop explícitamente para que parezcan semánticamente plausibles, (ii) porque B y O presentan distribuciones demasiado similares, y O no presenta anomalías en su distribución que puedan verse reflejadas en su asimetría o curtosis (ver figura 7) y (iii) porque no todos los estudios obtienen los mismos resultados cuando buscan medir efecto en la latencia de movimiento sacádico, lo cual no garantiza de que incluso si es el caso de que en las fotografías utilizadas haya algunas semánticamente implausibles, estas vayan a afectar significativamente la latencia de la primera sacada hacia el rostro.

6.7 Recapitulando ¿por qué no se observaron diferencias entre las condiciones B y O?

Las posibles razones son varias, y por eso mismo, es importante que puedan realizarse experimentos más finos a futuro para tener más claridad respecto al fenómeno de la captura de la mirada por rostros en sujetos sanos. (1) En primer lugar, cabe la posibilidad de que el valor del rostro en el procesamiento extrafoveal sea muy alto como para que el cuerpo le agregue significativamente valor extra; (2) en segundo lugar, es posible que el cuerpo de todos modos no represente un estímulo significativo en ningún caso, esto podría testearse comparando cuerpos sin rostro, semánticamente plausibles, en escenas naturalistas, donde los rostros se vean casualmente ocultados por un objeto, y se mida la velocidad de sacada hacia los cuerpos (y dentro de estos, qué parte recibe sacadas más

tempranas), y la secuencia ordinal de la primera fijación que caiga en el área de los cuerpos; (3) es posible que el cuerpo sí tenga la capacidad de disminuir la latencia hacia los rostros o la secuencia ordinal de la primera fijación, pero que esto se haya visto enmascarado por algún tipo de efecto indeseado, esta opción nos parece poco factible por razones arriba expuestas con detalle, no obstante, las posibles variables confundidoras que no fueron exhaustivamente testeadas en este experimento, son las siguientes: (i) Implausibilidad semántica de algunas fotografías O, (ii) que el efecto de reconocimiento de rostros disminuya significativamente más la latencia de sacada que el efecto contextual del cuerpo en sacadas ordinarias hacia rostros desconocidos; esto sería relevante ya que aproximadamente la mitad de los voluntarios del estudio conocía a la mitad o más de los sujetos que aparecen en las fotografías, no obstante, Agnati, Guidolin, Cortelli, Genedani, Cela-Conde y Fuxe (2012) estiman que el tiempo mínimo para que un percepto pueda ser consciente, es de 200 ms; además, los tiempos de reacción mínima para reconocer rostros –en el caso del estudio, rostros de famosos–, es de 360-390 ms (Barragan-Jason et al., 2013); habiendo obtenido nosotros varios resultados menores a esas cifras. Sin embargo, no se descarta un efecto tipo *priming* o no-consciente, por lo que sería relevante estudiar experimentalmente si los rostros conocidos disminuyen significativamente el tiempo de reacción hacia estos, en escenas naturalistas.

Conclusiones

Pero hace mucho tiempo que he aprendido a ponerme en guardia cuando alguien cita a Pascal. Es una cautela de higiene elemental.

