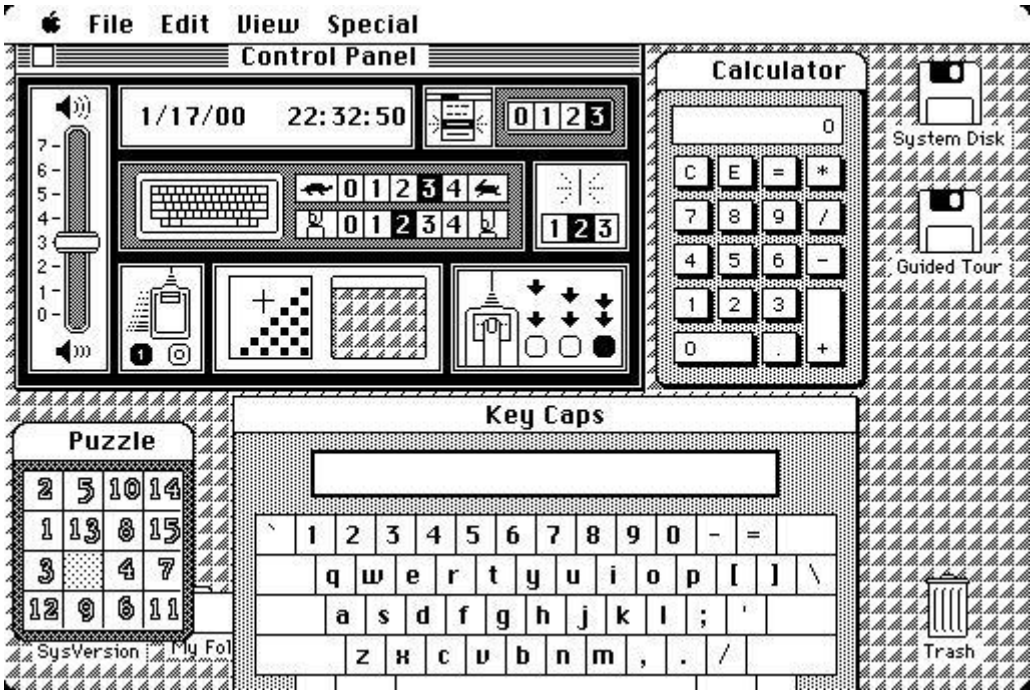
LA TEORÍA DE LA

CONSCIENCIA DE LA INTERFAZ

DE DONALD HOFFMAN



PROYECTO FIN DE MÁSTER DEL MÁSTER EN CIENCIAS COGNITIVAS DE LA UNIVERSIDAD DE MÁLAGA

Curso: 2017/18

Autor: Santiago Sánchez-Migallón Jiménez

Director: D. Antonio Diéguez Lucena

A mis padres

CONTENIDO

1. INTRODUCCIÓN 4

1.1. POR UNA TEORÍA DE LA CONSCIENCIA NATURALIZADA 4
1.2. POR UNA TEORÍA DE LA CONSCIENCIA EVOLUCIONISTA 9

2. LA TEORÍA DE LA PERCEPCIÓN DE LA INTERFAZ 12

2.1. TIPOS DE ESTRATEGIAS PERCEPTIVAS 12
2.2. JUEGOS EVOLUTIVOS 15
2.3. ALGORITMOS GENÉTICOS 17
2.4. LA TEORÍA DE LA PERCEPCIÓN DEL INTERFAZ 19
2.4.1. EL MARCO BAYESIANO ESTÁNDAR PARA LA VISIÓN Y SU LIMITACIÓN 21
2.4.2. PERCEPCIÓN COMPUTACIONAL EVOLUTIVA 22
2.4.3. EL BUCLE PERCEPCIÓN-DECISIÓN-ACCIÓN (PDA) 24
2.4.4. LAS OBJECCIONES DEL MUNDO OBJETIVA E INTERSUBJETIMANTE MEDIBLE 26
2.4.5. ILUSIONES Y ALUCINACIONES 29

3. LA TEORÍA DE LA CONSCIENCIA DE LA INTERFAZ 31

3.1. EL PROBLEMA MENTE-CUERPO 31
3.1.1. EL ERROR DE ENTENDER LA PERCEPCIÓN COMO UNA REPRESENTACIÓN FIEL 33
3.1.2. LAS INTERFACES DE USUARIO COMO OCULTAMIENTO DE LA COMPLEJIDAD CAUSAL 34
3.1.3. LA HIPÓTESIS DE LA INTERFAZ DE USUARIO MULTIMODAL (MUI) 35
3.1.4. REALISMO CONSCIENTE 40
3.4.5. EL PROBLEMA MENTE-CUERPO 45
3.4.6. OBJECIONES DESDE LA TEORÍA DE LA EVOLUCIÓN 48

4. REVISIÓN CRÍTICA Y CONCLUSIONES 50

4.1. LA PARADOJA DE EPIMÉNIDES 50
4.2. IRREDUCTIBLES QUALIA 51
4.3. ESCAPANDO DEL TEATRO CARTESIANO 52
4.4. PERO ENCERRADOS EN LA CÁRCEL EPISTÉMICA DEL SOLIPSISMO 53
4.5. VALORACIÓN FINAL 54
4.5.1. EL GRAN ACIERTO 54
4.5.2. LA TESIS PROBLEMÁTICA 56
4.5.3 EL TRABAJO QUE QUEDA POR HACER 58

REFERENCIAS 61

BIBLIOGRAFÍA 66

1. INTRODUCCIÓN

1.1. POR UNA TEORÍA DE LA CONSCIENCIA NATURALIZADA

A lo largo de la historia del conocimiento humano, la consciencia ha sido tratada, o bien desde la filosofía, o bien desde perspectivas no científicas, en el sentido de no utilizar una metodología propiamente científica (método hipotético-deductivo-experimental) para abordarla ¿Por qué? En primer lugar, porque la ciencia moderna no apareció hasta, aproximadamente, el siglo XVII (y la psicología como disciplina propiamente científica llegó mucho más tarde, no naciendo como ciencia independiente hasta finales del siglo XIX,) y, en segundo lugar, porque la ciencia se volcó en el estudio del mundo físico, lugar del que la filosofía moderna había expulsado la mente.

Descartes, punto de partida de la Modernidad, diferenciaba radicalmente la *res extensa* (el mundo medible, cuantificable, tridimensional) de la *res cogitans* (una especie de *cajón de sastre* en el que Descartes metía todo lo que pudiese ser denominado como mental). Mundo y mente eran dos sustancias diferentes, en un sentido aristotélico, es decir, dos entidades completamente independientes una de la otra. De aquí que el propio Descartes tuviese, después, enormes problemas para relacionar cuerpo y mente, y su solución al problema, la celebérrima *glándula pineal*, fuera absolutamente insatisfactorio (Descartes, 1977).

Descartes había abierto la Edad Moderna estableciendo un abismo ontológico insalvable entre cuerpo y mente. Desde entonces, el cuerpo era un objeto material y mecánico, propio de la naturaleza, como cualquier otro; y la mente, al no ser extensa, no pertenecía a este mundo y, además, al ser tradicionalmente identificada con el alma inmortal, se la circunscribía al ámbito de lo sobrenatural, a los ámbitos metafísico y

religioso. Descartes rompía con la visión aristotélica en la que el *ánima* era el principio vital (causa eficiente de movimiento) de los seres vivos, que los diferenciaba de los seres inertes. Para Aristóteles el dualismo, a pesar de existir en cierta medida en su distinción materia/forma heredada del platonismo, no estaba tan marcado: Materia y forma son físicamente inseparables. Solo cuando conocemos podemos abstraer la forma en nuestra mente, pero en la realidad sería absurdo una mente sin cuerpo, de la misma manera que sería absurdo contemplar la *altura en sí*, sin darse un *objeto alto* en el que observarla: los accidentes son inseparables de la sustancia, incluso si hablamos de atributos esenciales o sustancias segundas.

Además, la mente, tiene una peculiaridad que la hace especialmente susceptible de ser expulsada de la naturaleza: solo podemos acceder a ella directamente en primera persona, solo yo tengo acceso directo a mis contenidos mentales (a los contenidos mentales de los demás solo tengo acceso indirecto, a través de *indicios externos*. Por ejemplo, si alguien está llorando puedo inferir que siente *interiormente* tristeza, aunque nunca puedo tener certeza de ello). Es, más, solo tengo acceso a los míos, no pudiendo saber de ninguna forma que otros los tienen. Este es el, también muy conocido, *escepticismo hacia las otras mentes*: si llevamos la idea al extremo, solo puedo tener certeza de que yo tengo estados mentales, cabiendo la posibilidad de que nadie más en todo el universo los tenga. Esta radical idea es el *solipsismo* del que Descartes nunca pudo salir.

Sin embargo, con los objetos propios del mundo natural no existe ese problema: son intersubjetivos, todos podemos tener acceso directo a ellos. Todos podemos ver, y observar pormenorizadamente, una roca, pero solo yo puedo tener acceso a mi sentimiento de tristeza o a mi creencia en que Londres es la capital de Gran Bretaña. Esta peculiaridad, en principio solo epistemológica, ha terminado por volverse ontológica: si

el acceso a los objetos mentales es diferente al acceso a los objetos físicos o naturales, será porque son ontológicamente diferentes.

Así mismo, dada la dificultad de tratar lo mental, una forma de hacerlo fue *por oposición a lo físico*: si A es una propiedad de los objetos físicos, A no es propiedad de los objetos mentales y, es más, ¬A será una propiedad de los objetos mentales. Así, por ejemplo, si la propiedad más patente de los objetos físicos, es su extensión, los objetos mentales serán inextensos. Si los físicos son corpóreos, los mentales incorpóreos; si los físicos corruptibles, los mentales incorruptibles; los físicos divisibles en partes, los mentales indivisibles, etc. Esto se vio además favorecido por la visión cristiana, que, por preceptos religiosos, necesitaba que la mente fuera inmortal. Si todos los objetos físicos son corruptibles y duran solo un tiempo, la mente no ha de ser física si queremos que sea inmortal. Por eso el cristianismo necesita apoyarse en una perspectiva dualista, y dado que el cristianismo ha sido la religión que ha dominado occidente en toda la Edad Moderna, el dualismo quedó muy respaldado.

Sin embargo, cuando la Edad Moderna avanzó y la revolución científica fue consolidándose, las posturas materialistas, fisicalistas y, en consecuencia, monistas fueron cobrando más fuerza y, este clásico dualismo cartesiano fue poniéndose en duda. De hecho, los más insignes racionalistas que continuaron la estela cartesiana dedicaron bastantes esfuerzos a intentar solucionar el problema mente-cuerpo que Descartes no logró resolver. Spinoza lo hace mediante un monismo naturalista y panteísta (Spinoza, 2011), Leibniz con su conocidísima *teoría de la armonía preestablecida* (Leibniz, 1992) o Malebranche con su *ocasionalismo* (Malebranche, 2009).

Desde la perspectiva opuesta pronto llegó la crítica. En general, al dualismo, se le criticó desde un empirismo que pone en duda cualquier entidad que no proceda de la más

preclara experiencia. La crítica a la metafísica de Hume (Hume, 2004) llegará incluso a la misma noción de *cogito cartesiano*, la primera verdad indudable sobre la que el francés edificaba todo el edificio del saber.

Tenemos exponentes en el lado más radical de la ilustración como La Mettrie y su *El hombre máquina* (La Mettrie, 1748), donde se rompe radicalmente con cualquier forma de dualismo y de espiritualismo. La Mettrie defenderá un materialismo mecanicista estricto que concebirá al ser humano como una máquina más. La Revolución Científica, sobre todo en el XVIII, supuso el gran triunfo del mecanicismo, con la mecánica newtoniana (Newton, 1962). Era la primera vez en la historia que una única ley, la de la gravitación, explicaba de una forma sencilla, todos los movimientos del universo (se rompía de una vez por todas con la distinción aristotélica entre mundo sublunar y supralunar). Para comprender el movimiento de los astros no hacía falta recurrir a ninguna fuerza o energía sobrenatural, sino que, entendiendo el universo como un gran reloj, solo con materia y fuerzas ejercidas sobre ella, podemos comprenderlo todo. Así, el proyecto filosófico de Hume, y que en cierta medida compartirá toda la gnoseología y la filosofía de la mente de su época, fue explicar la mente con el mismo rigor, y el consecuente éxito, que Newton había tenido para explicar el universo (Newton se considera el modelo de ciencia estricta. Ya el mismo Kant no se pregunta si la ciencia es posible, sino solo cómo es que es posible. Kant nunca pone en duda la ciencia de Newton). Así, da la impresión, de que Hume habla de las ideas y de las impresiones (los componentes de la mente) como si de planetas o astros se tratara (los componentes del universo), y de las normas que los rigen (principios de asociación) como si de leyes naturales estuviera hablando. La mente se hace mecánica.

Sin embargo, el mecanicismo clásico entra en crisis en el siglo XIX, y, sobretodo, a principios del XX, con la llegada del paradigma cuántico y relativista. A nivel filosófico,

se encontró, entre otros, con el problema de ser incapaz de explicar concluyentemente la emergencia de la consciencia de la materia inerte. Hace unos años el filósofo australiano David Chalmers expuso magistralmente el problema, denominándolo, ya de un modo canónico, como el *hard problem of consciousness* (Chalmers, 1996). Chalmers sostenía que, mientras ciertas cualidades de nuestra mente eran, relativamente, fáciles de comprender y emular computacionalmente desde un paradigma científico, propiamente materialista-naturalista (eran *easy problems*), existía una cualidad, la *consciencia fenoménica*, que se escapaba a nuestras explicaciones. Thomas Nagel en su célebre artículo *"¿Qué se siente al ser un murciélago?"* (Nagel, 2000), llegó a la misma conclusión: nuestra experiencia consciente, nuestros *quale*, parecen irreductibles a la explicación científica. En la actualidad, el reciente libro de Markus Gabriel (Gabriel, 2018), sigue la línea de retomar argumentos clásicos en contra de esta naturalización de la consciencia.

En el lado opuesto de la actualidad, tendríamos a aquellos que, de algún modo, entienden que la consciencia es una ilusión, como Gilbert Ryle (1949) y su famosa hipótesis del *fantasma en la máquina*, o su discípulo Daniel Dennett (2017). También tenemos las teorías de la identidad, el materialismo reductivo o eliminativo, o el fisicalismo de tipo, postulados inicialmente por Place (1970) o Smart (2014), y muy popularmente defendidas por Patricia Churchland (2013). Estas posturas vienen a sostener *grosso modo*, que existen estados físicos tal que son idénticos a estados mentales, de tal modo que la consciencia es, totalmente, reductible a la materia. No obstante, desde nuestro juicio, no llegan a solucionar el *hard problem*, y terminan por encallarse en los clásicos problemas en los que siglos atrás fueron cayendo los filósofos que se enfrentaron a ellos.

Es por ello que la teoría de Hoffman que desarrollamos en este trabajo constituye una nueva forma de enfrentarse al problema, no en el sentido de ser un nuevo tipo de idealismo (que, para nada es nuevo), ni porque a nadie se le hubiera ocurrido antes decir que los objetos de nuestra consciencia no tienen por qué corresponderse con los objetos de la realidad (tesis ésta tan vieja como la misma filosofía), ni siquiera por utilizar la ilustrativa metáfora del interfaz de un PC que suele identificar su teoría (la idea de que los objetos de la consciencia son *esquemas útiles* también es bastante vieja), sino porque la teoría de Hoffman intenta fundamentarse científicamente desde perspectivas tan, relativamente, novedosas como la teoría de juegos o los algoritmos genéticos. Hoffman pretende una muy interesante naturalización de la mente, a la que, además, quiere dar una buena base matemática, y la máxima evidencia empírica posible. Como veremos, Hoffman recurre a diversos principios de física cuántica para ello. La teoría de la interfaz constituye un nuevo abordaje pretendidamente científico de la consciencia[1].

