**Maestría en Ciencias Humanas**

**Opción Filosofía Contemporánea**

---

Tesis para defender el título de Magíster en Ciencias Humanas

opción Filosofía Contemporánea

# *Como el silbato de una locomotora a vapor*

***Sobre la posibilidad de que la consciencia fenoménica sea un accidente evolutivo***

Maestrando: Sebastián Peralta Rodríguez

Directores de tesis: Dr. Reinaldo Bernal Velásquez | Dr. Matías Osta Vélez

Montevideo, 2024

# Aval de los directores de tesis

07 de octubre de 2024

Sr. Decano de la Facultad de Humanidades y Ciencias de la Educación,

Prof. Dr. Pablo Martins,

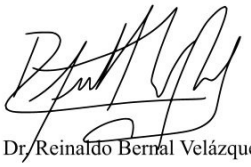
Por medio de la presente, los abajo firmantes, en calidad de supervisores de tesis, hacemos constar que hemos revisado y aprobamos la presentación de la tesis de maestría elaborada por el estudiante Sebastián Peralta Rodríguez. El trabajo mencionado se titula: “Como el silbato de una locomotora a vapor: Sobre la posibilidad de que la consciencia fenomenal sea un accidente evolutivo”. Consideramos que dicha tesis cumple con los requisitos académicos y de investigación establecidos por nuestra institución para proceder a su defensa.

Sin otro particular, ATTE

Dr. Matías Osta Vélez

Prof. Adj. de la UA de Filosofía

FHCE, UdelaR.



Dr. Reinaldo Bernal Velázquez

Profesor Asistente de Filosofía

Pontifica Universidad Javeriana, Colombia.

# Tabla de contenidos

**Aval de los directores de tesis.....I**
**Tabla de contenidos.....II**
**II.- Resumen.....V**
**III.- Abstract.....VII**
**IV.- Palabras clave.....IX**
**1.- Introducción: la hipótesis *BAV*.....1**
**2.- Enmarcando la discusión.....7**
**2.1.- El marco de la teoría de la evolución.....7**
2.1.1.- Genes, mutaciones, variabilidad.....8
2.1.2.- Selección natural.....9
2.1.3.- Deriva genética.....11
2.1.4.- Enjutas.....13
**2.2.- La consciencia: problemas de término y problemas de concepto.....14**
2.2.1.- Consciencia de creatura, consciencia de estado.....15
2.2.1.1.- Consciencia de creatura: creaturas sintientes, creaturas en alerta, creaturas autoconscientes, creaturas para las cuales hay algo que es ser como ellas.....16
2.2.1.2.- Consciencia de estado: estados autoconscientes, estados cualitativo-fenoménicos, estados para los que hay algo que es como estar en ese estado.....18
2.2.2.- Consciencia-P / Consciencia-A.....20
2.2.3.- Otros usos del término consciencia.....21
**2.3.- Consciencia fenoménica y consciencia de acceso.....22**
2.3.1.- Consciencia-P.....23
2.3.2.- Consciencia-A.....25
2.3.3.- Relaciones entre consciencia-P y consciencia-A y sus respectivas funciones.....28
2.3.3.1.- Blindsight / Petit mal / Prosopagnosia.....28
2.3.3.2.- El modelo de Schacter.....31
2.3.3.3.- Por qué estos razonamientos ignoran la distinción P-A.....33
2.3.3.4.- Consciencia-A sin consciencia-P y viceversa.....35
2.3.3.5.- Imbricación P-A.....38
**2.4.- El concepto de *función*.....40**
2.4.1.- Teleo-funciones.....41
2.4.2.- Funciones de Cummins.....42
**3.- Funciones de la consciencia.....44**
**3.1.- Inesencialismo / Epifenomenalismo.....46**
3.1.1.- Inesencialismo.....46
3.1.1.1.- Un paréntesis sobre los zombis.....48

3.1.2.- Epifenomenalismo....49
3.1.3.- Relaciones y distinciones entre estas perspectivas....51
**3.2.- Funcionalismo....53**
**3.3.- Retomando dos distinciones (creatura-estado y consciencia-autoconsciencia)....54**
**3.4.- Algunas funciones propuestas....55**
3.4.1.- La consciencia es el motor de la acción....55
3.4.2.- La consciencia fenoménica habilita un determinado tipo de razonamiento....57
3.4.3.- El consenso de la integración....58
3.4.4.- La consciencia fenoménica le otorga un carácter intrínsecamente valioso a ciertos estados mentales....59
**3.5.- El problema *fácil* y el problema *difícil*....60**
**3.6.- La brecha explicativa....63**
**4.- Volviendo sobre la hipótesis BAV....67**
**4.1.- Relatos adaptacionistas sobre la consciencia fenoménica....68**
**4.2.- Criterios para que un rasgo sea considerado una enjuta o un accidente evolutivo....70**
4.2.1.- La consciencia fenoménica no presenta una desventaja neta....71
4.2.2.- La consciencia fenoménica no es, a la vez, adaptativamente ventajosa y explicable a través de mecanismos de selección natural....73
4.2.3.- Debe haber otro rasgo del cual es un subproducto....79
**4.3.- La hipótesis BAV y su relación con la metafísica de lo mental....80**
4.3.1.- El epifenomenalismo implica la hipótesis BAV....81
4.3.2.- El dualismo de propiedades implica la hipótesis BAV....83
4.3.3.- El fisicalismo no reduccionista implica la hipótesis BAV....85
4.3.4.- Algunas intuiciones funcionalistas implican la hipótesis BAV....87
4.3.5.- El fisicalismo reduccionista es compatible con la hipótesis BAV....89
**5.- Eludiendo la determinación de una función....91**
**5.1.- Argumentos comparativos....95**
**5.2.- Argumentos por la complejidad....97**
5.2.1.- Complejidad funcional....97
5.2.2.- Complejidad anatómica....99
5.2.2.1.- Carácter unificado....100
5.2.2.2.- Múltiples canales de ingreso....101
**5.3.- Problemas con el argumento de la complejidad....105**
5.3.1.- Respuestas epifenomenalistas al argumento de la complejidad....105
5.3.2.- Otros cuestionamientos a los argumentos por la complejidad....107
5.3.2.1.- La distinción entre dos “sistemas”....107
5.3.2.2.- El problema de la energía....110
**6.- Sobre el enfoque de la pregunta por la adaptación y un problema en el argumento del funcionalismo a la hipótesis BAV .... 113**

**6.1.- La evolución de la nocicepción y la evolución de la consciencia fenoménica.....117**
**6.2.- El problema con los zombis hipotéticos.....122**
**6.3.- Inesencialismo, epifenomenalismo, y una nueva especie zombi.....126**
**7.- Una réplica a la hipótesis BAV que pretende disolver el problema difícil.....128**
**7.1.- Tres supuestos sobre el rol funcional de la consciencia fenoménica. 131**
**7.2.- La experiencia subjetiva tiene un valor intrínseco.....132**
7.2.1.- Valor subjetivo y toma de decisiones.....135
7.2.2.- El valor subjetivo de las emociones.....137
7.2.3.- Valor subjetivo y percepción consciente.....138
**7.3.- Lo que esta hipótesis permite explicar.....139**
7.3.1.- Comportamiento creativo y flexible.....140
7.3.2.- Impulso interno para la conducta.....141
7.3.3.- Unificación de la experiencia.....142
**7.4.- Implicancias de esta hipótesis para la disolución del *problema difícil* .....144**
**7.5.- Algunos problemas con esta hipótesis.....146**
**8.- Conclusión.....151**
**Referencias bibliográficas.....160**

# II.- Resumen

Llamamos *hipótesis BAV* (por el inglés *by-product or accident view*) a la hipótesis de que la consciencia fenoménica -esto es: la capacidad de tener experiencias subjetivas- no es una adaptación biológica que haya surgido como consecuencia de la evolución por selección natural, sino que se trata o bien de un subproducto de otra u otras características del funcionamiento de lo mental (una *enjuta*), o bien un accidente evolutivo que no otorga ninguna ventaja adaptativa a los organismos que exhiben tal capacidad.

En el presente trabajo se elabora un marco general de análisis para poder comprender la naturaleza y el alcance de esta hipótesis, presentando el encuadre biológico que permite operar con el concepto de enjuta o accidente evolutivo, analizando las distinciones analíticas que permiten delimitar la consciencia fenoménica y separarla de otras clases de consciencia, aclarando el concepto de función, y recorriendo las principales funciones que se han propuesto para la consciencia fenoménica.