— José Ortega y Gasset

De la revisión de la literatura actual acerca del tema y el experimento realizado, es posible extraer las siguientes conclusiones: (i) los rostros son estímulos complejos, compuestos por características de bajo y alto nivel, que presentan una captura atencional potente e involuntaria, (ii) dicha captura atencional no depende de la saliencia bruta de sus componentes, y se presenta a los 6 meses de edad, aunque ya hay captura atencional a rostros (aunque no tan fuerte ni sostenida) a las 6 semanas de edad (Cashon y Cohen, 2003; Di Giorgio et al., 2012), sería importante que futuras investigaciones puedan precisar los umbrales donde se va desarrollando cada característica de la captura atencional de rostros en adultos (que Golarai et al. (2007) estiman que se alcanza a los 16 años en promedio), y bajo qué condiciones este desarrollo se inhibe o modifica, para citar un par de ejemplos entre los más notorios: en prosopagnosia (Lê et al., 2003; Mannan et al., 2009) y en esquizofrenia (Benson et al., 2012; Chen et al., 2008; Chen y Ekstrom, 2016; Egaña et al., 2013). (iii) Las latencias menores para iniciar un movimiento sacádico hacia un rostro, reportadas por investigadores, son de 100-110 ms (Crouzet et al., 2010); en el presente experimento, excluyendo datos atípicos, estuvieron en el rango de 115-135 ms, (iv) no se encontraron diferencias en las latencias, que pudiesen ser atribuidas a la presencia de los cuerpos en las escenas observadas, (v) la captura atencional de los rostros no puede ser definida como transitoria, y además, presenta inhibición del retorno [IdR] (Jingling et al., 2015; Theeuwes y Van der Stitchel, 2006, p.606; Silvert y Funes, 2016), no obstante, es necesario que futuras investigaciones testeen este fenómeno en escenas naturalistas y con protocolos lo más parecidos al comportamiento usual de los ojos en el mundo real, ya que la mayor parte de la bibliografía en torno al fenómeno utiliza experimentos poco naturales con contextos empobrecidos (que contienen pocos objetos y dispuestos de maneras no naturales).

Dado lo anterior, es posible hipotetizar que la captura atencional de los rostros es descomponible en dos fenómenos con tiempos distintos, el primero (100-150 ms aprox.) sería principalmente atención implícita [*covert*] hacia las características de bajo nivel de los rostros en el procesamiento extrafoveal de la escena y calzaría con el ERP occipital (P1); luego, el segundo (250-300 ms aprox.) sería el movimiento sacádico que dirige la mirada hacia el rostro (atención explícita [*overt*]), guiado por sus características de alto nivel o semánticas, que calzaría con el ERP occípito-temporal (N170), el cual responde a la configuración holística de los rostros (Joseph, DiBartolo y Bhatt, 2015; Rossion y Caharel, 2011). Es interesante que ambos potenciales están presentes en sujetos diagnosticados con esquizofrenia (Wynn et al., 2008), para quienes los rostros no presentan una captura atencional demasiado alta. Sin embargo, los estudios realizados a la fecha imponen tareas y utilizan escenas poco ecológicas (e.g. un rostro flotando en fondo gris o blanco); por lo que sería importante investigar el fenómeno en un protocolo de exploración libre y escenas naturalistas.

Finalmente, dentro de algunas formas de realizar la clasificación top-down / bottom-up, la captura atencional de rostros podría ser etiquetada como top-down, y dentro de otras como bottom-up. En la clasificación que se centra en la distinción voluntario/involuntario (Al-Aidroos et al., 2012; Carrasco, 2011; Pinto et al., 2013), la captura a rostros es bottom-up, y surge como problema explicar su baja saliencia física (Birmingham et al., 2008; Cerf et al, 2009), claro, esto puede evitarse señalando que son top-down e invocando la existencia de metas o prioridades no-conscientes, pero esta interpretación sería poco parsimoniosa y ambigua. Ya que la saliencia física es la característica definitoria de lo bottom up (Borji et al., 2013b; Pinto et al., 2013), los rostros tendrían que ser un estímulo top down, sin embargo, queda como problema explicar que presenten IdR, que incluso cuando están borrosos tengan captura atencional. En vista de lo cual, es necesario entender las particularidades de los rostros como estímulos: baja saliencia física, captura que no depende por completo de sus características de alto nivel, independencia de la voluntad o las metas del sujeto, dependencia de la historia de selecciones (socialización) y quizá de la historia de recompensas –cariño, protección, sexualidad, los cuales podrían causar por condicionamiento apetitivo el hecho de que algunas caras activen áreas relacionadas al placer en el cerebro (Kampe et al., 2001; O'Doherty et al., 2003), y puedan funcionar como estímulo reforzador en un contexto complejo–. Los rostros cumplen con sólo una (la tercera) de las tres características (dependencia de: metas actuales,