1.2. POR UNA TEORÍA DE LA CONSCIENCIA EVOLUCIONISTA

La verdadera revolución darwiniana no vino con *El Origen de las Especies*, sino con una obra algo posterior: *The Descent of Man, and Selection in Relation to Sex* publicado en 1871. Lo crucial no era tanto que las especies evolucionaban unas de otras, como que el mismo hombre era un animal más, y por tanto nada especial; y que su idolatrada mente no era más que un producto de la evolución como cualquier otro. La teoría de la evolución darwiniana supuso un nuevo golpe al dualismo sobrenaturalista. Tanto fue así que muchos evolucionistas, como el mismo Alfred Wallace, aceptaban la evolución para todo rasgo o capacidad del organismo menos para la mente (Wallace, 2007).

[1] Para un abordaje de la naturalización de la consciencia véase Petitot et al. 1999.

A pesar de todo, la idea fue para adelante y se intentaron explicar todos los rasgos de la mente como adaptaciones o ventajas evolutivas. En algunos casos, esto encaja perfectamente: parece evidente que una mejor coordinación entre la percepción y el movimiento, función primordial de todo sistema nervioso, otorga mejores posibilidades de supervivencia. Del mismo modo, una mejor memoria, mejor capacidad de planificación, mejor capacidad de aprendizaje, etc. suponen, a todas luces, ventajas en la competitiva *struggle of life*. De hecho, ciertas definiciones actuales de inteligencia, no dudan en subrayar un importante componente adaptativo (Sternberg & Sternberg, 1985).

Pero pronto empiezan los problemas: ¿todas las características de la mente se pueden explicar claramente como adaptaciones? ¿En qué puede aumentar mi fitness escribir un poema, pintar un cuadro o dedicarme a reflexionar sobre el sentido de la existencia? Las facultades que, culturalmente, consideramos de más altura parecen, precisamente, las que presentan una mayor inutilidad adaptativa. Además, parece que tenemos un *excedente cognitivo*, es decir, tenemos unas cualidades que, aunque, en su origen, tuvieran una clara función adaptativa, en la actualidad han superado con mucho esa función. Por ejemplo, no cabe duda de que nuestra capacidad matemática excede en mucho lo necesario para la supervivencia y la obtención de pareja. Resolver ecuaciones diofánticas no tiene ninguna utilidad adaptativa directa por mucho que se quiera retorcer el tema. Es más, ¿en qué puede aumentar mi fitness conductas autodestructivas que pueden llegar, incluso, al suicidio?

En una crítica decidida al programa adaptacionista, es muy famoso el artículo de Lewontin y Jay-Gould (1979) en donde se defiende que no toda cualidad fenotípica puede explicarse como una adaptación. Gould y Lewontin ponen el claro ejemplo de las enjutas o pechinas de la catedral de San Marcos: surgen como una necesidad estructural de la cúpula, como un subproducto de una auténtica adaptación (supongamos que la cúpula es

una adaptación), y que luego se han utilizado con fines ornamentales (un subproducto se convierte luego en una auténtica adaptación). Fenómenos como la fijación aleatoria de alelos, la alometría, la pleiotropía, la retribución material o la correlación forzada mecánicamente, las exaptaciones, o los epifenómenos que luego se han reconvertido en auténticas adaptaciones, no pueden explicarse en los términos de optimización típicos del programa adaptacionista.

Si aplicamos esta crítica a la mente podemos solucionar, *grosso modo*, el problema de las funciones difícilmente adaptativas de nuestra mente, pero encontramos nuevas dificultades: si una función mental no es adaptativa, ¿para qué vale? ¿Qué hace ahí? Dennett sostiene que esto nos lleva a una *hipótesis vacía*, si bien creemos que se equivoca: no estamos ante una hipótesis vacía, solo que la respuesta se hace mucho más compleja: ahora no solo hay que fijarse en la posible función adaptativa del rasgo a estudiar, sino en sus relaciones con los demás rasgos y en su historia evolutiva. Y aquí surge una cuestión crucial: vuelve a aparecer el *hard problem* de Chalmers: ¿para qué vale la consciencia? ¿Cuál puede ser su función adaptativa? Si pudiésemos imaginar un zombi cuya conducta es completamente similar a la de un ser humano, pero careciendo de consciencia, ¿para qué la evolución habría seleccionado la consciencia?

Pero, ¿Acaso es posible imaginar los zombis de Chalmers? En un sentido, claramente sí: las computadoras. Nuestros sistemas de IA son capaces de conductas sumamente inteligentes sin consciencia alguna y, es más, son capaces de imitar conductas en las que nosotros utilizamos la consciencia, sin tenerla. Entonces, la consciencia quedaría como un epifenómeno, un subproducto de ventajas evolutivas que carece de ninguna función adaptativa. Empero, por otro lado, la consciencia parece una cualidad compleja y costosa de mantener: ¿por qué la evolución habría mantenido un subproducto tan caro? Parece difícil sostener que la consciencia no tiene ninguna función adaptativa,

más cuando las sensaciones de placer y dolor explican muy bien la optimización de fitness en términos adaptacionistas.

Enmarcada en esta problemática, aunque de pretensiones mucho más humildes, la teoría de la interfaz de Hoffman parte desde una perspectiva completamente evolucionista: va a entender la percepción, la cognición y la consciencia desde un enfoque evolutivo. La teoría de Hoffman va a constituir un intento de gnoseología evolucionista que va a partir, justamente, de una cuestión fruto de la teoría de la evolución: ¿Es la teoría del conocimiento de corte realista la más acorde con la evolución? ¿Qué estrategia evolutiva otorga más fitness: el realismo o un *ficcionalismo útil*? Es muy llamativo el hecho de que la mayor parte de la comunidad científica, y filosófica en el ámbito anglosajón, mantiene posturas realistas, mientras que, como veremos, lo más acorde con la teoría evolutiva no es, precisamente, el realismo.

2. LA TEORÍA DE LA PERCEPCIÓN DE LA INTERFAZ

2.1. TIPOS DE ESTRATEGIAS PERCEPTIVAS

Comenzamos examinando las diferentes posibles estrategias perceptivas que un organismo podría adoptar en un nicho ecológico en el que compite con otros organismos por optimizar su fitness. Seguiremos aquí las definiciones y líneas argumentativas expuestas por Hoffman et al. en "The Interface Theory of Perception" (2016).

En primer lugar, se define lo que es una estrategia perceptual, dándonos dos definiciones (una libre de dispersión, o ruido, y otra no):

> *Definition 1 A (dispersión-free) perceptual strategy, P is a measurable function P:W→X wheres (W, W) denotes a measurable space of states of the world and (X, X) denotes a measurable space of perceptual experiences.*

Definition 2 A perceptual strategy with dispersion is a Markovian Kernel P:WxX→[0,1], where W denotes a measurable space of states of the world an X denotes the events for a measurable space X of perceptual experiences.

Hoffman, sencillamente, define las estrategias perceptuales como funciones que relacionan un mundo W con un número mesurable de estados *W*, con un mundo perceptivo X, con un número mesurable de experiencias perceptivas *X*. Estas funciones pueden dibujarse en una matriz estocástica cuyo núcleo de transición es un núcleo de Markov que asigna valores finitos entre 0 y 1.

Pasamos a describir las posibles estrategias perceptivas que pueden adoptarse:

Definition 3 An omniscient realist strategy is a perceptual strategy for which X=W and P is an isomorphism.

Es la estrategia de un observador omnisciente que percibe toda la realidad tal y como es. Para Hoffman es evidente que no es un buen modelo para explicar la percepción ya que parece claro que ningún organismo tiene una percepción omnisciente del mundo. Estaríamos ante "el ojo que todo lo ve" bien descrito por Koenderink (2014). La percepción absoluta sería un derroche de recursos completamente absurdo en términos evolutivos ya que a cualquier organismo le basta con muchísima menos información del mundo para sobrevivir en él.

Definition 4 A naïve realist strategy is a perceptual strategy for wich X ⊂W and P is an isomorphism on this subset that preserves all structures on W.

El observador conoce una parte de la realidad, si bien tiene un conocimiento perfecto de esa parte.

> *Definition 5 A critical realist is a perceptual strategy for which X need not be a subset of W, but P is nervertheless a homomorfism that preserves all structures on W.*

Las percepciones no necesitan ser un subconjunto del mundo real, pero en las relaciones entre percepciones se preservan las relaciones entre los estados en el mundo objetivo.

> *Definition 6 A hybrid realist strategy is a critical realist strategy that requires that there exists a strict subset $X \subset X$ that satisfies $X \subset W$ and requires that P is an isomorphism on this subset that preserves all structures.*

Es la clásica distinción de Locke entre cualidades primarias y secundarias. Hay una parte de la realidad que se percibe perfectamente y de la otra, aunque no se la perciba con veracidad, se conserva homomorfismo estructural.

> *Definition 7 An interface perceptual strategy is a perceptual strategy that does not require X to be a subset of W and for which the mapping P has no restrictions other than being measurable (so that the probabilities of perceptions are systematically related to probabilities of events in W).*

Por último, la estrategia ganadora: las percepciones no son un subconjunto del mundo real y ninguna estructura del mundo se preserva en las relaciones entre percepciones. La estrategia de la interfaz no percibe ninguna información real. Entonces, cabe preguntarse: ¿cómo es posible que la selección natural eligiera esta estrategia? ¿Cómo podría sobrevivir un organismo que no perciba nada verídico del mundo que le rodea? Hoffman lo demuestra recurriendo a la teoría de juegos aplicada a la evolución.

2.2. JUEGOS EVOLUTIVOS

Hoffman compara la estrategia perceptiva realista con la de interfaz mediante un juego. Imaginemos un territorio en el que existen recursos cuyo valor oscila entre 0 y 100, y donde las percepciones de cada participante están limitadas a cuatro colores: rojo, amarillo, verde y azul. Los "objetos" de cada color tendrán un premio y la cuantía de los premios está ordenada en función de los colores de la siguiente forma: azul > verde > amarillo > rojo. La función de premios es una función gaussiana en la que ganan los valores intermedios y pierden los valores excesivos (los que más se alejen de 50 por encima o por debajo). Hoffman intenta emular de modo muy simplificado la homeostasis típica de los organismos biológicos: intentar siempre volver a un valor inicial intermedio.

La estrategia del realismo crítico percibiría de modo realista los cuatro colores (fig.1) y pondera perfectamente el intervalo de premios que puede dar cada color (por ejemplo, el azul da de 75 a 100). También vemos que es homomórfica porque preserva la estructura ascendente del orden de premios. Entonces, la estrategia realista escogería, la mayor parte de las veces, los colores que dan premios intermedios (amarillo y verde), y menos los que dan muchos o pocos (rojo y azul). Esta estrategia será óptima percibiendo la realidad (ya que percibe correctamente los colores y el intervalo de premio asociado a cada color) pero no será óptima a la hora de obtener beneficios.

Hoffman la compara con una segunda estrategia de tipo interfaz (fig.2). En ella el mapeo de los recursos no es un homomorfismo ya que no respeta el orden ascendente, sino que, para el mismo color, utiliza ambos órdenes. El homomorfismo se da solo con la estructura de beneficios, por lo que, necesariamente, termina por ser más óptima que la realista. Hoffman, además, se refiere a simulaciones de Monte Carlo hechas con distintas versiones de este juego (Mark, 2013 y Mark *et al*. 2010) en las que se aumenta la complejidad (se amplía el espacio, el número de recursos y de jugadores) y que, vuelven

a mostrar que la estrategia realista pierde ante la de interfaz. Es más, la complejidad la perjudica, ya que la realista tiene que almacenar mucha más información irrelevante para obtener beneficios. Estas conclusiones refuerzan mucho la tesis de Hoffman, dado que los entornos simulados son muy simples en comparación con la realidad natural, infinitamente más compleja, por lo que, si las estrategias realistas tienen ya problemas a esta escala simple, cuántos más tendrán en escalas varios órdenes de magnitud más complejas. El único entorno en donde las estrategias realistas son mejores que las de interfaz es cuando los premios varían monotónicamente con la percepción de la realidad. Cuando rompemos esa monotonía, las realistas siempre pierden. Y eso, de nuevo, es una poderosa razón a favor de la teoría de la interfaz porque ¿qué razón hay para que exista tal monotonía? Lo habitual en el entorno natural serán escenarios complejos en los que no exista tal ajuste monotónico.

La estrategia de interfaz optimiza mejor que la realista los beneficios porque, sostiene Hoffman, siendo ésta su tesis fuerte, la selección natural ajusta la percepción a los beneficios y no a la verdad.

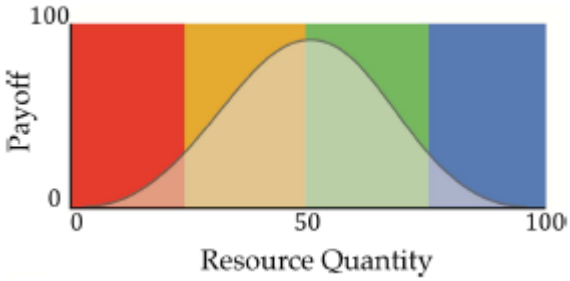


Fig. 1. Gráfico que representa la función de pagos para una estrategia de realismo crítico. Como vemos la función es aproximadamente gaussiana. Sacado del original Hoffman, D. et al. “The Interface Theory of Percepción” (2016. p.7).

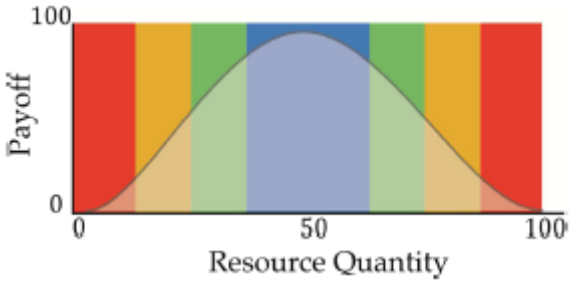


Fig. 2. Gráfico que representa la función de pagos para una estrategia de la interfaz. Como vemos la función de pagos daría más ganancias que la del realismo crítico. Sacado del original Hoffman, D. et al. "The Interface Theory of Percepción" (2016. p.7).

2.3. ALGORITMOS GENÉTICOS

Hoffman se hace una nueva pregunta: si vemos que las estrategias realistas son pobres en comparación con las de interfaz en el terreno de juego, ¿la evolución las habrá puesto, si quiera, a jugar o ni siquiera fue rentable hacerlo desde el principio?

Para resolver esta cuestión Hoffman acude a los algoritmos genéticos y, concretamente, nos pone el ejemplo del robot virtual Robby (Michell, 1998). Robby se encargaba de recoger refrescos distribuidos aleatoriamente en una malla 10 x 10. Alrededor de ella hay un muro que se modela como un perímetro de cuadrados, por lo que el mundo de Robby crece y se representa por una malla 12 x 12. Cada casilla se valora de la siguiente forma: 0 si está vacía, 1 si hay una lata de refresco y 2 si hay un muro. Robby puede solamente percibir las cuatro casillas adyacentes a la que se encuentra en cada momento (no percibe en diagonal). El objetivo del algoritmo genético es hacer evolucionar a Robby para conseguir que sea más eficiente a la hora de recolectar latas de refresco, desconociendo éste el estado de la malla. Con este fin, Robby tiene un conjunto (G) de siete acciones primitivas que puede ejecutar: permanecer donde está, recoger una lata, avanzar hacia el norte, avanzar hacia el sur, avanzar hacia el este, hacia el oeste o dar un paso al azar. El algoritmo genético debe aprender qué combinaciones de estas acciones son las óptimas en conjunción con los distintos estados perceptivos de Robby.