Una vez enmarcada la discusión, se procede a analizar la hipótesis BAV, evaluando el modo en el que la consciencia fenoménica cumple o no con los criterios habituales que se utilizan para establecer el carácter de enjuta o accidente de un rasgo biológico, así como también evaluando las relaciones de implicación lógica que pueden existir entre la hipótesis BAV y diferentes posiciones metafísicas respecto de lo mental.

Por último, se analizan algunas perspectivas críticas respecto de la hipótesis BAV con el objetivo de mostrar algunos de los diferentes enfoques desde los cuales puede ponerse en duda esta hipótesis. Sobre la base de estos

cuestionamientos, se sopesa críticamente la solidez de la hipótesis y se sugiere, en virtud de esta evaluación, que la hipótesis no puede descartarse por completo, a pesar de que deberían retirarse algunos de los argumentos que se ofrecen en su defensa.

# III.- Abstract

We refer to the BAV hypothesis (from "by-product or accident view") as the hypothesis that phenomenal consciousness - that is: the capacity to have subjective experiences - is not a biological adaptation resulting from evolution by natural selection, but rather is either a by-product of other characteristics of mental functioning (a spandrel), or an evolutionary accident that confers no adaptive advantage to organisms exhibiting such capacity.

In this research, a general framework of analysis is developed to understand the nature and scope of this hypothesis, presenting the biological framework that allows operating with the concept of spandrel or evolutionary accident, analyzing the analytical distinctions that allow delimiting phenomenal consciousness and separating it from other kinds of consciousness, clarifying the concept of function, and reviewing the main functions that have been proposed for phenomenal consciousness.

With the discussion framed, the BAV hypothesis is analyzed, evaluating how phenomenal consciousness does or does not meet the usual criteria for establishing the character of a spandrel or accident of a biological trait, as well as evaluating the logical implications that may exist between the BAV hypothesis and different metaphysical positions regarding the mental.

Finally, some critical perspectives regarding the BAV hypothesis are analyzed aiming to show some of the different approaches from which this hypothesis can be questioned. Based on these inquiries, the solidity of the hypothesis is critically weighed, and it is suggested, by virtue of this evaluation,

that the hypothesis cannot be completely dismissed, although some of the arguments offered in its defense should be withdrawn.

# IV.- Palabras clave

Consciencia fenoménica; enjuta; adaptación; metafísica de lo mental; funciones de la consciencia.

**Keywords**

Phenomenal consciousness; spandrel; adaptation; metaphysics of mind; functions of consciousness.

# 1.- Introducción: la hipótesis *BAV*

> *La consciencia de las bestias parecería estar relacionada con el mecanismo de su cuerpo simplemente como un producto colateral de su funcionamiento, y ser completamente carente de cualquier poder para modificar este funcionamiento, como el silbato de vapor que acompaña el funcionamiento de un motor a vapor no tiene ninguna influencia en su maquinaria. Su voluntad, si tienen alguna, es una emoción indicativa de sus cambios físicos, no la causa de tales cambios.*
>
> Thomas Huxley

> *Puerta y calle de pájaros / muchacho saliendo al día / mira luz que es un escándalo / mira sol ir de alegría / claro que / está claro que / eso no existe.*
>
> Eduardo Darnauchans

La hipótesis BAV (llamada así por el inglés *By-product or Accident View*) respecto del origen evolutivo de la consciencia fenoménica consiste en afirmar que la consciencia fenoménica es una *enjuta* o un accidente evolutivo (Robinson et al., 2015) y que, por tanto, no necesariamente cumple una función. Para poder sopesar la solidez de esta hipótesis, es necesario en primer lugar aclarar el sentido de lo que en ella se afirma, por lo que será necesario detenernos en la delimitación y discusión de conceptos como el de a) *consciencia fenoménica*, b) *enjuta* o *accidente evolutivo* y c) *función*.

Sin embargo, aun antes de estas tareas preliminares, puede observarse en términos generales el carácter contraintuitivo de la hipótesis si se la presenta en los siguientes términos. Demos por sentado que los seres humanos (y

posiblemente muchos otros organismos), como producto de nuestras interacciones con nuestro entorno y con nosotros mismos, tenemos ciertos estados mentales que *se sienten* de determinada manera. El estado mental en el que nos encontramos cuando percibimos el color azul, el estado mental en el que nos encontramos cuando olemos una flor, o el estado mental en el que nos encontramos cuando una astilla se clava en nuestro dedo, son todos estados mentales que *se sienten* de determinada manera: parecen poseer cualidades *fenoménicas*. La *consciencia fenoménica* puede entenderse como la capacidad de tener tales estados mentales. Parece claro que una roca no los tiene, y parece claro que una persona sí (Block, 1995). En otros términos, en tanto no hay nada *que sea como ser* una roca rodando por una pendiente, la roca no posee un estado mental con propiedades fenoménicas (no posee un estado mental de ningún tipo); existe sin embargo *algo que es como ser* una persona tropezando en una pendiente, por lo que esta persona está en un estado mental que sí tiene propiedades fenoménicas (Nagel, 1974). Pues bien, la hipótesis BAV propone que este rasgo en virtud del cual podemos experimentar el mundo *de determinada manera* (el hecho de que las cosas se sientan de una determinada forma, a saber: el poseer una consciencia fenoménica) no es un rasgo que haya evolucionado como consecuencia de la selección natural. La hipótesis BAV supone que este rasgo no ha sido seleccionado en virtud de la mayor adaptación o aptitud que confiere a los organismos que lo portan, sino que es o bien a) el subproducto de algún otro rasgo o conjunto de rasgos que sí confieren una mayor adaptación, o bien b) un accidente evolutivo que no responde a ninguna presión selectiva. Si este fuera el caso, no cabría atribuirle necesariamente ninguna función biológica al hecho de que tengamos experiencias en primera persona del mundo que nos rodea.

Esta somera presentación de la hipótesis BAV, aun a cuenta de una profundización en sus diferentes aspectos, es suficiente para poder tomar nota de su carácter contraintuitivo. La dificultad de tomar en serio la hipótesis obedece a numerosas razones, pero quizás la más prominente sea el hecho inobjetable de que

nuestra consciencia fenoménica es una de las características más importantes (si no la más importante) de nuestra experiencia humana, por lo que resulta contraintuitivo suponer que se trata de un rasgo que no fue seleccionado positivamente en el transcurso de la evolución, y mucho más contraintuitivo aún suponer que no cumple ninguna función.

Para poder sopesar los puntos de apoyo que puede tener esta hipótesis en nuestra comprensión general sobre los procesos evolutivos, será necesario establecer un marco que implica una discusión específicamente biológica vinculada a la preponderancia que diferentes mecanismos evolutivos tienen en el curso general de la evolución. Existen muchos mecanismos bien establecidos que inciden en el cambio biológico que exhiben las especies a lo largo del tiempo, pero es objeto de debate el lugar que ocupa cada uno de ellos en una explicación general sobre la evolución, las interacciones que pueden presentar, y el modo en el que cada uno de estos mecanismos implica ciertos sesgos en la elaboración de los programas de investigación, entre otros aspectos.

Desde una perspectiva *adaptacionista,* todas las características fenotípicas de una especie (aquellas características vinculadas con la estructura de los organismos y sus pautas de comportamiento) constituyen adaptaciones evolutivas y, por tanto, deben redundar -de una forma clara o de una forma subrepticia- en una determinada ventaja adaptativa para aquellos organismos que poseen esa característica respecto de los que no, a fin de poder explicar su continuidad en la historia filogenética de la especie como consecuencia de la selección natural. Desde este enfoque, una tarea central a la hora de explicar evolutivamente un rasgo biológico es la de determinar su función.

Como se ha dicho, el concepto de *función* ha sido objeto de una serie de precisiones analíticas que tendrán que ser tenidas en cuenta, destacándose en la literatura dos abordajes diferentes de éste: en primer lugar, aquello a lo que Peter Godfrey-Smith (1996) llamó teleo-funciones -a partir del tratamiento que del tema

realizó Larry Wright (1973)-; en segundo lugar, las llamadas *funciones de Cummins*, llamadas así a partir de un trabajo de Robert Cummins (1975).