expectativas y conocimiento) que Corbetta y Shulman (2002) le imputan a lo top-down; por ello rescatamos la clarificación conceptual que proponen Awh et al. (2012), donde ya no bastaría con llamarle fenómeno “top-down”, sino, se convierte en imperativo ver cuales elementos semánticos, emocionales y contextuales; sincrónicos y diacrónicos, de la historia de selecciones del sujeto operan a la base de la captura atencional. Claro está, estas hipótesis finales que aventuramos requieren ser contrastadas empíricamente por futuras investigaciones.

Bibliografía

Agnati, L., Guidolin, D., Cortelli, P., Genedani, S., Cela-Conde, C. & Fuxe, K. (2012). Neuronal correlates to consciousness. The "Hall of Mirrors" metaphor describing consciousness as an epiphenomenon of multiple dynamic mosaics of cortical functional modules. *Brain Research*, 1476, 3-21.

Andrews, T., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase–face illusion. *NeuroImage*, 17, 890–901

Awh, E., Belopolsky, A. & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Science*, 16(8), 437–43.

Barragan-Jason, G., Besson, G., Ceccaldi, M., & Barbeau, E. (2013). Fast and Famous: Looking for the fastest speed at which a face can be recognized. *Frontiers in Psychology*, 4, 100.

Benson, P., Beedie, S., Shepard, E., Giegling, I., Rujescu, D. & St. Clair, D. (2012). Simple viewing tests can detect eye movement abnormalities that distinguish schizophrenia cases from controls with exceptional accuracy. *Biological Psychiatry*, 72, 716-724.

Bindemann, M., Burton, M., Hooge, I., Jenkins, R., & de Haan, E. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, 12, 1048–1053.

Birmingham, E., Bischof, W., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49(24), 2992–3000.

Bonitz, V. & Gordon, R. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, 129, 255-263.

Borji, A. & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 185-207.

Borji, A., Sihite, D., & Itti, L. (2013a). Objects do not predict fixations better than early saliency: A re-analysis of Einhäuser et al.'s data. *Journal of Vision*, 13(10):18, 1–4.

Borji, A., Sihite, D., & Itti, L. (2013b) What stands out in a scene? A study of human explicit saliency. *Vision Research,* 91, 62–77.

Buswell, G. (1935). *How People Look at Pictures*. Chicago, IL: University of Chicago Press.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46, 4333–4345.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525.

Cashon, C., & Cohen, L. (2003). The construction, deconstruction, and reconstruction of infant face perception. En O. Pascalis y A. Slater (Eds.) *The development of face processing in infancy and early childhood: Current perspectives* (pp. 55–68). New York: NOVA Science Publishers.

Cerf, M., Frady, E. & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, 9(12):10, 1–15.

Chen, Y., Norton, D., Ongur, D. & Heckers, S. (2008). Inefficient face detection in schizophrenia. *Schizophrenia Bulletin*, 34(2), 367-374.

Chen, Y. & Ekstrom, T. (2016). Perception of faces in schizophrenia: Subjective (self-report) vs. objective (psychophysics) assessments. *Journal of Psychiatric Research*, 76, 136-142.

Chica, A., Bartolomeo, P. & Lupiáñez, J., (2013). Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural Brain Research,* 237, 107-123.

Chica, A., Botta, F., Lupiáñez, J. & Bartolomeo, P. (2012). Spatial attention and conscious perception: interactions and dissociations between and within endogenous and exogenous processes. *Neuropsychologia*, 50(5), 621–9.

Cook, Carvalho & Damasio, (2014). From membrane excitability to metazoan psychology. *Trends in Neuroscience*, 1085, 1-8.

Corbetta, M. & Shulman, G. (2002). Control of goal directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215.