La función de pagos es así: Robby recibe 10 puntos por cada lata que coge, pero pierde un punto por cada tiempo en que pasa buscando sin encontrar nada, y pierde 5 cada vez que intenta buscar en el muro perimetral. Cada robot Robby que se prueba tiene aproximadamente 240 genes que corresponden al conjunto G de acciones primitivas que pueden tomarse como respuesta a cada percepción. Mitchell empleó 200 robots en la primera generación con genes completamente elegidos al azar. Cada robot se enfrentaba a 100 mundos con latas distribuidas, igualmente, al azar, y podría ejecutar 200 acciones en cada mundo. El fitness de cada robot era el resultado de la función de pagos obtenida después de recolectar sus 100 mundos correspondientes. Todo este proceso se repite hasta 1000 generaciones.

En un principio, todos los robots son *realistas naïve*, ya que conocen el verdadero estado de, al menos, sus cuatro casillas adyacentes. Entonces Mark (2013) va a modificarlo para comparar estrategias perceptivas. Se permite que en cada casilla pueda haber hasta diez latas y la función de pago será ésta: (0,1,3,6,9,10,9,6,3,1,0). Por ejemplo, Robby recibirá un punto en las casillas que tengan una o nueve latas, o tres puntos si tiene tres o siete. Cada robot no puede ver el número exacto de latas que hay en cada casilla sino solo dos colores: rojo y verde. La estrategia perceptiva consistirá en intentar asociar dichos colores con la función de pago de modo que se obtengan los mismos beneficios. Para que la estrategia perceptiva coevolucionara con la estrategia de búsqueda de Mitchell, Mark añadió a los robots 11 genes más que equivaldrían a como cada Robby asigna un color a cada número posible de latas.

Como era de esperar, al principio, los robots son cómicamente estúpidos en sus estrategias, pero aproximadamente a partir de la generación 500, comienzan a surgir hábiles recolectores, y todos ellos utilizan dos estrategias: o las casillas se ven rojas si contienen 0, 1, 9 o 10 latas y si no se ven verdes o, lo mismo, pero a la inversa: las casillas

se ven verdes si contienen 0, 1, 9 o 10 latas y si no se ven rojas. Como vemos estamos ante una estrategia de interfaz estricta. La estrategia realista se basaría en, por ejemplo, ver como rojos las casillas que tengan entre 0 y 5 latas y ver como verdes las que tengan entre 6 y 10. De nuevo, la estrategia del interfaz es superior a la realista. Pero lo que Hoffman quiere comprobar aquí no es solo eso, sino si las estrategias realistas llegaron a aparecer alguna vez en el juego evolutivo. En la simulación de Mark hay espacio para 2048 (2^{11}) estrategias perceptivas, por lo que es probable que las estrategias realistas fueran probadas y desechadas durante las primeras 500 generaciones de evolución. Pero en un caso más complicado con 30 genes y 10 posibles colores el espacio de búsqueda subiría a 10^{30} estrategias perceptivas posibles, por lo que es posible que una estrategia realista, sin presiones selectivas a su favor, jamás apareciera en ninguna generación (ya que solo podrían hacerlo por casualidad). Según Hoffman, contra toda intuición inicial, son estrategias tan malas que ni siquiera merecería la pena intentarlas.

2.4. LA TEORÍA DE LA PERCEPCIÓN DEL INTERFAZ

Hoffman no se cansa de repetir la metáfora ilustrativa de lo que es la teoría del interfaz: tenemos que pensar que las relaciones entre nuestras percepciones y la realidad son análogas a las que se dan entre nuestro ordenador y el escritorio del sistema operativo (o de cualquier interfaz gráfica o GUI). Cuando yo ejecuto, por ejemplo, un reproductor de vídeo, el icono en el que hago clic para activarlo no me aporta ningún tipo de información realista acerca de todo el mecanismo de circuitos, voltajes y campos magnéticos (de todo el *hardware*) que se pone en funcionamiento para que yo pueda ver cómodamente un vídeo cualquiera. Sería una terrible pérdida de tiempo y recursos tener que percibir con precisión el funcionamiento interno del ordenador cada vez que quiero hacer algo con él. Así, la selección natural habría moldeado nuestros sistemas perceptivos

no para contemplar la realidad en sí, sino todo lo contrario: para ocultar su inmensa complejidad.

Hoffman le saca más partido a su metáfora: el espacio y el tiempo serían el escritorio (el continente donde ocurre todo contenido), mientras que los objetos percibidos y sus propiedades serían los iconos (Hoffman, 2016. p.9). Y así como las expresiones *escritorio* e *icono* no describen correctamente la estructura de una computadora, el espacio y el tiempo tampoco describen la auténtica estructura de la realidad. No obstante, eso no quiere decir que sean completamente falsas en el sentido de ser una alucinación. Su función es servir de guías adecuadas para las conductas adaptativas. Cuando hacemos clic en el icono del reproductor de vídeo, el reproductor funciona, ejecutamos eficazmente una acción.

Nuestras percepciones están ajustadas por la selección natural para maximizar el fitness de nuestro organismo. Así, la distinción entre verdad y fitness es esencial en la teoría evolutiva. El fitness es una función de la realidad objetiva, pero no depende sólo de ésta, sino de las necesidades biológicas del organismo. Por ejemplo, nos ilustra Hoffman, para una mosca hambrienta un montón de estiércol puede traducirse en excelente fitness, mientras que para un humano no. El fitness es, en términos generales, una función compleja del mundo objetivo que depende de un organismo, de su estado y de su acción (*Ibíd*. p.10).

Esta idea no es nada intuitiva ya que lo más natural es pensar que el mundo que me rodea es real y que lo percibo, como mínimo *aproximadamente*, tal cual es. Así, Hoffman cita a Bertrand Russell cuando el famoso filósofo británico sostenía que la disposición de los objetos en la realidad debería ser similar a la de la posición relativa de los datos sensoriales en nuestros “espacios privados”. Parece lógico pensar que, aunque los colores que percibimos no son reales, ya que varían según la iluminación o,

sencillamente, ya que lo que sabemos sobre la realidad no es que los objetos sean homogéneos, sino que son un vacío en el que transitan erráticamente minúsculas partículas; pero de la disposición espacial de los objetos parece más difícil dudar: ¿acaso no está delante de mi esa mesa que no solo veo, sino que puedo tocar, mover, incluso oler y saborear…? Para Hoffman no hay razón alguna para pensar que eso tenga que ser así. Como veremos más adelante, con la *metáfora de la realidad virtual* se puede desmontar con facilidad tal objeción.

2.4.1. EL MARCO BAYESIANO ESTÁNDAR PARA LA VISIÓN Y SU LIMITACIÓN

El marco estándar contemporáneo para investigar la visión enfoca el tema como si de un problema inductivo se tratara. Se considera que cualquier imagen en la retina es compatible con infinitas interpretaciones posibles. La misma imagen podría haber sido generada por infinitas escenas tridimensionales. La cuestión sería cómo el sistema visual converge en una única interpretación entre todas las posibles. Esta ambigüedad fundamental inherente a la percepción solo se puede resolver aplicando sesgos o restricciones adicionales, por ejemplo, con respecto a cuán probables sean las diferentes interpretaciones de escena. El entorno en el que evolucionó nuestra especie es un lugar altamente estructurado, que contiene muchas regularidades. La luz tiende a venir de arriba, hay una prevalencia de estructuras simétricas, los objetos tienden a ser compactos y se componen de partes que son en gran parte convexas, etc. En el transcurso de la evolución, tales regularidades han sido internalizadas por el sistema visual. Por lo tanto, ayudan a definir los sesgos probabilísticos que hacen que algunas interpretaciones de una imagen sean mucho más probables que otras.

Si lo formalizamos, tenemos que dado un input y_0 el sistema visual calcula las probabilidades posteriores $p\ (x \mid y_0\ )$, para las interpretaciones de la escena candidata.

Aplicando el Teorema de Bayes, la probabilidad posterior es proporcional al producto de la probabilidad de la escena con su probabilidad previa.

$$p\left(x\middle|y_0\right) \propto p\left(y_0\middle|x\right)p(x)$$

La probabilidad previa consiste en el conocimiento implícito del sistema visual dado en la experiencia onto y filogenética que causa que unas escenas tengan más probabilidad que otras. Habitualmente, según esta función de probabilidad se selecciona la escena con mayor probabilidad como la mejor interpretación.

Según Hoffman, el marco bayesiano estándar para la visión establece ciertas suposiciones clave que lo hacen demasiado limitado. En él, los inputs *x* recibidos representan tanto estados del mundo como interpretaciones posibles de éste. Es decir, que se presupone que el conjunto de hipótesis del observador es idéntico al mundo objetivo. Y aquí está el error fundamental: el marco bayesiano parte del supuesto de que el individuo percibe verídicamente la realidad. Recordando las estrategias perceptivas, el marco bayesiano supone que X=W (o que X es isomorfo a W), lo cual nos lleva a un realismo ingenuo que, como ya mostramos, no sería seleccionado evolutivamente. Hoffman muestra la necesidad de un nuevo marco que incorpore la estrategia perceptiva del interfaz.

2.4.2. PERCEPCIÓN COMPUTACIONAL EVOLUTIVA

Hoffman generaliza el enfoque bayesiano en un nuevo marco que denomina *Percepción Computacional Evolutiva* (ECP). Dada la naturaleza intrínsecamente inductiva de la percepción, la ECP incorpora la inferencia probabilística de una forma fundamental, pero sitúa el mundo objetivo *W*, fuera del aparato inferencial bayesiano. *X* e *Y* van a ser dos espacios de representación que no presuponen en ningún momento

ninguna equivalencia con *W*. Por ejemplo, *Y* podría ser un tipo de representación de nivel inferior (por ejemplo, una estructura de una imagen 2D) y *X* podría ser otra representación, pero de un nivel más superior (por ejemplo, involucra una estructura de una imagen 3D), sin que, de ninguna forma, se de alguna equivalencia entre ellas y *W*.

De la misma forma Hoffman va a introducir el fitness, con la intención de introducir el factor evolutivo que, además, es la clave de las relaciones entre el organismo y el mundo objetivo. Dichas relaciones van a estar mediadas por lo que Hoffman llama *canales*, que no son más que cada una de las estrategias perceptivas que vimos anteriormente. Así, un canal es la estrategia perceptiva que un organismo usa para interactuar con el mundo objetivo. Y como ya vimos, la estrategia perceptiva idónea se consigue calculando el fitness obtenido mediante ella. El fitness depende no solo del estado objetivo del mundo, sino también del organismo en cuestión, su estado y el tipo de acción que se está considerando. Hoffman define entonces una función de adecuación global f: W × O × S × A → $\mathbb{R}$ + donde *O* es el conjunto de organismos, *S* sus estados posibles y *A* sus posibles clases de acción. Una vez que tenemos un organismo particular en *O*, su estado está en *S*, y su clase de acción en *A*, tenemos una función específica de aptitud $f_{o,s,a}$: W → $\mathbb{R}$ + que asigna puntos de fitness (números reales no negativos) a cada *w* en *W*.

Para calcular el fitness se define el concepto de *observador ideal darwiniano*:

> *Definition 8 Given a specific fitness function $f_{o,s,a}$ a Darwinian ideal observer consists of a representational space X, and a perceptual channel P_x: W→X that maximizes the expected-fitness payout to the organism.*

Sin embargo, la evolución no suele producir canales perceptivos que optimicen el pago del fitness, sino únicamente *soluciones satisfactorias*. Ya sabemos que la evolución, más

que como un competente ingeniero, suele actuar como un aficionado al bricolaje. Entonces Hoffman introduce la definición de *Observador Darwiniano*, cambiando optimización por satisfacción.

> *Definition 9 Given a specific fitness function* $f_{o,s,a}$*, a Darwinian observer consists of a representational space X, and a perceptual channel* P_x*: W→X that has been shaped by natural selection as a satisficing solution to the problem of increasing expected-fitness payout to the organism.*

En las páginas siguientes Hoffman comenta someramente cómo evolucionarían los respectivos canales, llegando siempre a la conclusión de la baja probabilidad de que en la historia evolutiva se generara algún tipo de estrategia realista. Por ejemplo, analiza si sería más efectivo tener representaciones muy específicas o, por el contrario, de propósito general, llegando a la conclusión de que lo más probable sería una estrategia mixta.

2.4.3. EL BUCLE PERCEPCIÓN-DECISIÓN-ACCIÓN (PDA)

Pero la cuestión sigue en pie: ¿cómo un organismo cuyo canal perceptivo le transmite información que no guarda ningún homomorfismo o isomorfismo con el mundo objetivo puede tener interacciones exitosas con el mundo? ¿Cómo, por ejemplo, puede moverse en un entorno plagado de obstáculos sin percibir, objetivamente, dónde están ubicados tales objetos?

Según Hoffman, para que esto ocurra, deben darse tres condiciones:

1. Hay canales perceptivos estables.

2. Hay una regularidad en las consecuencias de nuestras acciones en el mundo.

3. Hay coherencia entre las percepciones y las acciones.

El planteamiento de Hoffman parte de que nuestros sistemas perceptivos han sido configurados, de manera crucial, por nuestros sistemas motores, de manera que manifiestan una total coherencia. El canal perceptivo informa sobre la función de pago para nuestro fitness, por lo que nuestro interfaz (el escritorio de PC) nos dice que podemos llevar a cabo una acción que será valiosa en términos evolutivos y, en consecuencia, nosotros decidimos hacerla. Pero pensemos además que lo que importa para conseguir fitness es ejecutar correctamente una determinada acción dado el estado tanto del organismo como del mundo, por lo que, en cualquier caso, nuestro sistema perceptivo, incluso si fuera realista, no tendría por qué informarnos del estado objetivo del mundo, sino del estado objetivo de las relaciones entre el mundo, mi organismo y la acción que realizo. Esta coherencia queda muy bien representada en lo que Hoffman denomina bucle PDA.

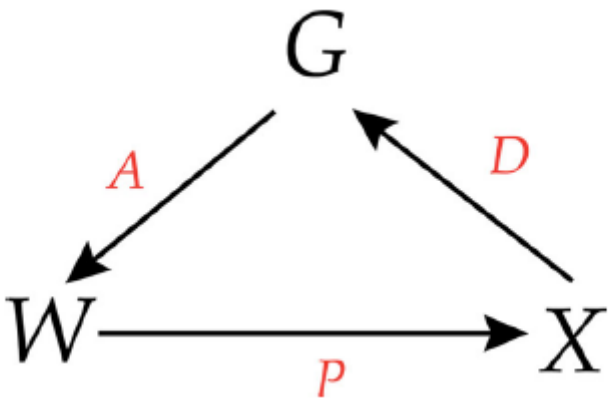


Fig.3. Bucle PDA. Sacado del original Hoffman, D. et al. “The Interface Theory of Percepción” (2016. p.17).

En la Fig.3 vemos los tres ejes (tres cadenas de Markov) que configuran el bucle PDA: Está el mundo objetivo (W) percibido (P) en una representación (X), desde donde se toma la decisión (D) en función de un espacio de acciones posibles (G) de realizar una acción (A). Es interesante que el bucle PDA puede aplicarse no solo a la percepción humana, sino a la de todos los organismos. Además, un organismo puede tener una

cantidad indefinida de estos bucles, incluso unos se pueden anidar dentro de otros. Según Hoffman, el bucle PDA es un poderoso marco abstracto para el modelado cognitivo.