La necesidad de encontrar para cada rasgo una función ha sido uno de los blancos de críticas de los enfoques adaptacionistas, principalmente desde enfoques más estructuralistas que advierten sobre las constricciones a las que está sujeto el proceso de cambio evolutivo, así como a la existencia de rasgos fenotípicos de los cuales parecería imposible sostener que se trata de adaptaciones, o que confieren a sus portadores una ventaja adaptativa de ninguna índole. El sonido rítmico que produce el latido del corazón humano es un rasgo fenotípico. Sin embargo, parece claro que no se trata de un rasgo que aumente las posibilidades de éxito de aquellos organismos que lo porten, en relación con otros congéneres de latidos silenciosos. Por tanto, no es a través de la búsqueda de una función que podremos explicar la presencia de este rasgo, y será necesario para ello tomar en consideración que, como éste, existen numerosos rasgos fenotípicos cuya presencia no puede explicarse a través del análisis de su ventaja adaptativa. Como en otros casos, puede tratarse de rasgos cuya presencia se explica porque son el resultado inevitable del funcionamiento de otro rasgo fenotípico que sí puede otorgar una ventaja adaptativa a su portador, como tener un órgano que bombee sangre a través del aparato circulatorio, por ejemplo. Puede incluso tratarse de meros *accidentes*: rasgos que no solamente no poseen ninguna función, sino que tampoco son el subproducto de otro rasgo que sí la posea. Serían el resultado de procesos azarosos como los que puede presentar cualquier dinámica compleja de poblaciones, esto es, la llamada *deriva genética*. Es en estos rasgos fenotípicos, que no pueden ser explicados correctamente como adaptaciones, donde debemos ubicar las enjutas y los accidentes evolutivos.

El concepto de enjuta adquirió relevancia biológica gracias a un célebre artículo de los biólogos Stephen Jay Gould y Richard Lewontin (1979), en el que desarrollan y sistematizan la idea de que es un error tratar de atribuir a cada rasgo

fenotípico una función adaptativa, cuando pueden existir muchos rasgos que se expliquen como el sub-producto de alguna otra característica presente en la especie, o como un simple *accidente* debido a la *deriva genética*. En este artículo, los autores denuncian la ingenuidad *panglossiana*[1] de lo que denominan "el programa adaptacionista", enfocado precisamente en atribuir funciones adaptativas a cada característica de los organismos vivientes. Si bien es cierto que el debate entre *adaptacionistas* y *pluralistas* es muy anterior a este artículo de Gould y Lewontin (Mayr, 1982), no es menos cierto que allí se delinea con precisión el concepto biológico de enjuta, término que proviene de la jerga arquitectónica, y que refiere a la forma cuasi triangular que resulta de construir un arco curvo dentro de un marco recto. Las enjutas, opulentamente ornamentadas en la Basílica de San Marcos que el artículo toma como ejemplo, no fueron "diseñadas" con la función de ser un soporte para esta ornamentación; antes bien son un resultado inevitable de una constricción constructiva del edificio, sin ninguna función específica que pueda atribuírseles en su origen. La hipótesis BAV, como se ha dicho, propone que la consciencia fenoménica, análogamente a lo que puede ocurrir con una enjuta o un accidente evolutivo, no es el producto de un proceso de selección natural y, por tanto, no supone ningún reparo para explicar el origen evolutivo de la misma el hecho de que carezca de cualquier función adaptativa que pueda haberla hecho objeto de esta presión selectiva.

Habiendo presentado en términos generales la hipótesis BAV vemos con claridad que su carácter contraintuitivo no solamente no opaca su relevancia, sino que por el contrario fundamenta en todo derecho el legítimo interés en su análisis. Constituye una tarea ineludible en cualquier intento por comprender la naturaleza de la consciencia el interrogarnos acerca de cómo, a partir de componentes materiales que no parecen tener consciencia en ningún sentido relevante (como los átomos), se configuraron a lo largo del tiempo sistemas complejos (como nuestros

1 Se refieren al Dr. Pangloss, personaje del *Cándido* de Voltaire que, en tanto sátira a Leibniz, afirma que vivimos en el mejor de los mundos posibles, y que toda característica del cosmos ha sido "diseñada" con un propósito específico.

cerebros), que sí parecen tener consciencia -y consciencia *fenoménica*- aun si esta indagación nos conduce a especulaciones más contraintuitivas todavía, como por ejemplo la puesta en duda de la distinción entre átomos y cerebros -en términos de consciencia fenoménica- que ilustró este párrafo.

Quedó planteada al inicio la necesidad de abordar la hipótesis BAV a través de la consideración de tres ángulos: a) las enjutas o accidentes evolutivos, b) la consciencia fenoménica, y c) el concepto de función. Una vez trazadas las coordenadas de estas tres discusiones podremos contar con una base sobre la cual considerar críticamente la hipótesis BAV, sus apoyos y los argumentos que se esgrimen en su contra.

# 2.- Enmarcando la discusión

## 2.1.- El marco de la teoría de la evolución

¿Por qué existen los seres vivos? ¿Por qué son como son y se comportan como se comportan? ¿A qué se debe su extraordinaria diversidad? ¿Cómo se explican sus similitudes? Son estas algunas de las preguntas que han guiado el trabajo de biólogas y biólogos durante los últimos siglos, y representan la voluntad de ir más allá de una simple descripción y clasificación de los organismos vivientes. La idea de que la biología debe explicar coherente y consistentemente la existencia de los seres vivos tal como los conocemos (su variabilidad, estructura y comportamiento) y no sólo limitarse a describirlos y clasificarlos, es una pretensión central de la biología evolutiva.

La teoría de la evolución, delineada por Charles Darwin en “El origen de las especies” de 1859, constituye el primer cuerpo de hipótesis que permite explicar los más variados fenómenos biológicos. Si bien la teoría se ha desarrollado incesantemente desde los trabajos germinales de Darwin, la llamada *síntesis moderna* o *neodarwinismo* (la articulación entre los postulados darwinistas, especialmente el mecanismo de selección natural, y los descubrimientos en el campo de la genética, a través de modelos matemáticos y químicos), llevada adelante en la década de 1930 por biólogos como Sewall Wright, John Haldane y Ronald Fisher, así como el descubrimiento de la estructura molecular del ADN llevada adelante por James Watson y Francis Crick sobre las investigaciones de Rosalind Franklin, continúan siendo hasta el día de hoy la base fundamental de la comprensión biológica contemporánea. Theodosius

Dobzhansky, uno de los principales exponentes de esta síntesis moderna, lo expresó con claridad en el título de un ensayo publicado en 1973: “nada en biología tiene sentido excepto a la luz de la evolución”. Así, la teoría de la evolución constituye la herramienta fundamental para dar el salto entre descripción y explicación en el campo de la biología. (Gould, 2004)

Una de las peculiaridades de esta teoría es que constituye un cuerpo de hipótesis que pretenden explicar la existencia y variedad de organismos vivientes sobre la tierra a través de un conjunto relativamente pequeño de mecanismos. Comparada con la complejidad de lo que se pretende explicar, la simplicidad de la explicación es llamativa. Intentaremos ilustrar brevemente cómo funciona el núcleo central de la teoría evolutiva, tomando como ejemplo paradigmático el caso de los animales de reproducción sexuada.

### 2.1.1.- Genes, mutaciones, variabilidad

En el caso de la reproducción sexual en animales, cada organismo (todas sus estructuras orgánicas, incluido su cerebro y su “cableado” original, en el caso de poseer uno) está organizado siguiendo instrucciones que se encuentran codificadas en el material genético. Este “plan maestro” consta de dos copias -cada una proveniente de uno de los progenitores- y el resultado final será una combinación de ambos repertorios (las posibilidades combinatorias son múltiples, no siempre lineales, no siempre aditivas, ni siempre consistentes en un “punto intermedio”), y su materialización concreta en el marco de ciertas condiciones medioambientales.