Coull, J. T., Frith, C. D., Buchel, C., & Nobre, A. C. (2000). Orienting attention in time: Behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*, 38(6), 808–819.

Crouzet, S., Kirchner, H., & Thorpe, S. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4), 1–17.

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective visual attention.

Annual Reviews of Neuroscience, 18, 193-222.

Di Giorgio, E. Turati, C., Altoè, G. & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3- and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, 113, 66-77.

Egaña, J., Devia, C., Mayol, R., Parrini, J., Orellana, G., Ruiz, A. & Maldonado, P. (2013). Small saccades and image complexity during free viewing of natural images in schizophrenia. *Frontiers in Psychiatry*, 4(37), 1-13.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, 8(14):18, 1–26.

Emery, N., (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24, 581–604.

Esber, G. R. & Haselgrove, M. (2011) Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proceedings of the Royal Society B.*, 278, 2553-2561.

Fletcher-Watson, S., Findlay, J., Leekam, S., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37(4), 571-583.

Foulsham, T.; Barton, J.; Kingstone, A.; Dewhurst, R. & Underwood, G. (2011). Modeling eye movements in visual agnosia with a saliency map approach: bottom-up guidance or top-down stratregy? *Neural Networks*, 24, 665-677.

Gallant, J., Connor, C. & Essen, D. (1998). Neural activity in areas Vl, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, 9, 85–90.

Golarai, G.,Ghahremani, D.,Whitfield-Gabrieli, S.,Reiss,A., Eberhardt, J., Gabrieli, J., et al. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, 10, 512–522.

Hahn, B., Ross, T. & Stein, E. (2006). Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. *NeuroImage,* 32, 842-3.

Henderickx, D., Maetens, K. & Soetens, E. (2012). The involvement of bottom-up saliency processing in endogenous inhibition of return*. Attention Perception, & Psychophysics*, 74, 285–299.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504.

Henderson, J. & Hollingworth, A. (1999). High-level scene perception. *Annual Review of*

Psychology, 50, 243–271.

Henderson, J., Weeks, P., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210–228.

Hubel, D. & Wiesel, T. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, 26, 994–1002.

Hwang, A., Wang, H. & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51, 1192-1205.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts ç of visual attention. *Vision Research*, 40, 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254–1259.

Jingling, L., Lin, H.-F., Tsai, C.-J. & Lin, C.-C. (2015). Development of Inhibition of Return for eye gaze in adolescents. *Journal of Experimental Child Psychology*, 137, 76-84.

Joseph, J., DiBartolo, M. & Bhatt, R. (2015). Developmental changes in analytic and holistic processes in face perception. *Frontiers in Psychology*, 6:1165, 1-16.

Kayser, C., Nielsen, K.. & Logothetis, N. (2006). Fixations in natural scenes: an interaction of image saliency and content. *Vision Research*, 46(16), 2535-2545.

Kanwisher, N., McDermott, J., y Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–11.

Kampe, K., Frith, C., Dolan, R. J. & Frith, U. (2001). Reward value of attractiveness and gaze. *Nature*, 413, 589.

Kastner, S. & Ungerleider, L. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and top-down attention: different processes and overlapping neural systems. *Neuroscientist*, 20(5), 509-21.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.

Kollmorgen, S., Nortmann, N., Schröder, S. & König, P. (2010). Influence of Low-Level Stimulus Features, Task Dependent Factors, and Spatial Biases on Overt Visual Attention. *PLoS Computational Biology*, 6(5), e1000791.

Kornmeier, J., Hein, C., & Bach, M. (2009). Multistable perception: When bottom-up and top-down coincide*. Brain and Cognition*, 69(1), 138-147.

Krzywinski, M. & Altman, N. (2014). Points of significance: Nonparametric tests. *Nature methods*, 11, 467-8.

Kubota, J., & Ito, T. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology*, 43, 738–752.