2.4.4. LAS OBJECCIONES DEL MUNDO OBJETIVA E INTERSUBJETIMANTE MEDIBLE

Hoffman sigue peleando contra argumentos a favor de la objetividad del mundo. El *argumento del mundo medible* consiste que en que percibimos correctamente el mundo ya que podemos medirlo desde diferentes instrumentos de medición muy precisos, dándonos, a todos los observadores, los mismos resultados. A esta segunda parte lo llama *argumento del consenso*: todos los observadores concuerdan en las mismas medidas: el mundo es *intersubjetivo*.

Para Hoffman ambos argumentos fallan. El primero porque, en muchas ocasiones, nuestras mediciones no coinciden con nuestra percepción del mundo. Por ejemplo, si observamos el cielo, nada en nuestra percepción nos dice que el sol está cuatrocientas veces más lejos que la luna, o que Próxima Centauri esté 250.000 veces más lejos que el sol. Hoffman cita a Koenderink (2014), quien llegó a la conclusión de que el concepto de veracidad (*veridicality*), tan usado en estudios sobre la visión, es un concepto vacío que termina por ser un obstáculo para una correcta comprensión de la percepción. Y, en segundo lugar, aun cuando nuestras percepciones coindicen con nuestras mediciones, lo único que estamos haciendo es, simplemente, extender nuestras representaciones perceptivas. Por ejemplo, la noción de espacio euclídeo mediante la que solemos describir el mundo objetivo, surge de introducir suposiciones de simetría, como la traducción e invariancia de rotación. Las mismas unidades de medida que utilizamos para calcular la profundidad de un espacio (por ejemplo, el metro), son, precisamente, aquellas que no serían elegidas por la selección natural como ya vimos en los juegos evolutivos.

El *argumento del consenso* yerra igualmente: el acuerdo en cualquier afirmación no implica, en ningún momento, la veracidad de dicha afirmación. Podría darse el caso de que todos los seres humanos tuviésemos un sistema perceptivo que no captara la auténtica realidad pero que nos ofreciera, a todos, la misma visión falsa de la realidad. Hoffman ilustra jocosamente esta idea diciéndonos que las moscas están todas de acuerdo con que el estiércol es muy sabroso.

Otro argumento a favor del realismo lo dio Bertrand Russell (1912). Pensemos en una compañía de soldados que desfilan por un camino. Muchos individuos pueden observarles desde diferentes puntos de vista y, por tanto, verán diferentes escenas (verán a unos soldados más cerca, y, por lo tanto, más grandes, que a otros, mientras que a otros ni si quiera los verán), pero, sin embargo, ciertas cualidades estructurales de la compañía permanecerán invariables para todo observador (por ejemplo, su orden en filas, la dirección del desfile, etc.). Para Russell, las diferentes formas cambiantes podrían ser subjetivas o no verídicas, mientras que las estructuras invariantes para todo observador serían la auténtica realidad. Para Hoffman este argumento también es incorrecto, debido a que falla la mayor: que los observadores observen una invarianza estructural en una acción grupal no quiere decir que el mundo tenga, objetivamente, esta invarianza. El mundo no necesita tener la estructura que el observador percibe, sin importar lo compleja que ésta sea, o lo bien que puedan predecirse sus transformaciones en la medida en la que el observador actúa. Para probarlo Hoffman propone el Teorema de la Invención de la Simetría:

> ***Invention of Symmetry Theorem***. *Let an observer have at its disposal a group G of actions on the world W, such that its own perceptual space X is a G-set. This means that G acts on X via the kernel PA = P(A(g)) = ∫P(w,dx)A(g,dw), i.e., the action of A followed by that of P; moreover G acts on X by a transitive group*

> *action, so that G is a symmetry group of X. Moreover, let G act on W in such a way that the observer's perceptual channel mediates this action: P(g.w) = g.P(w), where the dot signifies the action of G on each set. Then,the perceptual experiences X of this observer will admit a structure with G as its group of symmetries.*

Nosotros percibimos el mundo como un espacio euclidiano en el que los objetos sufren transformaciones mientras ellos mismos, o lo observadores, cambian, por ejemplo, de posición. El *Teorema de la Invención de la Simetría* diría que las invarianzas y transformaciones coherentes y sistemáticas que parece mostrar el mundo no tienen por qué ser cualidades del propio mundo objetivo. Para que un individuo perciba esas cualidades (o simetrías) sólo hace falta que éstas sean coherentes con sus acciones, de modo que, por ejemplo, el espacio o el tiempo, podrían ser únicamente presupuestos del observador (Hoffman se basa aquí en los estudios de Terekhov y O'Regan de 2013) sobre la invención del espacio, quienes sostienen que las categorías euclidianas podrían aprenderse al interactuar con un mundo no euclidiano. En ese artículo en concreto explican como un organismo podría crear la noción de espacio, concretamente la de *desplazamiento rígido*, solo a partir de invariables sensoriomotores llamados cambios sensoriales "compensables").

Para la pregunta lanzada por *Edge* en 2017: *¿Qué concepto o término científico debería ser más conocido?*, Donald Hoffman respondió que el *Principio Holográfico*[2]. Es un principio de física cuántica que sostiene, a grandes rasgos, que toda la información contenida en un cierto volumen de espacio, se puede codificar sin pérdida solo conociendo toda la información del perímetro de esa región (Hooft, 2001; una explicación más extensa en Bousso, 2002 o Susskind, 1995). La descripción de dos dimensiones solo

[2] https://www.edge.org/response-detail/27026

requiere un grado discreto de libertad por área de Planck y, sin embargo, es lo suficientemente rica como para describir todos los fenómenos tridimensionales. La información que contiene un sistema no depende del volumen sino solo del área. A partir de ahí se infiere que nuestro mundo aparentemente tridimensional puede ser explicado, en su totalidad, como una proyección bidimensional similar a una imagen holográfica. Hoffman utiliza esta idea como claro ejemplo de que, aunque a todo el mundo nos parezca la tridimensionalidad del mundo como una invariante de Russell, no tiene por qué ser así: un mundo bidimensional podría contener toda la información de nuestro mundo. Es más, el espacio-tiempo como contenedor fundamental de la realidad queda desbancado hacia una realidad más profunda.

¿Y a qué debería deberse esa invención? ¿No sería más sencillo pensar que esas invarianzas representan un orden intrínseco al mundo? No porque, de nuevo, si lo que el organismo busca es satisfacer la función de fitness, su percepción del espacio será solo real en la medida en la que contribuya a satisfacer dicha función. Pero como ya vimos, el realismo no es una buena estrategia para optimizar el fitness, por lo que la evolución no la seleccionaría. En este sentido, pone como ejemplo Hoffman, nuestra percepción del espacio podría solo representar el costo, en términos de locomoción, que cuesta moverse en él.

2.4.5. ILUSIONES Y ALUCINACIONES

Hoffman, citando a Gregory (1997), admite la dificultad de definir una ilusión perceptiva, más cuando las explicaciones científicas van en la dirección de sostener que toda percepción es, en cierto sentido, una ilusión. Lo más fácil es decir que una ilusión es una discrepancia visual que se desvía de nuestras mediciones habituales. Sin embargo, desde la teoría de la interfaz, al considerar que toda percepción es una ilusión, se subraya la necesidad de una nueva teoría de la ilusión.

En vez de entender la ilusión como lo que se aleja de la medición verídica de la realidad, habría que entenderla como lo que se aleja de ser una guía para la adaptación biológica, es decir, como percepciones que perjudican el éxito en la obtención de fitness. Hoffman nos pone el ejemplo del cubo de Necker. En general, cuando contemplamos un dibujo tridimensional dibujado en un plano sufrimos la ilusión de pensar que el cubo tiene tres dimensiones cuando en realidad tiene solo dos. Pero esto es lo que nos diría una teoría realista. Desde la teoría de la interfaz, no diríamos que es una ilusión porque se aleja de la realidad, sino que es una ilusión porque no nos permite guiar el comportamiento adaptativo. Creer que un objeto tiene tres dimensiones y no dos podría perjudicar alguna de nuestras acciones para incrementar el fitness.

Hoffman nos pone otro ejemplo cambiando de modalidad sensorial de la vista al gusto. Existe una glucoproteína llamada miraculina que está presente en las bayas rojas de la *Richadella dulcifica*. Cuando alguien come dichas bayas, la miraculina produce durante aproximadamente una hora que el sabor ácido se convierta en dulce, de modo que, si comemos, por ejemplo, limones, éstos nos sabrán dulces. La teoría realista nos diría que el dulzor es una ilusión porque se aleja del sabor dulce real. Sin embargo, aquí habría algo raro porque ¿qué sentido tiene hablar del sabor real del algo? ¿Las moléculas tienen, objetivamente, algún sabor? De hecho, objetivamente, mi sensación de sabor es dulce y esa sensación sería indistinguible objetivamente de otra que muestre más sabor dulce ¿Por qué, entonces, podemos decir que ese dulzor es una ilusión?

La teoría de la interfaz no tendría ese problema: el dulzor sería una ilusión si perjudica la optimización de la función de fitness. Si, por ejemplo, ese dulzor me hiciera creer que estoy comiendo alimentos ricos en azúcares mientras, realmente, estoy comiendo alimentos poco nutritivos, el dulzor sería una auténtica ilusión.

3. LA TEORÍA DE LA CONSCIENCIA DE LA INTERFAZ

3.1. EL PROBLEMA MENTE-CUERPO

La posición de Hoffman con respecto al viejo problema mente-cuerpo parte de una atrevida premisa. Dadas la enorme cantidad de correlatos entre la actividad consciente y la actividad cerebral que la neurociencia nos ofrece, parece preocupante no disponer de una teoría concluyente de la consciencia. Lo habitual es responder que dicha actividad neuronal es la causante de la consciencia (cuando no se dice que actividad mental y neuronal son una sola y misma cosa). Suele mencionarse que la consciencia emerge de la actividad neuronal. Hoffman invierte la idea: *va a ser la actividad consciente la que cree actividad neuronal y no al contrario* (Hoffman, 2008). De hecho, la actividad consciente creará todas las propiedades objetivas del mundo físico. Para defender esta tesis va a basarse en dos teorías: la teoría del interfaz que hemos desarrollado y la teoría de los agentes conscientes que pasamos ahora a desarrollar.

Hoffman pone como ejemplos de correlaciones entre actividad consciente y base neuronal, diferentes evidencias centradas en las áreas visuales. Por ejemplo, el daño en el área cortical V1 se correlaciona con la pérdida de la visión consciente (Celesia *et al*, 1991), o el daño a las circunvoluciones linguales y fusiformes se correlaciona con la acromatopsia, una pérdida de sensación de color (Collins 1925, Critchley 1965), entre muchas otras (Hoffman, 2008: 88 y ss.). Dada esta cantidad de evidencia, parece darse un consenso en la comunidad científica sobre que esta actividad neuronal crea, genera o produce la actividad consciente. Además, han surgido, en número creciente, un buen número de teorías para explicar esto. Hoffman cita varias de ellas, como la teoría de identidad de tipos de Smart y Place, el funcionalismo reduccionista de Block y Fodor, el funcionalismo no reduccionista de Chalmers, la teoría representacionalista de Tye, y el naturalismo emergentista de Searle (para una introducción general pero muy rigurosa a

las principales teorías filosóficas sobre la mente véase Moya, 2011). Hoffman advierte que no va a tratar posiciones importantes como el emergentismo de Broad, el monismo anómalo de Davidson o la teoría de la superveniencia de Kim, debido a la brevedad de su artículo. Sin embargo, un punto que sí ve necesario tocar, es la falta de precisión de todas estas teorías al establecer la correlación. Por ejemplo, la teoría de la identidad simbólica no establece con claridad qué elementos de un estado neuronal corresponden al símbolo. Se hace patente que hay un enorme hueco en las relaciones entre estado físico y mental que necesita llenarse mediante una teoría genuinamente científica.

Los intentos ya se han puesto en marcha. Crick y Koch propusieron que ciertas oscilaciones neurales (concretamente de entre 35 y 5 Hz) dadas en la corteza cerebral eran la base de la experiencia consciente. También propusieron el *claustrum* como sede de la experiencia unificada de las experiencias mentales para solucionar el testarudo *binding problem*. Edelman y Tononi, igualmente, propusieron su *teoría de la reentrada*, sosteniendo que un grupo de neuronas puede contribuir directamente a la experiencia consciente solo si es parte de un grupo funcional distribuido que, a través de interacciones reentrantes en el sistema talamocortical, logra una alta integración durante cientos de milisegundos. También tenemos la *teoría del espacio global de trabajo* de Baars, los polémicos *microtúbulos* de Penrose y Hameroff o el *colapso de la superposición de plantillas de acción* de Stapp.

Sin embargo, todas estas teorías solo parecen ofrecernos indicios de lugares en donde buscar, sin que ninguna de ellas se acerque a la exigencia de precisión, rigor y capacidad predictiva, que toda teoría científica debe poseer. Por ejemplo, un mínimo de precisión exigible sería el que la teoría nos permitiera, a partir de las bases físicas, diferenciar dos experiencias conscientes diferentes como podría ser el olor de una flor del olor a pintura. Sin embargo, ninguna de las teorías antes expuestas permite nada

semejante. Esta lamentable situación ha llegado a algunos, como el filósofo británico Collin McGinn (1989) a llegar a negar la posibilidad de que los seres humanos seamos capaces de comprender nuestra propia consciencia. No obstante, la mayoría de los investigadores no llegan tan lejos y, sencillamente, sostienen que todavía hay que investigar más. Necesitamos más datos empíricos que sustenten nuevos avances teóricos.

Otra respuesta ha sido negar la misma existencia del problema mente-cuerpo. Esto se ha hecho en dos direcciones: o bien negando la existencia de la mente desde los clásicos materialismos eliminativos de Churchland o Dennett, o bien negando la existencia (o al menos cierta idea de) del cuerpo al que pueda reducirse la mente. Esta segunda postura la ha defendido Chomsky, quien ha argumentado que no existen una noción consistente de la relación entre cuerpo y mente desde que Newton introdujo la noción de fuerza a distancia, lo que imposibilita cualquier demarcación entre cuerpo y mente. Chomsky concluye que la consciencia es una propiedad de la materia organizada exactamente igual que, por ejemplo, el electromagnetismo.

La propuesta de Hoffman irá en una dirección completamente diferente: establecer una teoría de la consciencia no dualista pero matemáticamente rigurosa, que no asuma la tesis de Chomsky de que la consciencia surge de la materia organizada.

3.1.1. EL ERROR DE ENTENDER LA PERCEPCIÓN COMO UNA REPRESENTACIÓN FIEL

Hoffman distingue dos grandes grupos de teorías acerca de cómo percibimos la realidad: las indirectas y las directas.

Las *indirectas* afirman que el objetivo de la percepción es aproximarse lo más posible a las propiedades útiles del mundo físico objetivo. El mundo es objetivo en la medida en que se entiende que es independiente del sujeto. La información transducida

por los receptores sensoriales no es lo suficientemente rica para establecer la aproximación a las propiedades útiles del mundo, por lo que el sujeto debe inferirlas utilizando suposiciones restrictivas. Esta inferencia podría formularse en el marco matemático de la teoría de la regularización (Poggio *et al*., 1985) o en la inferencia bayesiana (Knill y Richards, 1996).

Las *directas* coinciden con las indirectas en el objetivo de la percepción, pero difieren en que la información tansducida es lo suficientemente rica para establecer la aproximación, de modo que no hace falta inferencia alguna.