La generación de los gametos (las células reproductoras, en general espermatozoides y óvulos) es un proceso que, generalmente, ocurre sin mayores

problemas. No obstante, como en todo sistema de copia, pueden ocurrir errores. El más común de estos errores es la mutación fija (una letra del código es sustituida por otra), aunque existen otros tipos de errores, cada uno con consecuencias también interesantes. Una mutación fija puede tener efectos deletéreos sobre el individuo que la portará, puede ser neutral (por ejemplo, ocurrir en una porción del código genético que no cumple ninguna función "codificadora") o, en algunos casos, puede ser beneficiosa. La existencia de errores de copia, sumada al entrecruzamiento (el proceso mediante el cual un organismo genera gametos virtualmente únicos que poseen una determinada combinación de información genética proveniente de cada uno de sus progenitores), y la combinación entre diferentes repertorios como consecuencia de la reproducción sexual, genera una gran variación entre los individuos de una población, no solo a nivel genotípico (la información genética con la que cuentan), sino también fenotípico (sus estructuras, caracteres, aspecto, pautas innatas de conducta, etc.). Por ejemplo, en una determinada generación de pinzones, existirá entre los individuos una importante variabilidad intraespecífica: algunos con picos ligeramente más anchos, o más largos, o más fuertes; algunos con alas ligeramente más aerodinámicas, o menos; algunos con una mayor resistencia a determinado patógeno, o con una mayor vulnerabilidad a cierto ectoparásito; algunos con un comportamiento de cortejo más arriesgado, o más conservador, etc. (Dobzhansky, 1997)

### 2.1.2.- Selección natural

En cada generación de una especie cualquiera, nacerán muchos más individuos de los que alcanzarán con vida la edad reproductiva y podrán, a su turno, tener descendencia. Las causas por las que muchos individuos no llegan a reproducirse y transmitir sus genes a la próxima generación pueden ser muy

variadas: algunas pueden deberse al mero azar, otras pueden tener que ver con cambios medioambientales extremos, y otras seguramente tengan relación con las características específicas de cada organismo. En términos estadísticos, los organismos que en mayor proporción llegarán a la edad reproductiva y lograrán procrear son, lógicamente, los mejor adaptados para tal fin, a saber: los más rápidos, los más sanos, los más fuertes, los más atractivos, los más inteligentes, los mejor camuflados, y un largo y a veces contraintuitivo etcétera[2]. Se trata, como se ha dicho, de una realidad estadística: podría pasar que el nido en el que se encuentran los huevos de pinzón con la mejor combinación genética que la especie ha visto hasta el presente, esté ubicado en una rama que será destruida por un rayo.

En tanto algunas de estas características (que hacen a algunos organismos más preparados para la supervivencia y la reproducción que otros) son heredadas genéticamente, es de esperar que cada generación de individuos exhiba cambios en relación con la anterior que no pueden explicarse sólo por el azar, sino por la actuación no azarosa de la selección natural sobre la azarosa variación intraespecífica. De este modo, las asombrosas características que podemos observar en los seres vivos, y particularmente su apariencia de haber sido deliberadamente diseñados por una inteligencia creadora superior, responden en realidad a múltiples procesos de selección natural (que no es otra cosa más que la supervivencia diferencial de algunos individuos sobre otros), operando a lo largo de miles de millones de años sobre un repertorio virtualmente infinito de combinaciones genotípicas.

2 A menudo ocurre que, dado el nicho ecológico en el que se encuentra una especie (su dieta, sus depredadores, su contexto climático, etc.), lo más beneficioso para un organismo no es ser veloz sino parsimoniosamente lento, como la lentitud metabólica que caracteriza a las tortugas. En ocasiones es más beneficioso ser llamativo que inconspicuo, como el caso de los pavos reales machos. Las características que hacen a un organismo más exitoso en términos de supervivencia y reproducción, no siempre encajan en el prototipo de “más rápido, más grande, más fuerte”.

Sin embargo, no todas las características fenotípicas que podemos observar en los seres vivos responden a este proceso de selección natural. Además de los procesos de selección natural, existen otros dos fenómenos que permiten explicar el resto de los rasgos fenotípicos: la deriva genética y las enjutas.

### 2.1.3.- Deriva genética

Llamamos “deriva genética” al proceso estocástico mediante el cual algunas características que surgen como una variación dentro de la especie paulatinamente ven modificada su frecuencia en la población y, en ocasiones, terminan fijándose como una característica universal de todos los organismos de esa especie o, por el contrario, finalmente desaparecen. Este proceso es particularmente importante para explicar rasgos que no presentan ni una ventaja ni una desventaja para el organismo (son adaptativamente neutros), aunque también ocurre respecto de cualquier otro tipo de rasgo.

Tomemos como ejemplo hipotético el color del pico de una población de pinzones que habita en una isla. Imaginemos que el color del pico está controlado por un único gen, y que dentro del *pool* genético de la población existen dos variantes de este gen (llamamos a estas variantes *alelos*): un alelo para picos de color negro y un alelo para picos de color marrón. En un primer escenario, podría ocurrir que el color de pico fuera adaptativamente neutro; esto es: no implica ninguna ventaja ni desventaja diferencial poseer un determinado color de pico respecto del otro. En este caso, la selección natural no puede operar sobre este rasgo modificando la frecuencia de estos alelos en la población -o mejor aún: no puede *explicar* las variaciones en la frecuencia de los alelos- ya que no es el color del pico la causa de la supervivencia y reproducción diferencial de los individuos. Sin embargo, la frecuencia de estos alelos efectivamente va a variar a lo largo del

tiempo. Por razones azarosas (como la caída del rayo sobre una rama), la presencia de un determinado alelo en relación con otro podrá aumentar o disminuir a lo largo del tiempo de forma aleatoria. Si analizamos la serie de datos a lo largo de un período lo suficientemente largo, puede perfectamente ocurrir que ambos alelos se mantengan presentes variando su frecuencia relativa, o podría ocurrir también que un alelo termine *fijándose*, es decir: se vuelva universal dentro de la especie y desplace por completo al alelo alternativo. Así, toda la especie podría terminar presentando picos negros y se podría "extinguir" dentro de la especie el alelo para picos marrones, sin que esto pueda ser explicado por las presiones selectivas, sino por la deriva genética.

En un segundo escenario hipotético, podría ocurrir que un determinado color de pico implicara una leve ventaja adaptativa respecto del otro, por ejemplo, porque los picos de color negro se componen de ciertas substancias que los hacen más vulnerables a partirse que los picos de color marrón. Si el único proceso evolutivo en juego fuera la selección natural, es de esperar que los picos de color marrón aumenten paulatinamente su presencia en la población en desmedro de los picos de color negro, y que finalmente éstos últimos desaparezcan por completo. Sin embargo, si esta ventaja diferencial fuera leve, perfectamente podría ocurrir que la deriva genética empujara a los picos negros (aun siendo más débiles que los marrones) a fijarse en la población y desplazar a los picos marrones. En una población pequeña, en la que -por ejemplo- la mitad de los individuos tuvieran picos marrones y la otra mitad picos negros, un evento climático extremo podría diezmar a la población y que el conjunto superviviente presentara otra distribución de frecuencia (por ejemplo, 80% de picos negros y 20% de picos marrones). Ante este nuevo escenario, es posible que un rayo desafortunado (o cualquier otro evento aleatorio) reduzca la presencia de picos marrones a una mínima expresión y, por qué no, termine finalmente en su desaparición. Como puede intuirse, mientras mayor sea la población y más notoria sea la ventaja o la

desventaja de un alelo sobre otro, menor será la incidencia de la deriva genética respecto de este rasgo.

La combinación de selección natural y deriva genética podría explicar la enorme mayoría de los rasgos que vemos en los organismos. Algunos de ellos deberán ser considerados adaptaciones, en la medida en que responden a presiones selectivas de la selección natural en virtud de su ventaja adaptativa para la supervivencia o la reproducción. Otros deberán ser considerados *accidentes evolutivos:* rasgos que no implican ninguna ventaja adaptativa (o incluso presentan una leve desventaja), pero terminaron fijándose en una población como resultado de la deriva genética. (Kimura, 1983)