Land, M. (2011). Oculomotor behaviour in vertebrates and invertebrates. En S. Liversedge, I. Gilchrist y S. Everling (Eds.) *The Oxford handbook of eye movements* (pp. 3-15)*.* New York: Oxford University Press.

Land, M. & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41, 3559–3565.

Land, M. & Nilsson, D. (2012). *Animal eyes* (2nd ed). New York: Oxford university press.

Land, M. & Tatler, B. (2009). *Looking and acting: eye movements in everyday life*. Oxford: Oxford University Press

Lê, S., Raufaste, E., Roussel, S., Puel, M., & Demonet, J. (2003). Implicit face perception in a patient with visual agnosia? Evidence from behavioural and eye-tracking analyses. *Neuropsychologia*, 41 , 702–712.

LeMeur, O. & Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, 116, 152-164.

Leonards, U. & Scott-Samuels, N. (2005). Idiosyncratic initiation of saccadic face exploration in humans. *Vision Research*, 45, 2677–2684.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6, 9–16.

Liu, J.; Li, J., Feng, L., Li, L., Tian, J. & Lee, K. (2014). Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, 53, 60–77.

Liu, T., Stevens, S., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, 47(1), 108–113.

Loftus, G. & Mackworth, N. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565-572.

Mack, A., Pappas, Z., Silverman, M. & Gay, R. (2002). What we see: inattention and the capture of attention by meaning. *Consciousness and Cognition*,11(4):488-506.

Maldonado, P. (2008). Anatomía funcional de la percepción visual. En A. Slachevsky, F. Manes, E. Labos & P. Fuentes (Eds.), *Tratado de Neurología Cognitiva y Neuropsicología*. Buenos Aires: Editorial Akadia.

Mannan, S., Kennard, C. & Husain, M. (2009). The role of visual salience in directing eye movements in visual object agnosia. *Current Biology*, 19(6), 247-248.

Martinez, F. & Pons, A. (2004). *Fundamentos de visión binocular*. Valencia: Publicacions de la Universitat de València.

Mayfrank, L., Kimmig, H., & Fischer, B. (1987). The role of attention in the preparation of visually guided saccadic eye movements in man. En J. K. O'Reganm & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition,* (pp. 37–45). NY: North-Holland.

Mescher, M. & De Moraes, C. (2014). The role of plant sensory perception in plant–animal interactions. *Journal of Experimental Botany*, 66(2), 425-433.

McDowell, J., Dyckman, K., Austin, B. & Clementz, B. (2008). Neurophysiology and neuroanatomy of reflexive and volitional saccades: evidence from studies of humans. *Brain Cognition.* 68, 255–270.

Miyamoto, Y., Nisbett, R., & Masuda, T. (2006). Culture and physical environment: Holistic versus analytic perceptual affordance. *Psychological Science*, 17, 113-119.

Miyazaki, S. & Iwasaki, S. (2010). Do happy faces capture attention? The happiness superiority effect in attentional blink. *Emotion*, 10(5), 712-716.

Moreno, A. & Lasa, A. (2003). From Basic Adaptivity to Early Mind. *Evolution and Cognition*, 9, 12-30.

Morris, J., Green, S., Marion, I., & McCarthy, G. (2008). Guided saccades modulate face- and body-sensitive activation in the occipitotemporal cortex during social perception. *Brain and Cognition*, 3, 16-25.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45, 205-231.

Nemrodov, D., Anderson, T., Preston, F. & Itier, R. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97:81-94.

Noë, A. (2004). *Action in perception*. Cambridge, Massachussets: MIT Press.

Nyström, M. & Holmqvist, K. (2008). Semantic override of Low-level features in image viewing – both initially and overall. *Journal of Eye Movement Research*, 2, 1-11.

O'Doherty, J., Winston , J., Critchley, H., Perrett, D., Burt, D. & Dolan, J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia,* 41, 147-155.