Para Hoffman, lo interesante es que ambos enfoques coinciden acríticamente en que el objetivo de la percepción es intentar aprehender propiedades objetivas del mundo físico, lo que el mismo Hoffman llamará la hipótesis de la representación fiel (*Hypothesis of Faithful Depiction*: HFD). Y, precisamente, la teoría del interfaz va a negar taxativamente esta hipótesis.

3.1.2. LAS INTERFACES DE USUARIO COMO OCULTAMIENTO DE LA COMPLEJIDAD CAUSAL

Un icono cualquiera del escritorio de nuestra computadora no tiene ningún tipo de semejanza con la función que realiza, es más, sirve como *entorno amigable* para ocultar el complejísimo sistema de voltajes y magnetismos que hace, realmente, funcionar la máquina. El icono sirve para *ocular la complejidad causalidad* sin tener, él mismo, ningún poder causal. El icono solo es un conjunto de píxeles que no causan nada dentro de la cadena causal del proceso que se desencadena, es una simplificación que oculta una miríada de procesos causales.

El icono solo informa al usuario para que éste, al hacer clic con el ratón, active un cierto proceso causal, es decir, el icono forma un ciclo efectivo de información-acción sin

tener un poder causal directo en la computadora. El usuario puede operar como si el escritorio fuera la auténtica realidad sin que su conducta pierda efectividad. Una persona podría obrar como si detrás del ordenador no existiera *hardware* de ningún tipo y seguir utilizándolo eficazmente, sin diferencia ninguna con quien cree, acertadamente, en la existencia de *hardware*.

3.1.3. LA HIPÓTESIS DE LA INTERFAZ DE USUARIO MULTIMODAL (MUI)

Hoffman rechaza la HFD y, en su lugar, propone la percepción como una interfaz de usuario multimodal (*hypothesis of Multimode User Interface*, MUI). Mientras que la HFD propone cierta correspondencia entre el mundo y la experiencia perceptiva, la MUI no dice absolutamente nada de la ontología del mundo real, no tiene, por decirlo en términos de Quine, ningún compromiso ontológico. De hecho, hacerse la pregunta sobre la verosimilitud entre experiencia sensorial y mundo objetivo sería caer en un error categorial, sería no formular bien la pregunta. La cuestión adecuada es si la experiencia sensorial me transmite información útil.

Según la MUI los objetos del mundo cotidiano (sillas, mesas, árboles…) no son de acceso público. Hoffman pone el ejemplo (2008, p. 97) de dar un vaso de agua a otra persona. La MUI sostendría que el vaso de agua que yo doy y el que la otra persona coge, son numéricamente diferentes. Habría dos vasos y no uno, y si un tercer observador entrara en escena, habría tres vasos. Esta idea parece absurda y, de hecho, filósofos como Searle argumentan a favor del realismo perceptivo sosteniendo que si podemos comunicarnos es porque nuestros significados son públicos (al menos en cierta parte) y que las instancias a las que hacen referencia también lo son, es decir, como mínimo compartimos el acceso perceptivo a ciertos objetos.

Hoffman critica este argumento mediante un contraejemplo: pongamos que Bob y Tom están jugando a un videojuego de realidad virtual que simula un partido de tenis.

Cada uno está en el salón de su casa conectado a los clásicos dispositivos de realidad virtual (gafas, guantes, mandos…), pero el videojuego simula con mucho realismo una auténtica pista de tenis. Ambos podrían estar de acuerdo con la afirmación "La raqueta ha dado a la bola", mientras que el objeto "bola" no es numéricamente el mismo para los dos. Cada una de sus videoconsolas genera un mundo virtual en los que los píxeles que representan a la bola son completamente diferentes. Hay dos bolas, una proyectada en las gafas de realidad virtual de Bob y otra en la de Tom. Conclusión: podemos establecer comunicación sin compartir acceso perceptivo a ningún objeto. Lo único que hace falta es que Bob y Tom mantengan una cierta coordinación. Searle cree que para que se de esta coordinación hace falta que el objeto sea público, pero como vemos, no es necesario.

Según la MUI los objetos solo existen en tanto que son percibidos. Una silla solo existe en la medida en la que la percibo, dejando de existir en el mismo instante que cierro los ojos. Es una tesis que, aparentemente, parece absurda y fácilmente refutable. Un argumento trivial sería decir que puedo dejar de mirarla y tocarla con la mano para comprobar que sigue existiendo. O, igualmente, cuando vuelvo a abrir los ojos la silla sigue exactamente donde estaba, lo cual parece un buen indicio para sostener que la silla seguía existiendo mientras no la veía. Pero, la objeción se refuta de nuevo desde el tenis virtual: Bob y Tom pueden sostener que la pelota de tenis sigue existiendo cuando no la miran, cuando no es así: la pelota de tenis virtual solo existe cuando Bob o Tom gira el cuello y las gafas de realidad virtual apuntan a dónde debería estar la pelota siguiendo las expectativas virtuales del simulador.

Una segunda objeción: cuando ves un tren que se dirige hacia ti a toda velocidad, según la teoría de la interfaz solo estás viendo un *icono*, algo que no tiene nada que ver con la realidad. Entonces, ¿por qué te apartas? ¿Por qué no te mantienes en tu sitio dado que el hecho de que te atropelle un tren no será verdaderamente real? Hoffman responde

(p.98-99) que esta objeción viene de confundir tomarse algo *literalmente* de tomarse algo *en serio*. Cuando yo me aparto para que no me pille el tren me estoy tomando *en serio* el icono "tren" pero no me lo estoy tomando *literalmente*. Seguramente que en el mundo real no exista nada similar a lo que yo percibo en mi mente, pero si tengo un icono que me advierte que debo apartarme, conviene tomarme en serio esa advertencia y apartarse, porque los que hicieron caso a sus "iconos de advertencia" sobrevivieron en la evolución, mientras que los que no lo hicieron se extinguieron.

Una tercera objeción: yo habitualmente puedo pensar lo que me plazca. Cuando con mi imaginación pienso en un muro, si lo deseo puedo atravesarlo ¿Por qué no ocurre eso con los iconos de mi interfaz? La respuesta de Hoffman es que, evidentemente, no podemos hacer lo que queramos con nuestros objetos mentales. Pone el ejemplo de un cubo de Necker. Al observar un cubo de Necker como el de la fig.4 podemos ver que el vértice B forma parte del lado más cercano a nosotros o, por el contrario, podemos transformar la perspectiva y ver que el vértice A pertenece al lado más cercano. La clave está en que nos cuesta cambiar la perspectiva, de modo que no podemos hacer lo que deseemos con el cubo. Además, si miramos bien la imagen, realmente no hay ningún cubo. Podríamos decir que solo hay ocho círculos grises atravesados por franjas del mismo color que el fondo, pero a nuestra mente le cuesta muchísimo no ver un cubo. El hecho de que los objetos mentales sean una construcción de la evolución no implica que sean maleables a nuestra voluntad. Es más, los objetos mentales no son una mera construcción que surja de la propia mente o de un entorno puramente subjetivo sino más bien lo contrario: los objetos mentales surgen de la interacción de nuestro organismo con la realidad y según una serie de reglas probabilísticas, y eso hace que no sean arbitrarios ni moldeables, sino más bien *obstinados*.

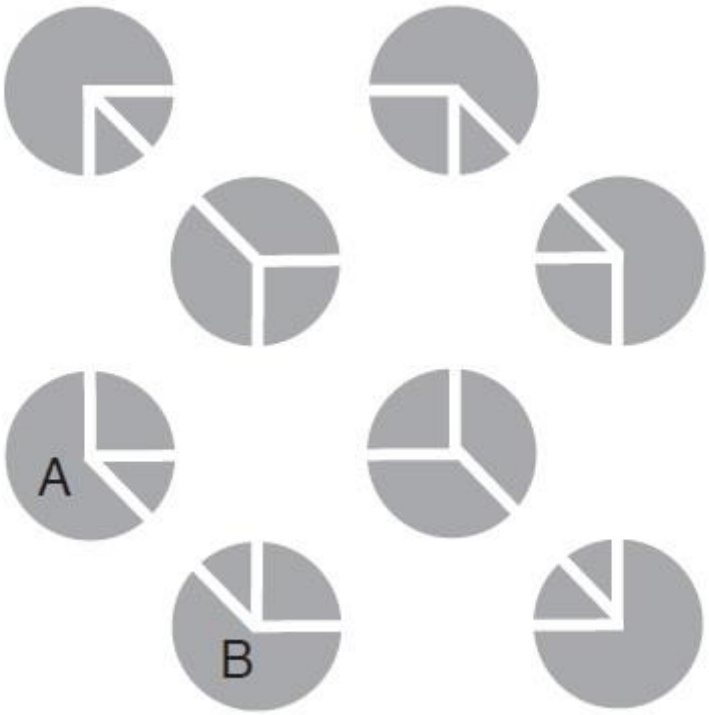


Fig. 4. Sacado del original “Conscious Realism and the Mind-Body Problem” pág. 99.

Cuarta objeción: Los físicos llevan desde hace mucho tiempo diciéndonos lo mismo. La solidez con la que solemos ver los objetos físicos es solo una ilusión, ya que lo que realmente hay es solo espacio vacío en el que revolotean leptones y quarks ¿Qué habría de nuevo en la MUI? Para Hoffman en esto no hay ningún problema, es más, lo único que se ve es que la MUI concuerda perfectamente con la física de partículas. Según la interpretación de Copenhague de la física cuántica, las partículas elementales tienen propiedades dinámicas que solo adquieren valores reales en el acto de la observación, afirmación perfectamente compatible con la teoría de la interfaz.

Quinta objeción: La MUI se parece mucho al idealismo, una filosofía que tuvo una inmensa influencia en el pasado, pero que hoy en día no goza de demasiada aceptación. Para Hoffman esto es un error porque la MUI y el idealismo no son lo mismo. La MUI no afirma que todo lo que existe son ideas o percepciones subjetivas, sino que

nuestras percepciones no tienen por qué parecerse al mundo objetivo. La MUI, al contrario del idealismo, sostiene que existe un mundo real, objetivo, externo a nosotros.

Sexta objeción: La MUI sostiene que lo que percibimos no es real sino creado por una interfaz que existe entre nosotros y el mundo real ¿No parece algo demasiado rebuscado, inverosímil? ¿De verdad que *el mundo real no es real*? Para Hoffman esta objeción viene del error de no utilizar con precisión una palabra tan ambigua como "realidad". Que algo sea real puede significar que ese algo existe, o también que existe con independencia de cualquier observador. Un dolor de cabeza es real en el primer sentido, pero no en el segundo. Así, la MUI no dice que todo lo que observamos sea irreal en el sentido de ser un "velo místico" entre la realidad y el observador, sino solo que todo lo que observamos es irreal solo en el segundo sentido, es decir, solo como independiente del observador. Igual puede decirse con respecto al significado de la palabra mundo: ¿para la MUI el mundo no es real? Hoffman responde de la misma forma: que el mundo sea real puede significar tanto que existe como que existe con independencia de nosotros. La MUI no dice en ningún momento que el mundo no existe, solo dice que, al contrario que la HFD, es muy probable que nuestras percepciones del mundo no se parezcan demasiado al propio mundo.

Séptima objeción: La MUI es falsa porque, sencillamente, los iconos de las interfaces de usuario de las computadoras sí que se asemejan a los objetos reales. Por ejemplo, la papelera de reciclaje donde "tiro a la basura" los archivos que ya no necesito, se asemeja mucho a una papelera real (su icono es el dibujo de una papelera), y se usa de un modo muy similar a cuando tiramos la basura en el mundo real. Hoffman responde: estos íconos no imitan los diodos, resistencias, voltajes y campos magnéticos que representan dentro de la computadora. Los íconos ocultan a propósito toda esta complejidad, de modo que los usuarios de computadoras pueden continuar con su trabajo.

Borrar un archivo, realmente, no tiene nada que ver con tirar la basura. El programador de la interfaz de nuestro ordenador utilizó la metáfora de la papelera como icono para borrar archivo, sencillamente, porque de esa forma se nos hace muy amigable y sencillo el uso del ordenador.

Octava objeción: la MUI parece conceder cierta semejanza entre sus iconos y el mundo real en el sentido de reconocer que el mundo *causa* los iconos. Así, la pelota de tenis virtual se comporta causalmente, al menos de forma aproximada, como una pelota de tenis real. Sin embargo, responde Hoffman, de nuevo estamos confundiendo la relación que se establece: lo que debería tener una semejanza sería la pelota de tenis virtual y el funcionamiento del computador; y, evidentemente, la pelota virtual no tiene ninguna semejanza ni gráfica ni causal ni estructural con el funcionamiento interno del ordenador.

Novena objeción: se puede cuestionar la metáfora completa de la realidad virtual. Parece difícil de aceptar que vivamos en una especie de realidad súper virtual similar al de la película *Matrix*. Además, no parece existir suficiente evidencia científica que respalde tan atrevida tesis. Hoffman recurre a la mecánica cuántica para buscar esa base científica. Desde la interpretación de Copenhague, a partir del entrelazamiento cuántico y las violaciones de las desigualdades de Bell, se puede negar el realismo local y, particularmente, sostener que ciertas propiedades físicas de un sistema no existen hasta que son observadas. No obstante, admite Hoffman, hay otras interpretaciones posibles. Por ejemplo, los defensores de la decoherencia no la aceptarían, así como que la misma interpretación de Copenhague solo acepta la dependencia del observador para propiedades microscópicas, no siendo válida para macroscópicas.

3.1.4. REALISMO CONSCIENTE

Para solucionar el problema mente-cuerpo, Hoffman va a unir a su teoría de la interfaz, lo que él denomina como *realismo consciente*. Si bien la teoría de la interfaz no

dice nada acerca de la ontología del mundo objetivo, el realismo consciente sí lo hace, y muy rotundamente: la realidad está compuesta, en su totalidad, por agentes conscientes. Estaríamos hablando de un monismo no fiscalista que niega el materialismo tradicional: la realidad no está compuesta de partículas físicas inertes. Las partículas y campos físicos son solo iconos en las MUI de los agentes conscientes, no siendo en sí mismos habitantes del mundo objetivo. Hoffman invierte el orden de la concepción materialista tradicional. Para el materialismo, la materia es lo primero y la consciencia surge de ella, mientras que para el realismo consciente la consciencia es lo primero y la existencia de la materia depende de ella.

Según el realismo consciente, cuando observo una mesa no es que mi mente este interactuando con un objeto material, sino que mi consciencia está interactuando con otros agentes conscientes, cuyo resultado es el icono o símbolo "mesa". Esto diferencia al realismo consciente del panpsiquismo. Este segundo diría que todos los objetos del universo son conscientes. Así, la mesa que he observado sería un ser consciente. Pero el realismo consciente no diría que la mesa es una entidad consciente, sino tan solo un símbolo, un icono, una representación mental. El realismo consciente no identifica los agentes conscientes con los objetos físicos.

De la misma manera, el realismo consciente no es el igual que el idealismo trascendental de Kant. Hoffman reconoce que la exégesis de la obra kantiana es muy difícil, pero se basará aquí en la de Strawson (Strawson, 1966: 38). Según ésta, Kant defiende que la realidad es suprasensible y, por lo tanto, incognoscible. Solo podemos conocer la realidad en tanto que fenómeno, es decir, cuando a la realidad se le han añadido las condiciones a priori tanto de la sensibilidad como del entendimiento. La *cosa-en-sí*, la *realidad nouménica* es completamente inaccesible, por lo que no puede hacerse ninguna ciencia acerca de ella. El realismo consciente, por el contrario, sí que afirma que

es posible una ciencia de la realidad, ya que la realidad está formada por agentes conscientes a los cuales tenemos acceso, y podemos establecer un modelo matemático de dichos agentes y de sus interacciones. Hoffman acepta que cierta interpretación de la filosofía kantiana, a saber, la más idealista, concuerda con su teoría, pero reconoce que esa interpretación es controversial y no es la única aceptable (p.104).