### 2.1.4.- Enjutas

Adicionalmente, como se ha dicho, algunos rasgos pueden ser subproducto de otros. Por ejemplo, por razones físicas los picos de color negro absorben más energía proveniente de la luz solar que los picos de color marrón. Aceptemos que esta diferencia de energía implica una variación en la temperatura promedio de ambos picos, e imaginemos que esta variación en la temperatura afectara de algún modo el tono o el timbre del canto de estas aves. Como consecuencia de estos procesos térmicos y fonéticos, el canto de los pinzones de color marrón resulta más agudo que el canto de los pinzones de color negro. Ahora bien, a la hora de estudiar la evolución de estos tipos de canto, quizás el análisis de cómo esta variación afecta el éxito reproductivo, o el análisis de si los depredadores de los pinzones tienen un umbral de detección mayor o menor en función del tono del canto de estas aves, no nos conduzcan a ninguna conclusión. Por ejemplo, podría ocurrir que el canto más agudo fuera desventajoso, tanto por el éxito reproductivo como por la vulnerabilidad a los depredadores, pero que, sin

embargo, persistan y aumenten los pinzones de canto agudo en la población. Esto sería perfectamente plausible si la presión selectiva respecto de la fortaleza del pico fuera mayor que la que ocurre respecto del tono del canto. Así, los pinzones de canto agudo podrían subsistir y tener éxito respecto de sus competidores de canto grave, en la medida en que este canto agudo es el subproducto de otro rasgo que confiere una ventaja adaptativa superior, a saber, la fortaleza de los picos de color marrón y su resistencia a la hora de partir la cáscara de los frutos de los que estos pinzones se alimentan. Diremos, finalmente, que el canto agudo de los pinzones no es una adaptación ni un accidente evolutivo, sino una *enjuta*: el subproducto adaptativamente neutro o ligeramente desventajoso de otro rasgo fenotípico cuya ventaja adaptativa explica su presencia en la población. (Gould & Lewontin, 1979)

Sin adentrarnos en las sofisticadas discusiones teóricas y empíricas que estas ideas suscitan, queda de este modo trazado el marco general de análisis que la teoría de la evolución proporciona a la hora de explicar las formas y comportamientos que exhiben los seres vivos en nuestro planeta. Teniendo en mente estas coordenadas, podremos ver más adelante el modo en el que la hipótesis BAV se articula con nuestra comprensión general sobre la evolución y, particularmente, sobre la consciencia fenoménica. Es momento, pues, de trazar las lineas generales que nos permitan operar con este esquivo concepto.

## 2.2.- La consciencia: problemas de término y problemas de concepto

“El shock me hizo perder la *consciencia*”. “Una computadora nunca podrá ser *consciente*”. “Las substancias psicoactivas inducen estados alterados de *consciencia*”. “Soy *consciente* de las consecuencias de mis actos”. “La hipnosis

hizo que algunos de mis recuerdos se tornaran *conscientes*". "¿Cómo pudiste ser tan *inconsciente*?".

Como puede verse, en las expresiones que a modo de ejemplo acabamos de indicar aparecen usos muy disímiles del término *consciencia*. El carácter polisémico del término no puede separarse del carácter problemático del concepto. ¿Qué es la consciencia? ¿Existen diferentes *tipos* de consciencia? ¿Se trata de una propiedad de las creaturas o de los estados mentales de algunas creaturas? ¿Puede considerarse una *entidad*? Todas estas preguntas están relacionadas, y se enfocan en dilucidar de qué hablamos cuando hablamos de consciencia. Será necesario, pues, determinar con claridad algunos de los usos que pueden dársele al término *consciencia,* en primer lugar, con el fin de evitar algunas de las confusiones que pueden presentarse al tratar el asunto y, en segundo lugar, porque algunas de las distinciones analíticas que surgen al intentar clarificar el sentido del término serán de relevancia para el problema que tenemos entre manos. A tales efectos, nos guiaremos en parte por la enumeración de acepciones que sistematiza Robert Van Gulick (2022).

### 2.2.1.- Consciencia de creatura, consciencia de estado

En primer lugar, podríamos utilizar el adjetivo *consciente* aplicándolo a determinadas creaturas o a determinados estados mentales de determinadas creaturas. Afirmamos cosas diferentes al decir que un ser humano es consciente o al decir que un determinado estado mental de ese ser humano es un estado mental consciente.

### 2.2.1.1.- Consciencia de creatura: creaturas sintientes, creaturas en alerta, creaturas autoconscientes, creaturas para las cuales hay algo que es ser como ellas

El primer término de esta distinción (la "consciencia de creatura" o *creature consciousness*) es un término también polisémico (Rosenthal, 2006). Por ejemplo, podríamos hablar de organismos conscientes en el sentido de *sintientes*, esto es: capaces de percibir el mundo que los rodea y, por tanto, responder a él. El uso de este término se ha vuelto frecuente en las discusiones vinculadas a la ética animal, con el fin de destacar la diferencia entre una máquina que responde de cierta forma a determinados estímulos, y un organismo viviente que no sólo responde sino que además *experimenta* tales estímulos. Desde luego, aplicar el término *consciente* en este sentido habilita una dimensión gradual del mismo, por lo que deberemos admitir diferentes grados de consciencia, lo que genera problemas vinculados a los márgenes. Dando por sentado que un mamífero es *sintiente* y que un protozoario no lo es, y basando esta distinción en la presencia o ausencia de un sistema nervioso central, cabría preguntarse en dónde ha de trazarse el límite si consideramos aquellos casos en los que existen estructuras nerviosas de tipo ganglionar pero que no constituyen un sistema central.

También podríamos utilizar el término en el sentido de "estar alerta", con el fin de referirnos no a la mera *capacidad* de experimentar sensaciones o estímulos, sino al ejercicio *actual* de tal capacidad. En este sentido, un ser es consciente sólo en la medida en que esté actualmente experimentando tales sensaciones, generalmente si se encuentra despierto y al tanto de lo que ocurre en su entorno, a diferencia de lo que ocurre cuando se está durmiendo o en un estado de coma. Nuevamente, existen problemas que surgen en los márgenes, por ejemplo, cuando queremos determinar si puede aplicarse este término en el caso de un ser que está soñando.

Un sentido más exigente del término implicaría no solamente la capacidad o el ejercicio de experimentar ciertas sensaciones o estímulos, sino la capacidad de saber que uno está experimentando tales sensaciones o estímulos; esto es: consciencia en tanto *autoconsciencia*. Este sentido del término aplicaría para creaturas que están al tanto *de que están al tanto* pero, como cabe esperar, surgen algunos problemas a la hora de incluir entidades en esta categoría dependiendo de la forma en la que definamos esta autoconsciencia, esto es, si es necesaria una autoconsciencia conceptual explícita como la de un humano adulto normal, o si cabe admitir en la categoría a creaturas con grados más rudimentarios de autoconsciencia como, por ejemplo, a aquellas que pasan con éxito pruebas estándar de autorreconocimiento (Kohda et al., 2019).

Un uso del término que revestirá particular importancia para esta investigación es el que deriva del germinal trabajo de Thomas Nagel, *"What is it like to be a bat?"* ("¿cómo es ser un murciélago?") (1974). De acuerdo con esta perspectiva, diremos que un ser es consciente si existe algo que es como ser ese ser. El ejemplo escogido por Nagel -el de los murciélagos- es particularmente ilustrativo, en tanto si bien es posible para los seres humanos imaginar cómo se percibe el mundo desde el repertorio sensorial habilitado por un sistema de eco-localización, es imposible que tal ejercicio de imaginación nos muestre efectivamente cómo se siente ser un murciélago. Sin embargo, parece claro que existe *algo* que es como ser un murciélago, un particular modo de experimentar la realidad para el cual el murciélago es *la primera persona*, por más que esta experiencia en primera persona nos esté vedada a los seres humanos. En este sentido, diremos que el murciélago es consciente en términos de que existe este *punto de vista* de la realidad desde la perspectiva del murciélago, algo que no podríamos decir, por ejemplo, de un libro o de un cortapapel.

El tipo de consciencia que nos interesa, sin embargo, no es el que se vehiculiza a través de un adjetivo que califica creaturas, sino a través de un

adjetivo que califica estados mentales. Ambos usos del término están, desde luego, profundamente imbricados. Podría utilizarse el término *consciente* para referirse a aquellas creaturas que pueden ser sujetos de estados mentales conscientes. Para ello, desde luego, es necesario definir primero qué entendemos por tales estados o, dicho de otra forma, que es lo que hace que un estado mental sea consciente. Asimismo, puede utilizarse el término consciente en un sentido *transitivo*: este es el uso que le damos al término cuando afirmamos que una creatura es consciente *de algo*, por oposición al sentido *intransitivo* de la consciencia que no tiene un objeto por referencia (Rosenthal, 1986).