Pastukhov, A., Vonau, V., Stonkute, S. & Braun, J. (2012).Spatial and temporal attention revealed by microsaccades. *Vision Research*, 85, 45-67.

Peelen, M., Heslenfeld, D. J., & Theeuwes, J. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *NeuroImage*, 22, 822–830.

Peil, K. (2014). Emotion: the self-regulatory sense. *Global Advances in Health and Medicine*, 3(2): 80–108.

Pinto, Y., van der Leij, A., Sligte, I., Lamme, V., & Scholte, H. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, 13(3):16, 1–14.

Pires, A., Vásquez, A., Carboni, A., & Maiche, A. (2014). Percepción visual. En D. Redolar (Ed.), *Neurociencia Cognitiva*. Madrid: Editorial Panamericana.

Pomerantz, J. (2003). Perception: Overview. En L. Nadel (Ed.) *Encyclopedia of Cognitive Science. Vol. 3.* (pp.527–537) London: Nature Publishing Group.

Posner, M. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.

Posner, M. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3-25.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A., McNamara, J. & Williams, S. (2008*). Neurociencia* (3ª. Ed.). Buenos Aires: Editorial médica panamericana.

Quine, W. (1975). On empirically equivalent systems of the world. *Erkenntnis*, 9, 313–28.

Rolls, E. (2008). Top–down control of visual perception: Attention in natural vision. *Perception*, 37, 333-354.

Ross, C. & Kirk, E. (2007). Evolution of eye size and shape in primates. *Journal of Human Evolution* 52(3): 294-313.

Rossion, B. & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, 51, 1297-1311.

Rousselet, G., Mace, M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3, 440–455.

Rouw, R., Kosslyn, S. & Hamel, R. (1997). Detecting high-level and low-level properties in visual images and visual percepts. *Cognition*, 63, 209-226.

Sarter, M., Givens, B. & Bruno, J. (2001).The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews*, 35(2), 146-160.

Schneider, K. & Kastner, S. (2009). Effects of sustained spatial attention in the human lateral geniculate nucleus and superior colliculus. *Journal of Neuroscience*, 29(6), 1784–1795.

Snowden, R. Thompson, P. & Trosianko, T. (2012). *Basic vision: An introduction to visual perception*. United Kingdom: Oxford University Press.

Stein, B., Stanford, T. & Rowland, B. (2009). The neural basis of multisensory integration in the midbrain: Its organization and maturation. *Hearing Research*, 258(1-2):4-15.

Sumner, P. (2011). Determinants of saccade latency. En S. Liversedge, I. Gilchrist y S. Everling (Eds.) *The Oxford handbook of eye movements* (pp. 413-24). New York: Oxford University Press.

Tallon-Baudry, C. (2012). On the neural mechanisms subserving consciousness and attention. *Frontiers in Consciousness Research*, 2, 397.

Tatler, B. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17.

Tatler, B. (2014). Eye Movements from Laboratory to Life. En M. Horsley, M. Elliot, B. Allen Knight y R. Reilly (Eds.) *Current trends in eye tracking research* (pp. 17-35). Suiza: Springer.

Tatler, B., Hayhoe, M., Land, M. & Ballard, D. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, 11(5):5, 1-23.

Theeuwes, J. & Van der Burg, E. (2011) On the limits of top-down control of visual selection. *Attention, Perception, & Psychophysics*. 73, 2092–2103.

Theeuwes, J. & Van der Stigchel, S. (2006). Faces capture attention: evidence from inhibition of return. *Visual Cognition*, 13, 657–665.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113, 766–786.

Tomonaga, M. (2010). Chimpanzee eyes have it? Social cognition on the basis of gaze

and attention from the comparative–cognitive–developmental perspective. En E. Lornsdorf, S. Ross, & T. Matsuzawa (Eds.), *The mind of the chimpanzee: Ecological and empirical perspectives* (pp. 42–53). Chicago, Illinois: University of Chicago Press.