La noción fundamental del realismo consciente es la noción de agente consciente, Hoffman sostiene que juega el mismo papel para la consciencia que el concepto de Máquina de Turing para la computación. Esto hace que su definición tenga un alcance muy amplio, aunque Hoffman va a enumerar una serie de sus cualidades esenciales:

1. El agente consciente no tiene por qué ser, necesariamente, una persona. Todas las personas son agentes conscientes, pero no a la inversa: no todos los agentes conscientes son personas.

2. De la misma forma las experiencias de un agente consciente no tienen que ser parte de las clásicas modalidades sensoriales propias de los seres humanos. Podrían, perfectamente, existir modalidades sensoriales. Totalmente ajenas y desconocidas para el hombre.

3. La dinámica de los agentes conscientes no tiene lugar, en general, dentro del espacio-tiempo ordinario de cuatro dimensiones. Se lleva a cabo en espacios de estado de observador consciente, y para estos espacios la noción de dimensión podría no estar bien definida. Ciertos agentes podrían emplear el espacio-tiempo, pero, como decimos, no es necesario.

En “The Origin of Time in Consciousness Agents”, se nos ofrece una definición de agente consciente basada en el bucle PDA (Hoffman, 2014. p. 507 y ss.). Como vemos en la fig. 5, el mundo está representado por un espacio mensurable W. Un agente

consciente comprende los seis componentes restantes del diagrama. Los mapas P, D y A pueden considerarse canales de comunicación. X es el espacio medible de las experiencias conscientes del agente, y G es el espacio mensurable de sus posibles acciones. Un número entero N cuenta las experiencias sucesivas del agente.

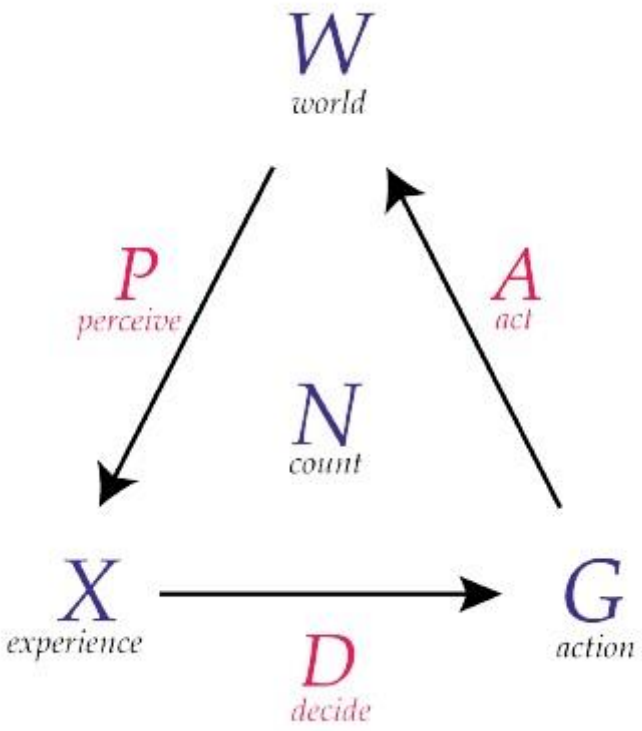


Fig. 5. Agente consciente. En “The Origin of Time in Consciousness Agents” p. 508.

Entonces, la definición formal de agente consciente sería:

Un agente consciente C, es a tupla de seis elementos tal que:

C = ((X, X), (G, G), P, D, A, N)), (1)

donde:

(1) (X, X) and (G, G) son espacios mensurables;

(2) P: W × X → [0,1], D: X × G → [0,1], A: G × W → [0,1] son núcleos de Markov; y

(3) N es un número entero.

Hoffman también desarrolla una dinámica de interacciones entre agentes conscientes que pueden ser representados mediante redes (o más bien pseudográficos). En la Fig. 6 tenemos una representación de la interacción entre dos agentes conscientes. En ella cada agente es parte del mundo W del otro agente. La parte inferior del diagrama representa al agente consciente C_1 y la parte superior al agente consciente C_2.

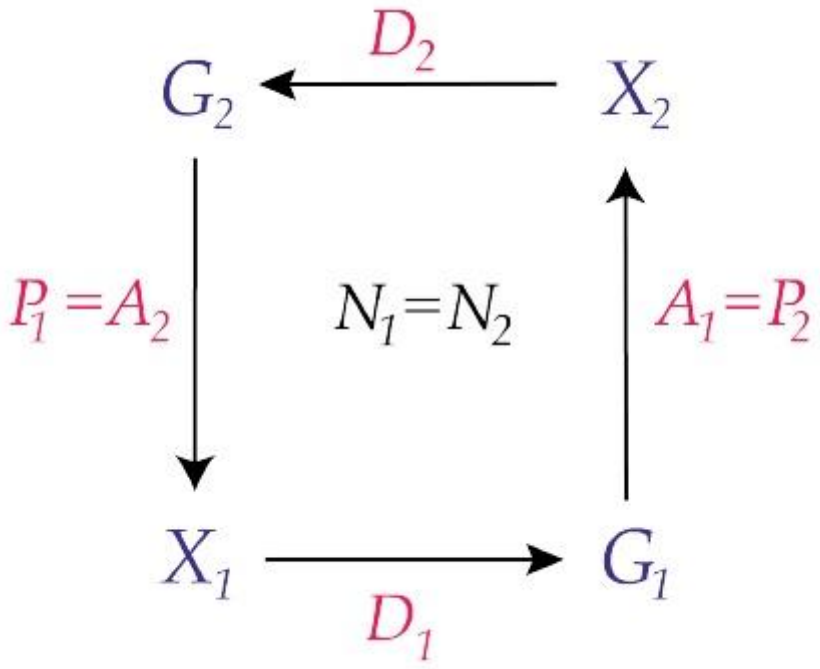


Fig. 6. Pseudográfico de la interacción entre dos agentes conscientes. En "The Origin of Time in Consciousness Agents" p. 509.

Hoffman subraya que ciertas combinaciones de agentes conscientes podrían llegar a tener la misma capacidad de cálculo de una Máquina Universal de Turing (Hoffman & Prakash, 2014), lo cual ofrece un riguroso modelo matemático tanto de agente consciente como de sus interacciones.

3.4.5. EL PROBLEMA MENTE-CUERPO

Hoffman reconoce que no existe ninguna teoría científica que nos sirva de base para afrontar el problema mente-cuerpo, así que él lo afrontará desde su MUI y su realismo consciente. Si, desde el fisicalismo lo que se trata es de describir cómo de la materia puede surgir consciencia, el camino del realismo consciente es inverso: hay que describir cómo los agentes conscientes construyen los objetos materiales.

Sin embargo, sí que tenemos ciertas teorías matemáticas que describen cómo un agente consciente construye formas visuales, colores, texturas y movimientos de objetos (Hoffman, 1998; Palmer, 1999; o Ullman, 1979, entre muchos otros). Hoffman acepta que los autores de todas estas teorías son, casi sin excepción, fisicalistas que aceptarían la HFD y que piensan que la mente reconstruye la realidad lo más fielmente posible. No obstante, sus aportaciones pueden reinterpretarse, sin ningún problema, como especificaciones de un método de construcción de objetos, y no de reconstrucción. Matemáticamente no hay que hacer ningún cambio para hacerlas compatibles con la MUI. Por ejemplo, Piaget afirma que los niños de alrededor de nueve meses de edad, aprenden la noción de permanencia, la creencia de que los objetos físicos siguen existiendo incluso cuando no se les mira (Piaget, 2013). Para la MUI la permanencia es una ilusión, un icono útil que el niño crea a partir de la interacción con otros agentes conscientes. Así, entre la afirmación de Piaget y la MUI habría una compatibilidad total.

El enfoque fisicalista tiene problemas para enfrentarse al problema de la causalidad: ¿cómo un estado mental puede causar un estado físico? Y si los estados mentales son estados físicos, ¿cómo es posible que tengan propiedades, aparentemente, tan diferentes a los objetos físicos? ¿Por qué no puedo ver, tocar, medir, etc. mis estados mentales? Una solución estaría en el funcionalismo, que afirmaría que los estados mentales son idénticos a estados funcionales, teniendo estos segundos plenas capacidades

causales. Sin embargo, según Hoffman, el funcionalismo ha quedado refutado por el "teorema de la aleatorización" (Hoffman, 2006), que afirma que los estados mentales y los estados funcionales no son numéricamente idénticos. Otra alternativa estaría en el epifenomenalismo, que afirmaría que, si bien los estados físicos causan estados mentales, los estados mentales no causan estados físicos. Sin embargo, esta idea choca con nuestra intuición de que mi mente causa mi conducta.

Hoffman sostiene que una mejor alternativa es lo que él denomina *epifisicalismo*: los agentes conscientes son el único *locus* de la causalidad, y dichos agentes son los que construyen los objetos físicos como elementos de sus MUI. Por el contrario, los objetos físicos no tienen poderes causales entre ellos (ni ningún otro poder causal), solo informan a los agentes conscientes para que éstos tomen sus decisiones. Cuando la raqueta virtual golpea la pelota virtual, la raqueta no causa realmente el movimiento de la pelota. La causalidad ocurriría a otro nivel: el del hardware del ordenador. Esta idea también implicaría que el cerebro no es la causa de la conducta, ya que el cerebro no deja de ser un icono más de la MUI. De hecho, ni siquiera yo soy el causante de mi conducta ya que el yo no es más que una compleja jerarquía de agentes conscientes, al que Hoffman llama *instanciación* ¿Significa esto que debemos detener la investigación sobre los correlatos neurales de la consciencia? De ninguna manera. Si queremos comprender la compleja jerarquía de agentes conscientes en instancias humanas, debemos usar los datos que proporcionan nuestras MUI.

Una objeción al realismo consciente por parte del fisicalismo sería la siguiente: si el mundo físico es inaccesible y está totalmente fuera del alcance de la mente, ¿dónde está esa interfaz de usuario? ¿En qué "pantalla mental"? ¿De qué está hecho su contenido? Estaríamos atrapados en el famoso problema del teatro cartesiano propuesto por Dennett. Según Hoffman, la clave de la objeción reside en el concepto de "mundo físico", ya que

en ella se está asumiendo una ontología fisicalista. Si se parte de la premisa de que todo es físico, la objeción tiene pleno sentido. Sin embargo, si partimos desde el realismo consciente, sosteniendo que los agentes conscientes son los contenidos fundamentales de la realidad, la cuestión de dónde están los contenidos mentales carece de sentido, ya que la materia o el mismo espacio-tiempo son componentes de los propios agentes conscientes.

Empero, esto lleva a más problemas. La consciencia se "expresa" mediante lo que se denominan *modalidades sensoriales*, las que, tradicionalmente se consideran irreductibles unas a otras. Por ejemplo, el color no puede explicarse en términos de sabor o de sonido (a no ser, claro está, que uno tenga sinestesia). La cuestión sería: ¿de dónde surgen estas modalidades sensoriales? ¿Serían los componentes últimos del universo? Para Hoffman cabría preguntarse incluso si esto nos llevaría a una nueva forma de misticismo. No, se responde a sí mismo, siempre que podamos establecer un modelo matemático de los agentes conscientes, de sus dinámicas y sus interacciones y, a partir de él, podamos hacer predicciones. De lo que se trata es de convertir el realismo consciente en una teoría científica.

Hoffman también enfrenta su teoría al problema del solipsismo (él lo denomina *cárcel epistémica*). Cuando observamos (creamos un icono) un objeto inerte y, después, observamos un agente consciente (otra persona, por ejemplo), realmente, no podemos diferenciarlos. Cuando observo a una persona solo veo sus rasgos faciales que, en el fondo, no son más que iconos en mi interfaz, exactamente igual que el color o la forma de una copa de vino que veo frente a mí. Yo no puedo acceder a la consciencia de los otros ¿Cómo escapar de esta *cárcel epistémica*? La respuesta de Hoffman es que nuestra interfaz está limitada a solo tener acceso a ciertas características de la realidad y no a otras. Así, no podemos acceder a otras consciencias. Sin embargo, a partir de conductas

o rasgos externos podemos inferir la existencia de otras consciencias. Si yo veo a alguien llorando, solo percibo su icono de interfaz, pero a partir de éste puedo inferir que hay una consciencia que siente tristeza.

3.4.6. OBJECIONES DESDE LA TEORÍA DE LA EVOLUCIÓN

Hoffman responde a una nueva objeción. Según sabemos el universo existió muchísimo antes de la aparición de la consciencia, la cual apareció en un momento concreto a lo largo de la evolución de los seres vivos. Si antes dijimos que la consciencia era el contenido fundamental del universo del cual se deriva la materia, debería darse temporalmente antes que la materia. Hoffman ofrece tres respuestas a esta objeción:

1. Aunque la teoría de la evolución ha sido interpretada, casi exclusivamente, en términos de una ontología fisicalista, los modelos matemáticos de la evolución no requieren ser fisicalistas. Pueden aplicarse perfectamente bien al realismo consciente. Por ejemplo, la teoría de juegos aplicada a la evolución (Maynard-Smith, 1982) encaja con él de una forma muy natural. Los sistemas de agentes conscientes pueden experimentar una evolución estocástica, y los agentes conscientes pueden sintetizarse o destruirse en el proceso (Bennett et al., 1989, 2002). En el fondo, lo que se dice, es que la teoría de la evolución no tiene ningún compromiso ontológico con ninguna teoría. Pero Hoffman hace un llamamiento a todos los realistas conscientes: hay que elaborar modelos matemáticos que partan de la consciencia, teniendo las leyes naturales como derivadas o secundarias de ella. Las leyes naturales han de entenderse como proyecciones de una concepción más sofisticada y profunda de la realidad.

2. No es cierto que la consciencia haya llegado más tarde en la historia del universo. La consciencia siempre ha sido primera y primaria. Según Hoffman nuestro error consiste en cometer la *falacia de la reificación*: suponer que los iconos que percibimos son objetos independientes de nosotros y fundamentales en el universo. Acatamos muy bien esta

falacia porque nuestra MUI guía de ese modo nuestro comportamiento de forma, evolutivamente, muy exitosa. Además, construimos estas ilusiones desde nuestra más tierna infancia y no es fácil prescindir de ellas. Por ejemplo, la ilusión de permanencia de los objetos comienza a los nueve meses de edad y, desde luego, no es nada fácil librarse de ella.

3. La teoría evolutiva estándar socava la idea de que percibamos la realidad tal y como es. La selección natural extingue las especies que no guían su comportamiento de manera útil para su supervivencia. (Bartley y Rarnitzky, 1987). El objetivo de los sistemas perceptivos es la supervivencia y no el conocimiento verdadero de la realidad. Por eso la teoría de la interfaz encaja mejor con la evolución que las teorías realistas. Además, el teorema de aleatoriedad (Hoffman, 2006) prueba que las propiedades de las experiencias conscientes no son numéricamente idénticas a las propiedades funcionales. Si, desde la teoría de la evolución decimos que todas las propiedades de la mente han de ser propiedades funcionales (cuya función es, a saber, aumentar el *fitness* de su poseedor), la teoría de la evolución no puede explicar la consciencia desde una perspectiva puramente funcionalista. Si bien, Hoffman reconoce que un funcionalista no reductivo podría salvar la objeción diciendo que los elementos de la consciencia que no son estrictamente adaptaciones, podrían, simplemente, ser el resultado *co-lateral* de adaptaciones; el realismo consciente no tiene que lidiar con tales problemas.