### 2.2.1.2.- Consciencia de estado: estados autoconscientes, estados cualitativo-fenoménicos, estados para los que hay algo que es como estar en ese estado

Hasta aquí nos hemos referido al término *consciente* tal como puede utilizarse para referirse a determinadas creaturas. Sin embargo, también puede emplearse para adjetivar determinados estados mentales. En estos casos, la expresión resulta igualmente polisémica. Por ejemplo, podríamos referirnos a un estado mental consciente como aquel estado mental en el cual no solamente nos encontramos, sino que estamos al tanto de que nos encontramos. El estado mental en el que me encuentro cuando siento el dolor de una astilla en mi dedo es un estado mental consciente no solamente porque estoy en ese estado, sino porque estoy al tanto de que estoy en él. Desde luego, este uso del término presupone una cierta dimensión *meta-mental*, en la medida en que pensamos sobre nuestros pensamientos, o mejor aún: tenemos estados mentales que *refieren* a nuestros estados mentales. Análogamente, esta perspectiva habilita la existencia de estados mentales no conscientes: aquellos que tenemos pero que no sabemos que tenemos, ya sea por falta de atención o por otras causas más estructurales.

Otros dos sentidos en los que puede decirse que un estado mental es consciente resultarán también de especial interés en el marco del presente trabajo y a menudo se presentan como intercambiables: se trata de los estados mentales *cualitativos* y *fenoménicos*. Puede decirse que un estado mental es cualitativo si implica ciertas propiedades cualitativas a las que nos referiremos como *qualia* o *propiedades fenoménicas*. Al probar una cucharada de mermelada de frutilla nuestra experiencia subjetiva tiene un aspecto cualitativo en virtud del cual *se siente* de manera diferente a la experiencia de probar una mermelada de naranja. Análogamente, la sensación visual que tenemos cuando observamos la cáscara roja de la frutilla difiere cualitativamente de la experiencia que atravesamos cuando observamos la cáscara anaranjada de la naranja. Estas propiedades cualitativas de nuestras experiencias que hacen que *se sientan de determinada manera* son a las que nos referimos cuando hablamos de los qualia (Tye, 2021). El concepto de qualia es radicalmente polémico, tanto en lo que respecta a su naturaleza metafísica, a sus características, a la distinción entre aquellos estados mentales que los involucran y lo que no, incluso respecto de su mera pertinencia como categoría analítica.

Por otra parte, en un sentido estrechamente vinculado con el anterior, se puede afirmar que un estado mental es consciente en tanto contiene propiedades *fenoménicas*, y por tanto se adscribe a un tipo de consciencia que llamaremos *consciencia fenoménica* o *fenomenal*. Estas propiedades fenoménicas incluyen a los qualia que podríamos llamar "sensoriales" (como la *rojez* de nuestra experiencia al observar la frutilla), pero también a otros aspectos más estructurales de la experiencia, como el formato espacio-temporal en el que ella se presenta. Se suele llamar a este aspecto de la consciencia, la *estructura fenoménica* de la misma. En lo que nos respecta, asumiremos por el momento que el hecho de poseer propiedades cualitativas o una estructura fenoménica son dos formas de referirnos a los mismos estados mentales, a los que llamaremos estados mentales conscientes en el sentido de *consciencia fenoménica*. Será fundamental para

nosotros la delimitación precisa de este tipo de consciencia y su distinción respecto de otros tipos de consciencia, como la consciencia *de acceso*.

Los estados mentales fenoménicos pueden vincularse directamente con las creaturas que anteriormente llamamos conscientes en el sentido que atribuimos a Nagel. Desde esta perspectiva, conscientes serán aquellos estados mentales para los cuales existe algo que es como estar en ese estado, haciendo nuevamente hincapié en la perspectiva "de primera persona" desde la cual se los identifica y conceptualiza. Estos tres últimos usos del término consciente para adjetivar estados mentales, a saber: a) estados mentales cualitativos, b) estados mentales fenoménicos y c) estados mentales para los cuales existe algo que es como estar en ese estado, son usos que claramente pueden yuxtaponerse. Por tanto, tomando en consideración el modo en el que el concepto de "consciencia fenoménica" es utilizado en el marco de la hipótesis BAV, las distinciones que cabría formular entre estos tres sentidos pueden soslayarse por el momento.

### 2.2.2.- Consciencia-P / Consciencia-A

En agudo contraste con este uso del término *consciente* en el sentido de consciencia *fenoménica*, es particularmente relevante la distinción que realiza Ned Block (1995) entre consciencia fenoménica y consciencia *de acceso*. Esta distinción implica la posible existencia de determinados estados mentales que, a pesar de no poseer propiedades cualitativas o fenoménicas (esto es, de no poseer esas cualidades que hacen que exista algo que es como estar en ese estado), contienen información que está disponible para ser utilizada por el individuo a efectos de guiar su conducta. Estos estados mentales conscientes en términos de *consciencia de acceso* implican una valoración de las relaciones intra-mentales entre estos estados y otros, por lo que dan cuenta de un aspecto funcional de la

consciencia que puede, a juicio de quienes suscriben a esta distinción que realiza Block, separarse de las propiedades intrínsecas que parecen tener determinados estados mentales para el sujeto que los experimenta. La distinción entre consciencia fenoménica (o *consciencia-P*) y consciencia de acceso (o *consciencia-A*) será objeto de un análisis particular.

### 2.2.3.- Otros usos del término consciencia

Por último, cabe consignar que puede utilizarse el término *consciente* para referirse a determinados estados mentales que forman parte de una *narrativa* o *flujo de la consciencia*[3], entendida ésta como la sucesión de episodios mentales tal como se le presentan en un continuo a un *espectador* (ya sea éste real o virtual) al que solemos denominar el *yo*. Esta intuitiva idea de un flujo de la consciencia se relaciona con la noción de que solamente algunas de las cosas que ocurren en la mente aparecen en el "espacio" de la consciencia fenoménica y lo hacen de manera sucesiva, por lo que el sujeto es consciente de estos estados en tanto ocupan momentáneamente este "locus" especial.

Hasta ahora, hemos referido los usos de la palabra *consciente* en tanto adjetivos que describen o bien creaturas, o bien estados mentales. Sin embargo, existe también la posibilidad lógica de utilizar el término *consciente* transformándolo en un substantivo que refiere a una determinada entidad: la consciencia. Es importante distinguir aquí entre un sentido débil del uso del término *consciencia*, como simple reificación de una abstracción ejecutada sobre las cualidades conscientes que estamos refiriendo (y hablaremos entonces de

3 Esta expresión ("stream of consciousness") tiene su origen en un trabajo de William James (1892) en el área de la psicología, pero sus desarrollos más conocidos se han dado en el campo de la literatura, particularmente a través una cierta técnica narrativa que pretende dar cuenta del flujo de la consciencia de los personajes. Véase, a este respecto, los monólogos interiores que elabora James Joyce en su novela *Ulises*.

consciencia fenoménica, consciencia de acceso, consciencia narrativa, etc., sin comprometernos con la existencia platónica de ninguna entidad que lleve estos nombres) o como una entidad “por derecho propio”, no en los términos quizás ingenuos del *élan vital* para los vitalistas, sino en el modo en que los campos electromagnéticos tienen existencia real y no solamente por referencia a la situación y propiedades de las partículas que en ellos se encuentran (Van Gulick, 2022). Esta perspectiva “realista” respecto de la consciencia, si bien no es preponderante en la literatura contemporánea, constituye una posibilidad lógica que será necesario distinguir de algunas posiciones más sofisticadas como el llamado dualismo de propiedades, o algunas perspectivas fisicalistas no reduccionistas.

En síntesis, el término “consciencia” es problemático. Si bien puede utilizarse para describir creaturas, nos enfocaremos en su uso para describir estados mentales. Estos estados mentales puede ser llamados *conscientes* en múltiples sentidos. Nos centraremos en estados conscientes en tanto *fenoménicamente* conscientes, y por tanto será necesario ahondar en la distinción entre consciencia *fenoménica* y consciencia *de acceso*.

## 2.3.- Consciencia fenoménica y consciencia de acceso

Una de las aproximaciones que podemos realizar al concepto de consciencia fenoménica es la distinción que realiza Ned Block (1995) entre este tipo de consciencia y el que él llama *consciencia de acceso*[4]. Si bien analizaremos el concepto de consciencia fenoménica también desde otras perspectivas, esta resulta de particular interés por la siguiente razón: la hipótesis BAV implica

4 Si bien la distinción está plenamente desarrollada y discutida en *On a confusion about a function of consciousness* (Block, 1995), también puede rastrearse en (Block, 1991, 1992, 1993, 1994).

reconocer explícitamente que la consciencia fenoménica debe ser distinguida de la consciencia de acceso en términos de su valor adaptativo. Ésta última posee un valor adaptativo claro (en tanto contribuye al cumplimiento de las funciones del sistema en el que se presenta), mientras que el valor adaptativo de la primera está aún indeterminado o, en el mejor de los casos, es objeto de debate (Robinson et al., 2015). Siguiendo la nomenclatura que el propio Block inició, les llamaremos *consciencia-P* (por *phenomenal consciousness* o *P-consciousness*) y *consciencia-A* (por *access consciousness* o *A-consciousness*).