Treisman, A. y Gelade, G. (1980). A feature-integration theory of attention*. Cognitive Psychology*,12, 97–136.

Tsao, D. y Livingstone, M. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, 31, 411–437.

Tseng, P., Carmi, R., Cameron, I., Munoz, D. & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of vision*, 9(7), 4.

Uttal, W. (1981). *A taxonomy of visual processes.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Valberg, A. (2005). *Light, vision, color.* West Sussex: Wiley and sons.

van Swinderen, B. (2011). Attention in Drosophila. *International Review of Neurobiology*, 99, 51-85.

Wang, D., Kristjansson, A. & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception & Psychophysics*, 67, 239–253.

Wei, L. y Luo, D. (2015). A biologically inspired computational approach to model top-down and bottom-up visual attention. *Optik*, 126, 522–529.

Wieser, M., Gerdes, A., Büngel, I., Schwarz, K., Mühlberger, A., & Pauli, P. (2014). Not so harmless anymore: How context impacts the perception and electrocortical processing of neutral faces. *NeuroImage*, 92, 74-82.

Wolfe, J., Butcher, S., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology*, 29(2), 483–502.

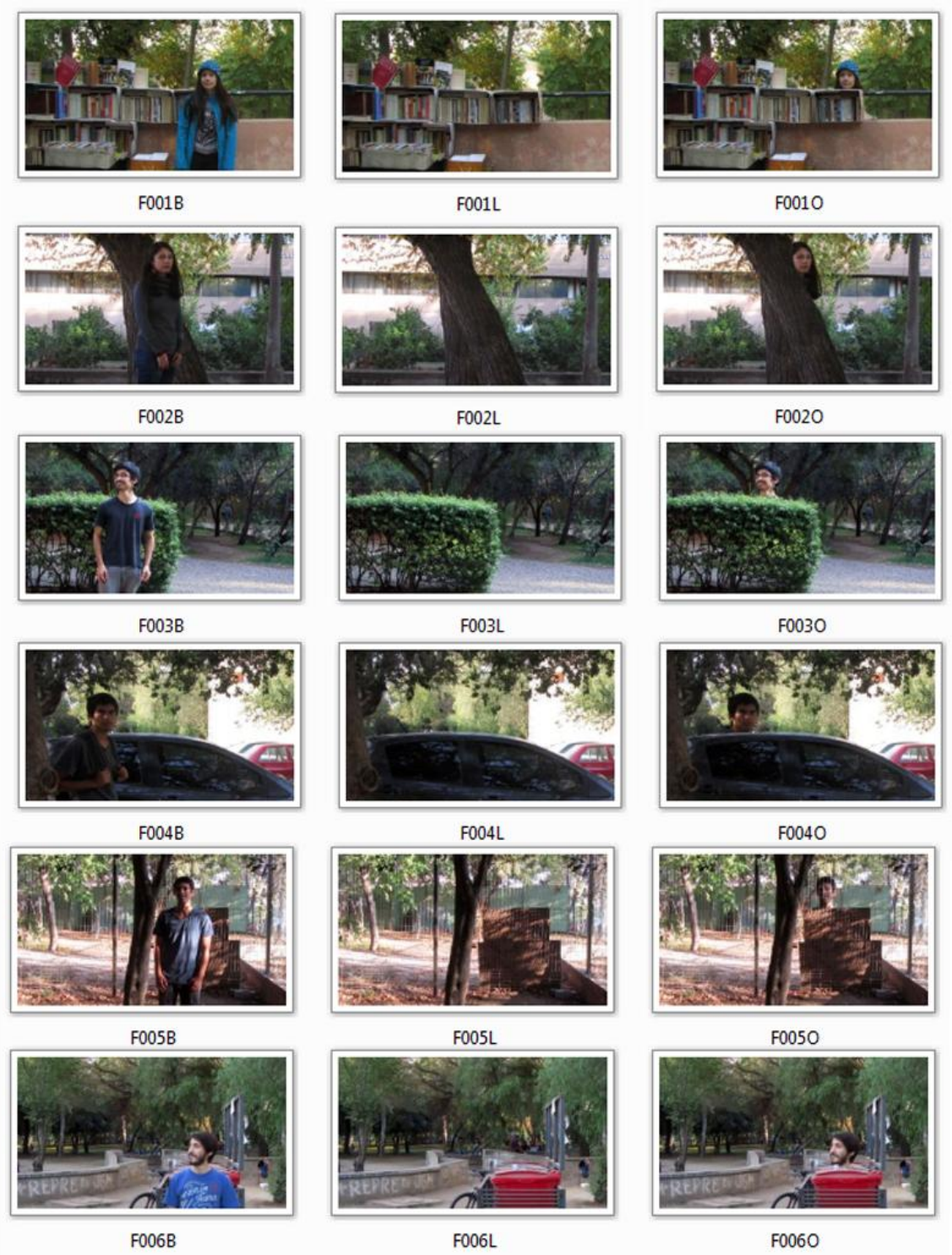
Wynn J., Lee J., Horan W. & Green M. (2008). Using event related potentials to explore stages of facial affect recognition deficits in schizophrenia. *Schizophrenia Bulletin*, 34, 679–687.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Yoshimura, K. (2011). Stimulus Perception and membrane excitation in unicellular alga Chlamydomonas. *Plant and Cell Biology*, 10(2), 79-91.