Otra gran objeción desde la teoría evolutiva, sería la de poner en duda los mismos criterios de utilidad en los que se basa la MUI para construir su interfaz. Si no tenemos acceso a la realidad de ninguna manera, ¿en qué nos basamos para decir que algo es útil o no? ¿Útil para qué? Hoffman responde someramente que no es que no tengamos acceso al mundo real, solo que nuestras percepciones no se asemejan a este mundo real.

Una tercera objeción diría así: cuando estoy observando una mesa parece que la mesa que observa un segundo observador, el coherente con la mía ¿Cómo podemos saber eso? Hoffman comenta que un lector irónico podría preguntar si estamos usando el mismo sistema operativo. Para responder a esto hay que tener en cuenta que la MUI no necesita que todas las MUI tengan la misma interfaz ni siquiera que las interfaces sean funcionalmente idénticos. La evolución sugiere que las interfaces sean similares dada una misma especie, pero también sugiere que sean ligeramente diferentes, ya que las variaciones aleatorias son esenciales en la evolución. De la misma forma, el ya citado teorema de la aleatoriedad demuestra que la identidad funcional no implica igualdad de nuestras experiencias conscientes.

4. REVISIÓN CRÍTICA Y CONCLUSIONES

4.1. LA PARADOJA DE EPIMÉNIDES

Una paradoja insalvable para toda teoría que afirme que la totalidad de lo real es falsa o ficticia es la famosa paradoja del mentiroso o de Epiménides. En su forma tradicional, se formularía así:

Epiménides, un cretense, afirma que todos los cretenses son unos mentirosos.

Si Epiménides dice la verdad y todos los cretenses son unos mentirosos, como Epiménides es un cretense no podría estar diciendo una verdad. Si, por el contrario, Epiménides miente, todos los cretenses dirían la verdad por lo que Epiménides, al ser cretense, debería estar diciendo la verdad.

La teoría de la interfaz de Hoffman sostiene que toda nuestra realidad es una ficción útil definida por la función de satisfacción de fitness en un entorno de evolución biológica. Por consiguiente, todo el *lenguaje* con el que expliquemos el funcionamiento de esa realidad será una explicación que solo puede partir de la ficción útil que es nuestra

realidad. Por tanto, el lenguaje para explicar la realidad no explicará el auténtico funcionamiento de la realidad. Dada la postura de Hoffman, no hay ninguna razón para sostener que podamos tener un *acceso privilegiado* a una realidad sorteando nuestra interfaz. Si solo podemos conocer la realidad a través de un interfaz, la propia teoría de la interfaz será también un conjunto de iconos o ficciones útiles, y no la realidad misma.

Hoffman debería dar alguna explicación, de cómo *él no es un cretense* y, de que se puede *salir del escritorio* y ver el auténtico funcionamiento de la cognición.

4.2. IRREDUCTIBLES QUALIA

La teoría de Hoffman no resuelve, en ningún aspecto, el problema de la especificidad y necesidad de la consciencia fenoménica. En “Objects of Consciousness” (Hoffman & Prakash, 2014, p. 14), Hoffman encabeza la lista de críticas posibles a su teoría, objetando que si lo que afirma para la consciencia también es aplicable para seres inconscientes, realmente, no se está diciendo nada de la consciencia. Hoffman responde que, aunque lo que se afirma sobre la consciencia pudiese aplicarse a seres u objetos inconscientes, eso no quita para que pueda aplicarse a los conscientes. Eso es tirar balones fuera. Si lo que dice de la consciencia se puede decir de la inconsciencia, realmente, no se estaría diciendo lo que tiene la consciencia de específico, por lo que, realmente, no se estaría definiendo consciencia. Y en esto Hoffman sí que falla, ya que él mismo consideraba lamentable que las actuales teorías sobre la consciencia no puedan diferenciar consciencia de inconsciencia, mientras que su propia teoría tampoco lo consigue.

4.3. ESCAPANDO DEL TEATRO CARTESIANO

Dennett propuso el famoso problema del teatro cartesiano (Dennett, 2017. p.101) que, haciendo referencia a la mencionada glándula pineal cartesiana, pone en aprietos muchas teorías de la mente actuales. Para Descartes, la glándula pineal era como una especie de pequeño observador (un *homúnculo*) que veía la realidad proyectada para él en una especie de "teatro mental". Esta idea sería absurda por dos razones:

1. Parece absurdo que nuestra cognición repita lo que hay en el exterior *otra vez*, dentro de nuestra mente ¿Por qué reconstruir de nuevo la realidad que está ahí fuera?

2. Si tenemos un homúnculo que observa la realidad como si estuviera en el teatro, a su vez, dentro de su mente debería haber un nuevo homúnculo, y así, sucesivamente *ad infinitum*.

La teoría de la interfaz de Hoffman podría parecer, en principio, que vuelve a caer en el problema del teatro cartesiano. La metáfora de que nuestra realidad es el escritorio de un PC apunta, desde luego, a un nuevo teatro. Sin embargo, no es así:

1. Nuestra cognición no repite lo que hay en el exterior de nuevo, sino que lo transforma en un *entorno amigable* reduciendo la complejidad. Es decir, la información se simplifica, no se repite.

2. Podríamos objetar que esa información se simplifica, precisamente, para que pueda ser utilizada por un torpe homúnculo, pero podemos esquivar esta crítica. Podemos sostener que nuestra mente es un conjunto de módulos funcionales encargados de las diferentes tareas cognitivas. Uno, o varios, de estos módulos serían los encargados de tomar decisiones en virtud de la información recibida. No sería absurdo pensar que la información se simplifica para facilitar su

utilización por los módulos encargados de tomar decisiones. Estos módulos no es que necesiten observar de nuevo la información como en un teatro, es que, sencillamente, utilizan la información para tareas y, en este sentido, una *información cocinada* para ser más operativa tiene todo el sentido del mundo.

4.4. PERO ENCERRADOS EN LA CÁRCEL EPISTÉMICA DEL SOLIPSISMO

Como no podría ser de otra manera, la teoría de Hoffman no puede escapar del clásico solipsismo cartesiano. Como ya vimos, solo percibimos iconos, que serán de la misma naturaleza mental para seres inertes que para otras consciencias. Entonces, no hay forma de saber si existen más consciencias a parte de la mía, ya que no tengo ningún tipo de acceso directo a la consciencia de los otros. Podría ser que yo fuera la única consciencia existente en el universo.

La respuesta de Hoffman consistía en sostener en que, a partir de la observación de rasgos y conductas externas de los objetos, podemos inferir si son conscientes o no. Efectivamente, así es como lo hacemos en la realidad, pero esto no garantiza que no podamos estar sujetos a cierto engaño. Cuando en la televisión veo el rostro de un hombre llorando, puedo inferir que ese hombre tiene consciencia, pero estaría cometiendo un error, porque detrás de la pantalla de mi televisor no existe ninguna mente. Los píxeles de la pantalla han creado una ilusión que ha engañado a mi sistema inferencial. Podríamos, tal y como ya planteó Descartes, estar soñando y creer que lo que soñamos es la auténtica realidad o, tal y como lo plantea en la actualidad Nick Bostrom (2003), estar viviendo en una simulación por ordenador.

No obstante, la respuesta de Hoffman nos parece suficiente. Es cierto que no podemos escapar absolutamente del solipsismo ya que es, prácticamente, irrefutable. Sin embargo, que algo sea irrefutable no quiere decir que sea cierto. Viendo que los demás seres humanos que viven a mi alrededor se comportan como si tuvieran mente, parece

razonable inferir, tal y como sostiene Hoffman, que la tienen, ya que la idea contraria, aunque irrefutable, nos parece proco probable.

No podemos saber si vivimos o no en una simulación, pero parece poco verosímil que vivamos en una. En este caso la carga de la prueba estaría en quienes defienden el argumento de la simulación, y a día de hoy no parece más que una arriesgadísima hipótesis sin la suficiente base. Por ejemplo, el argumento de Bostrom partiría de la premisa de que toda la mente humana es transferible a un ordenador, es decir, se acepta sin restricciones una teoría computacional de la mente que hoy en día ha sido criticada desde muchas perspectivas (por ejemplo, Lucas, 1961; o Searle, 1980). Para aceptar que vivimos en *Matrix* haría falta muchísima más carga tanto argumental como empírica y, a día de hoy no hay nada parecido. Parece más razonable pensar que vivimos en un mundo real y objetivo, aunque no lo percibamos tal cual es, sino a través de una interfaz.

4.5. VALORACIÓN FINAL

4.5.1. EL GRAN ACIERTO

La tesis fuerte de Hoffman parece sólida y un completo acierto utilizarla como punto de partida: desde una perspectiva evolutiva parece absurdo que percibamos la realidad tal y como es, idea que además no parece mayoritaria en la comunidad académica de corte anglosajón, en donde suele primar el realismo a la vez que se defiende enfáticamente el evolucionismo. Pero es que si queremos estar en total acuerdo con la teoría de la evolución, se antoja muchísimo más razonable pensar en un interfaz que solo recoge la información útil del entorno, que en un terriblemente ineficiente mecanismo de percepción realista. Del mismo modo intentar aportar evidencia experimental de esta tesis a partir de la simulación de entornos evolutivos y la utilización de algoritmos genéticos es muy apropiada. La simulación informática de entornos evolutivos se ofrece como un gran campo de pruebas para ensayar modelos cognitivos.

Hoffman consigue hacer de un tema tradicionalmente metafísico como es el *problema crítico del conocimiento* (o, más propiamente, de la percepción), un tema abordable desde una perspectiva científica. Eso es un gran paso hacia la naturalización de la filosofía.

No obstante, aun siendo desde mi punto de mi vista la parte más fuerte e interesante de su teoría, no deja de tener ciertos problemas o, como mínimo, de dejar abiertas ciertas cuestiones que cabría entrar a elucidar. Por ejemplo, Hoffman descarta el realismo de un modo tajante: dadas las simulaciones llevadas a cabo por Mark (2013), la estrategia perceptiva realista es tan mala que, prácticamente, ni se habría llegado a ensayar en la naturaleza. Hoffman, parece demasiado radical negar la posibilidad de cierto realismo ya que, aunque los organismos naturales solo pretendan optimizar su fitness, necesitan saber realmente, dónde y cuándo se encuentra tal fitness. Aunque su percepción ignorara todo lo demás por no ser evolutivamente pertinente, de algún modo, estaría percibiendo cierto componente de la realidad de un modo verídico. Cuando los escarabajos australianos se confunden y creen que el color dorado de las botellas de cerveza es el de la espalda de las hembras e intentan copular con las mismas botellas, realmente, están percibiendo el color dorado de una forma verídica, aunque este color dorado sea un "icono" que no representa a la hembra en su totalidad.

Otra objeción podría ser que, aunque parezca, con total lógica, que una percepción realista del mundo sería un derroche de recursos en comparación con una percepción más selectiva orientada exclusivamente al fitness, no siempre tiene por qué ser así. Cabe la posibilidad de que, en ocasiones, percibir la realidad *tal y como es* no suponga demasiado coste y, por tanto, no resulte necesario crear un "icono". Hoffman debería establecer un sistema de costes para evaluar con precisión los costes y ahorros de las diferentes estrategias perceptivas.

Tampoco creo convenientes las constantes referencias de Hoffman a la cuántica para buscar justificación a la idea de que no percibimos la realidad tal como es. En primer lugar, la física cuántica es un mundo muy complicado en el que fácilmente pueden darse interpretaciones erróneas. De hecho, se ha abusado muchísimo, por ejemplo, del principio de indeterminación de Heisenberg, del experimento de la doble rendija de Young, e incluso del experimento de Aspect, para justificar las tesis posmodernas más inverosímiles (véase la celebérrima obra de Sokal y Bricmont, 1998), por término general, siempre en la línea de menoscabar la objetividad del conocimiento para fundamentar tesis relativistas o constructivistas. Hasta que la cuántica no se esclarezca más, Hoffman debería, prudentemente, quedarse en el mundo a escala humana. Ir a la cuántica es intentar resolver algo muy complicado (el problema crítico del conocimiento) con algo mucho más complicado aún (el extrañísimo comportamiento de los elementos últimos de la realidad).

Y, en segundo lugar, hemos evolucionado para adaptarnos a un mundo macroscópico en el que los efectos cuánticos son completamente imperceptibles. Por lo tanto, todo lo que suceda a nivel atómico y sub-atómico no debería tener nada que ver con nuestra percepción. De hecho, nuestra percepción parece ser *perfectamente newtoniana*.

4.5.2. LA TESIS PROBLEMÁTICA

Hoffman establece esa radical inversión ontológica entre consciencia y mundo material. Mientras que la comunidad científica parte de que de la materia emerge la consciencia (o, en el peor de los casos, que se reduce a ella), Hoffman parte de que de la consciencia emerge la materia. Así la consciencia sería el componente fundamental del universo y su teoría, aunque él lo niega en cierto sentido, es un tipo de idealismo. En esto no tiene por qué haber ningún problema, solo que, si se defiende una postura impopular

en la actualidad, hay que dar argumentos sólidos para defenderla. Y Hoffman lo hace, aunque a nuestro juicio, de modo insuficiente. Su principal argumento es que no tenemos ninguna evidencia científica de cómo la materia produce consciencia, mientras que, del camino contrario sí que tenemos. En este punto, quizá también por la brevedad de los escritos en donde lo trata, Hoffman debería concretar muchísimo más el sentido de su tesis. Tenemos experimentos que nos dicen como construimos ciertos perceptos pero, siempre, partiendo del esquema de que la información sensorial parte de una realidad externa diferente a la propia consciencia. Es decir, la evidencia experimental que defendería la idea de que el sujeto construye la experiencia sensorial comienza con la premisa de que chocaría con la idea de que la consciencia es primigenia.

Es más, la propia teoría de la evolución (y la misma teoría de Hoffman) viene a sostener que la consciencia tuvo que surgir como una adaptación evolutiva (siendo estrictamente adaptacionistas, eso sí), y, precisamente, Hoffman utiliza esta idea para demostrar la tesis de la interfaz que le sirve, a la vez, para demostrar que la consciencia es primaria a la materia… Claramente estaríamos ante un círculo vicioso.

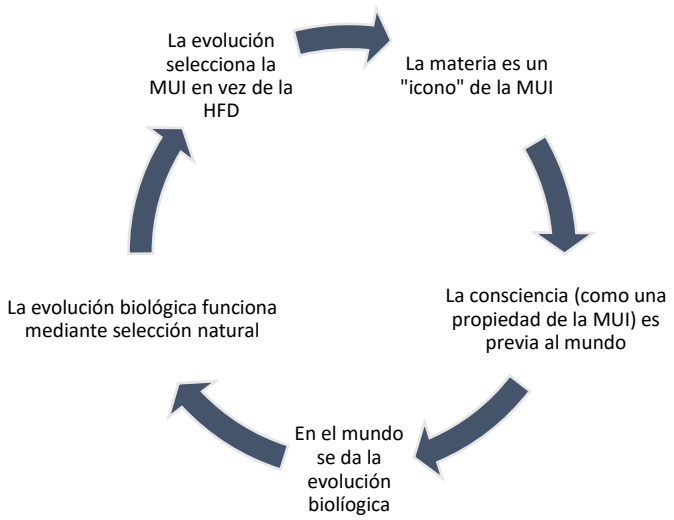


Fig. 7. Círculo vicioso del realismo de la consciencia de Hoffman. Elaboración propia.