Block construye esta distinción a partir de lo que reconoce como una característica del concepto de consciencia, al que denomina un concepto *híbrido* o *mestizo*, en el entendido de que bajo este concepto se agrupan diferentes tipos de fenómenos o propiedades, algunos de los cuales incluso se yuxtaponen total o parcialmente dependiendo del contexto y el uso que se le asigne al término. De este modo, en un mismo razonamiento pueden utilizarse premisas que aplican a un *tipo* de consciencia, junto a premisas que aplican a otro *tipo*, y generarse así innumerables confusiones.

### 2.3.1.- Consciencia-P

A la hora de definir la consciencia-P, Block reconoce que es muy dificultoso hacerlo sin caer en descripciones circulares, por lo que quizás lo mejor que podamos hacer es "apuntar" adecuadamente al fenómeno. Al respecto, toma como referencia de un señalamiento demasiado amplio la definición de consciencia fenoménica que presenta John Searle:

> Por consciencia simplemente me refiero a esos estados subjetivos de alerta o sintiencia que comienzan cuando uno se despierta en la mañana y continúan a través del período en el que uno está despierto hasta que uno

> se duerme sin soñar, o cae en coma, o muere, o está, como se dice, inconsciente. (Searle, 1990, p. 635)

A juicio de Block, el problema con esta clase de definiciones no es precisamente que sea circular (se está consciente cuando no se está inconsciente), sino que "apunta" a demasiadas cosas, a demasiados tipos o aspectos de la consciencia que no necesariamente son equivalentes unos a otros.

Con el objetivo de afinar la puntería respecto del fenómeno, Block define la consciencia-P de la siguiente manera:

> La consciencia-P es la experiencia. Las propiedades de la consciencia-P son propiedades *experienciales*. Los estados P-conscientes son *experienciales*, esto es, un estado es P-consciente si tiene propiedades *experienciales*. La totalidad de las propiedades experienciales de un estado es 'cómo es' ['what it is like'] tenerlo. Moviéndonos de sinónimos a ejemplos, tenemos estados P-conscientes cuando vemos, oímos, olemos, saboreamos y tenemos dolores. Las propiedades P-conscientes incluyen las propiedades *experienciales* de las sensaciones, sentimientos y percepciones, pero también incluiría las de los pensamientos, deseos y emociones. (Block, 1995, p. 230).

Es necesario destacar aquí que el aspecto polémico de la distinción de Block no es el de definir la consciencia fenoménica en la línea de lo que ya había hecho Nagel (1974), sino la de establecer que las propiedades P-conscientes o *experienciales* de un estado mental son distinguibles de las propiedades cognitivas (vinculadas con el pensamiento), intencionales (propiedades en virtud de las cuales un estado mental *refiere* o *es sobre* algo) o funcionales (las que pueden establecerse en virtud de sus relaciones de causa-efecto) de tales estados mentales. El estado mental en el que me encuentro cuando experimento en la cocina un olor a quemado es un estado mental que tiene muchas y variadas propiedades. Tiene

propiedades cognitivas, en tanto forme parte de un razonamiento que me conduce a concluir que algo se está quemando. Tiene propiedades intencionales, en tanto es un estado mental que *refiere* o es *sobre* algo: lo que se está quemando, o la alteración que produce este evento en el aire. Tiene propiedades funcionales, en tanto implica relaciones de causa-efecto que impactan en ciertos procesos neuroquímicos en mi cerebro que, a su turno, activan otros procesos cerebrales y, eventualmente, conductuales. Pero tiene, también, propiedades P-conscientes o experienciales: la experiencia de percibir ese olor a quemado se siente de determinada manera. Todas estas propiedades, si bien se presentan de manera concomitante en un estado mental puntual, deben sin embargo ser objeto de una distinción analítica.

Una observación interesante: no es necesario acordar con la descripción que hace Block de la consciencia-P para acordar con el diagnóstico de que se ha confundido esta noción (aun incongruente para sus detractores) con otras acepciones del término *consciencia* a la hora de hablar de su función, y esta confusión es en parte consecuencia de no distinguir correctamente entre las propiedades que se acaban de glosar.

### 2.3.2.- Consciencia-A

La principal distinción que traza Block, como se ha dicho, es entre la consciencia-P y la consciencia-A. Ésta última es definida en los siguientes términos:

> Un estado es acceso-consciente (A-consciente) si, en virtud de que uno lo tiene, una representación de su contenido es (1) inferencialmente promiscua (…), esto es, disponible para usar como una premisa en el razonamiento, (2) disponible para el control racional de la acción, y (3) disponible para el control racional del discurso. (Hablaré tanto de los

> estados como de sus contenidos como A-conscientes). Estas tres condiciones son en conjunto suficientes, pero no todas necesarias. Considero a (3) como no necesaria (y como no independiente de las otras), porque quiero habilitar a animales no lingüísticos, por ejemplos los chimpancés, a tener estados A-conscientes. Veo la consciencia-A como un concepto clúster, en el que (3) -a grosso modo, la reportabilidad- es el elemento del clúster con el menor peso, a pesar de que a menudo (3) es la mejor guía práctica hacia la consciencia-A. (Block, 1995, p. 231).

Tres diferencias principales traza Block entre la consciencia-A así definida y la consciencia-P. En primer lugar, el contenido de la consciencia-P es fenoménico, mientras que el contenido de la consciencia-A es representacional, en el entendido de que la naturaleza de los contenidos de la consciencia-A corresponde a jugar un papel en el razonamiento, y únicamente los contenidos representacionales pueden hacerlo. Sin embargo, algunos estados mentales tienen contenidos fenoménicos y también representacionales, por lo que es más preciso afirmar que "es en virtud de su contenido fenoménico o del aspecto fenoménico de su contenido que un estado es P-consciente, mientras que es en virtud de su contenido representacional, o del aspecto representacional de su contenido, que un estado es A-consciente". (Block, 1995, p. 232). De acuerdo con esta distinción, un mismo estado mental puede tener un aspecto P-consciente y un aspecto A-consciente, dado que puede poseer propiedades fenoménicas (o propiedades *experienciales* consistentes en "cómo es estar en ese estado") y propiedades representacionales. Los estados A-conscientes son, por tanto, necesariamente transitivos: son consciencia *de algo*. Los estados P-conscientes pueden ser transitivos, pero pueden también no serlo.

Una segunda diferencia es que la consciencia-A es un concepto funcional. Lo que hace que un estado sea A-consciente es lo que ese estado, en virtud de su contenido representacional, hace en un sistema determinado. En nuestro caso, lo que esa representación es capaz de hacer en lo que el psicólogo Daniel Schacter

llamará el *sistema de control ejecutivo (Schacter, 1989)*, esto es: el control racional de la acción y el discurso. En otros términos, lo que hace que un estado sea A-consciente es la capacidad del sistema cognitivo de poder utilizar las propiedades representacionales de ese estado en el marco de un razonamiento, o en el control racional de la conducta, lo que revela el carácter funcional del concepto.

La tercera diferencia es que existen estados mentales que son del *tipo* P-consciente: no pueden concebirse sin sus propiedades fenoménicas. El dolor, por ejemplo, es un estado mental del tipo P-consciente, dado que no puede concebirse un dolor que no *se sienta* de una determinada manera. Los estados mentales A-conscientes, por el contrario, no implican la existencia de un *tipo* particular de estado. Un pensamiento puede ser accesible en un momento y no serlo al siguiente, en la medida en que la información que conllevan sus propiedades representacionales puede estar disponible para su uso en el razonamiento o el control racional de la acción de manera contingente, no permanente.