Zmigrod, S. y Hommel, B. (2013). Feature integration across multimodal perception and action: a review. *Multisensory Research*, 26, 143-57.

Anexo 1: batería de fotografías utilizadas



F007B

F007L

F007O

F008B

F008L

F008O

F009B

F009L

F009O

F010B

F010L

F010O

F011B

F011L

F011O

F012B

F012L

F012O

F013B

F013L

F013O

F014B

F014L

F014O

F015B

F015L

F015O

F016B

F016L

F016O

F017B

F017L

F017O

F018B

F018L

F018O

Anexo 2: Prueba U de Mann-Whitney para todas las variables

Comparación condiciones B y O.

	Latencia primera Sacada hacia AI	Primera Fijación dentro del AI	Cantidad total de Fijaciones en el AI	Tiempo de Permanencia en el AI	% del ensayo Permaneciendo en el AI	% total de Fijaciones en el AI
Mann-Whitney U	123993.000	126487.500	128699.500	130400.000	131046.500	127047.000
Wilcoxon W	250749.000	255265.500	265202.500	266903.000	267549.500	263550.000
Z	-.274	-.057	-1.573	-1.199	-1.067	-1.889
Asymp. Sig. (2-tailed)	.784	.955	.116	.230	.286	.059

Comparación condiciones B y L.

	Latencia primera Sacada hacia AI	Primera Fijación dentro del AI	Cantidad total de Fijaciones en el AI	Tiempo de Permanencia en el AI	% del ensayo Permaneciendo en el AI	% total de Fijaciones en el AI
Mann-Whitney U	1903.000	1955.000	6914.500	3800.500	3607.500	5572.000
Wilcoxon W	128659.000	130733.000	143417.500	140303.500	140110.500	142075.000
Z	-15.445	-15.435	-27.319	-27.880	-27.921	-27.510
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

Comparación condiciones O y L.

	Latencia primera Sacada hacia AI	Primera Fijación dentro del AI	Cantidad total de Fijaciones en el AI	Tiempo de Permanencia en el AI	% del ensayo Permaneciendo en el AI	% total de Fijaciones en el AI
Mann-Whitney U	1714.000	1617.000	6741.000	4027.500	3965.500	5247.000
Wilcoxon W	125965.000	126867.000	143244.000	140530.500	140468.500	141750.000
Z	-15.532	-15.599	-27.370	-27.849	-27.862	-27.596
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000