Además, Hoffman estaría enfrentándose contra toda la neurociencia contemporánea. Parece, prácticamente, un axioma, buscar la causa de los fenómenos

mentales en sucesos físicos en el cerebro. Si el realismo consciente de Hoffman fuera cierto, gran parte de la neurociencia moderna estaría por completo equivocada: carecería de sentido cualquier investigación que intentara buscar, por ejemplo, cómo se genera una emoción a partir de neurotransmisores químicos como la dopamina o la serotonina. El camino correcto sería, tal y como sostiene Hoffman, buscar cómo la consciencia genera los neurotransmisores químicos, lo cual no deja de parecer muy extraño y contraintuitivo.

Y luego nos quedaría por saber el origen de la misma consciencia. Podemos pensar el origen de la materia en la consciencia, pero entonces, ¿cómo estudiamos el propio origen de la consciencia? ¿Cómo estudiarlo? Hoffman no aporta ningún dato de su posible origen y, peor aún, deja el tema sin ninguna salida, como una completa *hipótesis vacía*. Si situamos el origen de la consciencia en la materia tenemos una línea posible de investigación, pero si invertimos los términos tenemos una vía para investigar cómo la mente construye la materia, de acuerdo, pero estaríamos completamente perdidos para intentar encontrar de dónde surgió la propia mente. Como el mismo Hoffman menciona respondiendo a una de sus críticas, el realismo consciente quizá nos lleva a una especie de misticismo. Y, en este sentido, el realismo consciente no sería una teoría científica (al hacerse inverificable o infalsable) sino una postura metafísica, algo que parece estar muy lejos de las pretensiones de hacer una ciencia estricta de la consciencia.

4.5.3 EL TRABAJO QUE QUEDA POR HACER

El trabajo de Hoffman es, aún, muy pequeño y, prácticamente está todo por hacer:

1. Los ejemplos sacados del mundo natural que Hoffman utiliza son anecdóticos. Para que su tesis estuviese bien fundada en la naturaleza necesitaría muchísimos más ejemplos e, incluso, algún tipo de teorización acerca de la percepción animal. El ejemplo que no se cansa de repetir, el del escarabajo australiano (*Julodimorpha bakewelli*) estudiado por Gwynne y Renz (1982), que está en peligro de extinción porque el macho confunde a las

hembras con botellas de cerveza, es los que lo que los psicólogos denominan un *estímulo supernormal* (Tinbergen, 1951), que no es más que un estímulo artificial que causa una fuerte respuesta en el observador. Para que el ejemplo sirviera para justificar la teoría de la interfaz debería tener mucha más extensión. Es cierto que Hoffman (2016) pone algunos ejemplos más como el de la percepción de las libélulas o de las ranas, pero, igualmente, son evidencias insuficientes.

2. La cantidad y variedad de simulaciones de algoritmos genéticos en el que Hoffman se basa es muy escasa. Para que su hipótesis tuviera un respaldo más sólido, debería ser mucho más amplio, utilizando muchos más tipos de simulaciones que las ya citadas de Mark y Michell. A pesar de que Hoffman afirma que la complejidad perjudica al realista, sería muy conveniente establecer entornos evolutivos simulados más variados (más competidores con diferentes habilidades, distintas reglas, diferentes especies con distintos objetivos que se interfieren entre sí, entornos con muchos tipos de recursos y obstáculos diferentes etc.) o crear nuevos juegos con otras reglas. Sospechamos que juegos configurados de otro modo podrían no dar tan claramente la victoria a la estrategia de la interfaz. Por ejemplo, pensemos en un organismo A cuya mente tipo interfaz es muy eficiente y se convierte en un gran especialista en percibir un tipo de recurso que a él le proporciona un gran fitness. Siguiendo estrictamente el gradualismo darwiniano, poco a poco se va convirtiendo en un excelente especialista en percibir ese recurso, por ejemplo, cada vez a más distancia o, incluso cuando se encuentra oculto. Supongamos que ese recurso es otro organismo B el cual, igualmente, siguiendo la selección natural, evoluciona. En un momento determinado evoluciona cambiando una de sus cualidades de modo que despista al "icono" de A de forma que éste sea incapaz de detectarlo. La única manera que tendríamos para captar a B sería conociendo otras de sus características que, previamente A no captaba ya que no estaban relacionadas con ningún tipo de

aumento de fitness. Aquí entonces un organismo C que siguiera una estrategia realista sí que podría captar otras características de B y, por tanto, seguir obteniendo el recurso de fitness que A habría perdido. En este ejemplo la estrategia perceptiva realista ganaría a la de interfaz. Para elucidar casos como éste serían necesarias nuevas simulaciones. Además, servirían para librar a Hoffman de la siempre presente sospecha de que las premisas de sus juegos están *ajustadas a priori* para dar el resultado que se busca.

3. El aparataje conceptual de la teoría de la interfaz es pírrico. Hoffman solo utiliza la expresión "icono" para referirse a cómo la MUI esquematiza la realidad en el "escritorio" (que serían como las coordenadas espacio-temporales de la realidad) que, además, no define con precisión por ningún lado ¿Hay diferentes tipos de "iconos"? ¿Y diferentes "escritorios"? ¿Cómo se relacionan unos "iconos" con otros? Hoffman debería precisar mucho más este tema y convertir la metáfora del escritorio en algo más que una explicación muy gráfica y excesivamente simple.

REFERENCIAS

Bartley III, W. W., & Radnitzky, G. (1987). *Evolutionary epistemology, rationality and the Sociology of Science*. La Salle: Open Court.

Bennett B.M., Hoffman D.D., and Prakash C. (1989). *Observer Mechanics: A Formal Theory of Perception*, Academic Press, San Diego.

Bennett B.M., Hoffman D.D., and Prakash C. (2002). "Perception and evolution. En Perception and the Physical World: Psychological and Philosophical Issues" in *Perception*, ed. by D. Heyer and R. Mausfeld, Wiley, New York, pp. 229–245.

Bostrom, N. (2003). "Are we living in a computer simulation?". En The Philosophical Quarterly, 53(211), 243-255.

Bousso, R. (2002). "The holographic principle". En Reviews of Modern Physics, 74(3), 825.

Celesia G.G., Bushnell D., Cone-Toleikis S., and Brigell M.G. (1991). "Cortical blindness and residual vision: Is the second visual system in humans capable of more than rudimentary visual perception?" Neurology 41, 862–869.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford university press.

Churchland, P. M. (2013). *Matter and consciousness*. MIT press.

Collins M. (1925). *Colour-Blindness*. Harcourt, Brace & Co, New York.

Critchley M. (1965). "Acquired anomalies of colour perception of central origin". *Brain* 88, 711–724.

Dennett, D. C. (2017). *Consciousness explained*. Little, Brown.

Descartes, R. (1897-1913). *Oeuvres*. Edición de Ch. Adam y P. Tannery, 12 vols., Paris: Leopold Cerf, 1897- 1913. (Versión española de las *Meditaciones metafísicas* en Alfaguara. Madrid, 1977.)

Gabriel, M. (2018). *Yo no soy mi cerebro. Filosofía de la mente para el siglo XXI*. Barcelona: Pasado y Presente, 2016. ISBN 978-84-944950-7-6. Arbor, 193(786), 425.

Gwynne, D. T., & Rentz, D. C. F. (1983). "Beetles on the bottle: male buprestids mistake stubbies for females (Coleoptera)". En Australian Journal of Entomology, 22(1), 79-80.

Gould, S. J., & Lewontin, R. C. (1979). "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme". Proc. R. Soc. Lond. B, 205(1161), 581-598.

Gregory, R. L. (1997). "Knowledge in perception and illusion". *Philosophical Transactions of the Royal Society of London* B, 352, 1121–1128.

Hoffman D.D. (1998). *Visual Intelligence: How We Create What We See*, W.W. Norton, New York.

Hoffman D.D. (2006). "The scrambling theorem: A simple proof of the logical possibility of spectrum inversión". En Consciousness and Cognition 15, 31–45.

Hoffman, D. (2008). "Conscious Realism and the Mind-Body Problem" en Mind & Matter Vol. 6(1), pp. 87–121

Hoffman, D., Singh, M., y Mark, J. (2013). "Does evolution Favor True Perceptions?" en Proceedings of the SPIE 8651, *Human Vision and Electronic Imaging XVIII*, 865104.

Hoffman, D. (2014). "The Origin of Time in Cosncius Agentes" en Cosmology, 2014, Vol. 18. 494-520.

Hoffman, D. D., & Prakash, C. (2014). “Objects of consciousness”. Frontiers in Psychology, 5, 577.

Hoffman, D., Singh, M. y Prakash. Ch. (2016). “The Interface Theory of Perception” en *Current Directions in Psychological Science* Vol 25, Issue 3, pp. 157-161.

Hooft, G. T. (2001). “The holographic principle”. En Basics and Highlights in Fundamental Physics (pp. 72-100).

Hume, D. (2004*). Investigación sobre el entendimiento humano* (Vol. 216). Ediciones AKAL.

Knill D. y Richards W., eds. (1996). *Perception as Bayesian Inference*. Cambridge University Press, Cambridge.

Koenderink, J. J. (2014). “The all seeing eye?” *Perception*, 43, 1–6.

Leibniz, G. (1992). *Nuevos ensayos sobre el Entendimiento Humano*. Alianza, Madrid.

Lucas, J. R. (1961). “Minds, machines and Gödel”. En Philosophy, 36(137), 112-127.

Malebranche, N. (2009). *Acerca de la Investigación de la Verdad.* Sígueme, Madrid.

Maynard-Smith J. (1982). *Evolution and the Theory of Games*, Cambridge University Press, Cambridge.

McGinn C. (1989). “Can we solve the mind-body problem ?” en *Mind* 98, 349–366.

Mark, J.T. (2013). “Evolutionary pressures on perception: when does natural selection favor truth?” Ph.D. Dissertation, University of California, Irvine.

Mark, J. T., Marion, B. B., & Hoffman, D. D. (2010). “Natural selection and veridical perceptions”. Journal of Theoretical Biology, 266, 504–515.

Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge, MA: Bradford Books MIT Press.

Moya, C. J. (2011). *Filosofía de la mente*. Universitat de Valencia.

La Mettrie, J. O. (1748). *L'Homme machine*. De l'imp. d'Elie Luzac, fils (Traducción al castellano en Valdemar, 2000).

Newton, I. (1962). Sir Isaac *Newton's mathematical principles of natural philosophy and his system of the world* (Vol. 1). Univ of California Press.

Palmer S.E. (1999). *Vision Science: Photons to Phenomenology*, MIT Press, Cambridge.

Petitot, J., Varela, F. J., Pachoud, B., & Roy, J-M. (Eds.). (1999). Naturalizing phenomenology. Stanford, CA: Stanford University Press

Piaget, J. (2013). *The construction of reality in the child*. Routledge.

Poggio T., Torre V. and Koch C. (1985). "Computational vision and regularization theory". Nature 317, 314–319.

Place, U. T. (1970). "Is consciousness a brain process?" en *The Mind-Brain Identity Theory* (pp. 42-51). Palgrave, London.

Ryle, G. (1949). *The Concept of Mind*. The University of Chicago Press.

Russell, B. (1912). *The problems of philosophy*. New York: Oxford University Press.

Searle, J. R. (1980). "Minds, brains, and programs". En Behavioral and brain sciences, 3(3), 417-424.

Searle J.R. (2004). *Mind: A Brief Introduction*. Oxford University Press, Oxford.

Smart, J. J. C. (2014). *Philosophy and scientific realism*. Routledge.

Spinoza, B. (2011). Ethica. Alianza, Madrid.

Sternberg, R. J., & Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. CUP Archive.

Strawson P.F. (1966). *The Bounds of Sense, an Essay on Kant's Critique of Pure Reason*. Methuen, London.

Susskind, L. (1995). "The world as a hologram". En Journal of Mathematical Physics, 36(11), 6377-6396.

Terekhov, A., & O'Regan, K.O. (2013). "Space as an invention of biological systems", arxiv.org/abs/1308.2124.

Tinbergen, N. (1951). *The study of instinct*. Oxford, Clarendon Press.

Ullman, S. (1979). *The interpretation of visual motion*. Massachusetts Inst of Technology Pr.

Wallace, A. R. (2007). *Darwinism: an exposition of the theory of natural selection with some of its applications*. Cosimo, Inc.

BIBLIOGRAFÍA

Baars, B. J. (1998). *Metaphors of consciousness and attention in the brain. Trends in neurosciences*, 21(2), 58-62.

Blackmore, S. J. (2006). *Conversations on consciousness*. Oxford University Press.

Blackmore, S. (2013). *Consciousness: an introduction*. Routledge.

Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press.

Churchland, P. M. (2013). *Matter and consciousness*. MIT press.

Cornsweet, T. (2012). *Visual perception*. Academic press.

Crane, T. (2001). Elements of mind: an introduction to the philosophy of mind. Oxford Unitiversity Press.

Crick, F., & Koch, C. (1990). "Towards a neurobiological theory of consciousness". In Seminars in the Neurosciences (Vol. 2, pp. 263-275). Saunders Scientific Publications.

Damasio, A. R. (1994). *El error de Descartes: la razón de las emociones*. Andrés Bello.

Damasio, A. R. (2005). *En busca de Spinoza: neurobiología de la emoción y los sentimientos*. Grupo Planeta (GBS).

Dehaene, S., & Naccache, L. (2001). "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework". Cognition, 79(1-2), 1-37.

Dehaene, S. (2014). *Consciousness and the Brain*. New York, NY: Viking.

Edelman, G. M. (1989). *The remembered present: a biological theory of consciousness*. Basic Books.

Edelman, G., & Tononi, G. (2008). *A Universe Of Consciousness How Matter Becomes Imagination: How Matter Becomes Imagination*. Basic books.

Gallagher, S., & Zahavi, D. (2007). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. Routledge.

Gershon, R. (1985). "Aspects of perception and computation in color visión". Computer vision, graphics, and image processing, 32(2), 244-277.

Itti, L., & Koch, C. (2001). "Computational modelling of visual attention". Nature reviews neuroscience, 2(3), 194.

Jackendoff, R. (1987). *Consciousness and the computational mind*. The MIT Press.

Gibson, J. J. (1950). *The perception of the visual world*. The Riverside Press, Cambridge.

Holland, J. H. (1992). "Genetic algorithms". Scientific american, 267(1), 66-73.

Holland, J. H. (1995). *Hidden orderhow adaptation builds complexity* (No. 003.7 H6).

Kim, J. (2018). *Philosophy of mind*. Routledge.

Marcel, A. J., & Bisiach, E. E. (1988). "Consciousness in contemporary science" Clarendon Press/Oxford University Press.

McGinn, C. (1991). *The problem of consciousness: Essays toward a resolution*. Blackwell.

Poggio, T., Torre, V., & Koch, C. (1987). "Computational vision and regularization theory". In Readings in Computer Vision (pp. 638-643).

Schiffman, H. R. (1990). *Sensation and perception: An integrated approach*. Oxford, England: John Wiley & Sons.

Searle, J. R. (2000). *El misterio de la conciencia* (Vol. 118). Grupo Planeta (GBS).

Tanimoto, S. (Ed.). (2014). *Structured computer vision: machine perception through hierarchical computation structures*. Elsevier.

Tononi, G., & Edelman, G. M. (1998). "Consciousness and complexity". Science, 282(5395), 1846-1851.

Tononi, G. (2008). "Consciousness as integrated information: a provisional manifestó". The Biological Bulletin, 215(3), 216-242.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition* (Vol. 2). Cambridge, MA: MIT press.

Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL: Wolfram media.