Si bien es posible identificar a los estados P-conscientes con las sensaciones, y a los estados A-conscientes con actitudes proposicionales como los pensamientos, creencias y deseos, no es menos cierto que muchas sensaciones son estados mentales que pueden tener un contenido representacional (de este modo, la sensación de dolor que experimento en un diente puede representar -desde la perspectiva de los *representacionalistas*- un cierto daño en algún componente del diente) así como muchos pensamientos podrían tener un contenido experiencial (la forma en la que *se siente* estar pensando en la próxima visita al dentista)[5].

5 La idea de que los pensamientos pueden “sentirse de determinada manera” es muy polémica. Se vincula con la tesis de que existe una cierta “fenomenología cognitiva”, y puede ampliarse esta controvertida perspectiva en (Bayne & Montague (eds.), 2011), donde se presentan muy diferentes posiciones sobre el asunto.

### 2.3.3.- Relaciones entre consciencia-P y consciencia-A y sus respectivas funciones

#### 2.3.3.1.- Blindsight / Petit mal / Prosopagnosia

Dos condiciones clínicas sumamente intrigantes -y, sobre todo, las líneas de argumentación filosófica que pueden tejerse a partir de ellas- son de especial interés para la distinción que traza Block. En primer lugar, la llamada "visión ciega" (*blindsight*); en segundo lugar, un tipo especial de epilepsia. Con el término "visión ciega" nos referimos a la particular condición de algunas personas que, habiendo sufrido daños en la corteza visual primaria del cerebro, acaban teniendo un área de su campo visual respecto de la cual no pueden reportar ninguna experiencia visual; esto es: al presentárseles a estos pacientes un estímulo visual en esta zona "ciega" de su campo visual, reportarán no haber visto nada. Sin embargo -y aquí radica lo fascinante del fenómeno- en algunos casos se verifica que estas personas, si se les invita a "adivinar" algunas de las características del estímulo visual recibido (y respecto del cual no tienen registro consciente) como la forma, dirección, movimiento, orientación e incluso el color, acaban realizando conjeturas que tienen un nivel de acierto estadísticamente muy superior al que cabría esperar por el mero azar. Por ejemplo, pueden orientar su mano de la forma adecuada para asir un objeto que, paradójicamente, dicen no poder ver. Esta fascinante condición ha sido objeto de innumerables investigaciones y ha constituido un marco paradigmático para el abordaje de diferentes interrogantes en el campo de las neurociencias, la psicología y la filosofía (Lucero M. & Muñoz-Quezada, 2015).

A partir del caso de la visión ciega, algunos autores (Baars, 1988; Flanagan, 1991, 1993) han razonado de la siguiente manera: en tanto la consciencia está ausente en el fenómeno de la visión ciega (en lo que respecta al área del campo visual afectada), la función de la consciencia debe ser lograr que

información representada en el cerebro pueda estar disponible para ser utilizada por el sujeto para guiar *racionalmente* su conducta, así como para reportar acerca de esta información. Si bien la información que arriba al cerebro puede influir en la conducta (y de hecho lo hace) en términos *no racionales* cuando se trata de contenidos *no conscientes* (como la información que le permite a los pacientes "adivinar" la naturaleza del estímulo), parecería necesario que la información tomara estado consciente para poder ser objeto de un procesamiento *racional*.

> La idea es que cuando un contenido no es consciente -como en los contenidos perceptivos del área ciega del paciente con visión ciega- puede influir en el comportamiento de varias maneras, pero solo cuando el contenido es consciente juega de hecho un rol racional, de modo que la consciencia debe estar involucrada en promover este rol racional. (Block, 1995, p. 228).

La segunda de las situaciones clínicas a partir de la cual es posible reflexionar sobre la función de la consciencia es lo que ocurre con algunos pacientes aquejados de un tipo particular de epilepsia conocido como *petit mal*, sobre todo cuando sobreviene un episodio convulsivo o "ausencia" en el transcurso de una actividad que ha sido objeto de un alto grado de internalización, tal como caminar, conducir un vehículo o tocar el piano. En algunos casos, el individuo continúa realizando esta actividad de un modo mecánico, rígido, aparentemente automatizado, pero sin ninguna consciencia de estar haciéndolo. Consultado luego sobre ese lapso en cuestión, no parece haber registro consciente al respecto. A partir de esta interesante condición, algunos autores como Searle (1992) han elaborado un razonamiento como el que sigue. Si se toma como ejemplo el caso de algunos pacientes que pueden continuar caminando o conduciendo su vehículo de camino a casa estando totalmente "ausentes" durante un tramo considerable del recorrido -en el que logran esquivar adecuadamente a otros peatones o vehículos, pero cruzan de vez en cuando un semáforo en rojo- podríamos decir que se trata de una forma *no consciente* de guiar la conducta en atención a ciertos fines. Analizándolo en mayor detalle, vemos que se trata de

conductas altamente internalizadas por su carácter rutinario y repetitivo. Estas conductas perfectamente pueden haber sido incorporadas en la arquitectura neuronal del cerebro, incluso al punto de poder ser realizadas sin vestigio alguno de consciencia, siempre y cuando la convulsión sobrevenga una vez iniciada la acción. Sin embargo, la conducta humana exhibe un grado de *creatividad* y *flexibilidad* que no encontramos en los episodios de conducta mecánica de los pacientes aquejados de petit mal. Por tanto, lo que la consciencia le otorga al individuo es la capacidad de dotar a su conducta de esta creatividad y flexibilidad que parece caracterizarla. "Pero el comportamiento humano normal, consciente, tiene un grado de flexibilidad y creatividad que está ausente en los casos (…) del conductor inconsciente y el pianista inconsciente. La consciencia añade poderes de discriminación y flexibilidad incluso a actividades rutinarias memorizadas" (Searle, 1992, p. 115).

Se han bosquejado aquí dos razonamientos. El primero: en tanto la ausencia de consciencia impide que la información perceptual esté disponible para su uso racional, la función de la consciencia debe ser la de hacer disponible esta información. El segundo: en tanto la ausencia de consciencia habilita únicamente conductas mecanizadas y automáticas, la función de la consciencia debe ser la de otorgar creatividad y flexibilidad a la conducta[6]. Ambos razonamientos permiten ilustrar el modo en el que se traslada una función de la consciencia-A hacia la consciencia-P. El análisis de estos razonamientos, así como el experimento mental de la "súper visión ciega" que analizaremos más adelante, permitirán ilustrar también la pertinencia de la distinción conceptual entre ambos tipos de consciencia. Pero para ello, es necesario glosar brevemente otras condiciones clínicas relevantes, así como algunos modelos que se han propuesto acerca de cómo se relacionan las funciones mentales que podrían dar cuenta de estos y otros fenómenos.

6 Tal como se verá más adelante, el vínculo entre la consciencia fenoménica y el carácter creativo y flexible de la conducta constituye uno de los aspectos centrales de un interesante razonamiento que puede construirse contra la hipótesis BAV.

Una condición clínica también fascinante, y que resulta de interés para la presente discusión, es la *prosopagnosia*. Los pacientes aquejados de prosopagnosia no pueden reconocer rostros. Esto es: son conscientes de que aquello es un ojo y esto de más aquí una nariz; pueden describir el color de un iris y la forma de un labio; pero no pueden reconocer globalmente un rostro como *el rostro de* una persona particular, ni siquiera de sus familiares más próximos. Para estos efectos, necesitan información adicional sobre una persona (el timbre de voz, un accesorio distintivo, el modo de caminar o de moverse) para lograr identificarla correctamente (Sacks, 2017).

Dos fenómenos interesantes se han verificado en algunos pacientes con prosopagnosia. En primer lugar, existen pacientes a los que se les muestra la fotografía de alguien, y luego se les muestran algunas fotografías más de diferentes personas entre las cuales se encuentra la misma persona de la primera fotografía. A pesar de la diferencia en expresión, posición, iluminación y ángulo, algunos pacientes pueden "acertar" cuál es la fotografía indicada, a pesar de que su prosopagnosia es tan profunda que no pueden reconocer a sus parejas ni a sus hijos en ninguna fotografía (Etcoff et al., 1991). En segundo lugar, algunos pacientes con prosopagnosia tienen éxito a la hora de identificar dos rostros en una categoría ocupacional similar ("Bush" y "Reagan", por ejemplo) a pesar de reportar que ninguno de los rostros les es familiar (Sergent & Poncet, 1990). En ambos casos, parece haber un acceso a cierta información que no necesariamente se presenta bajo la modalidad de una consciencia fenoménica.

### 2.3.3.2.- El modelo de Schacter

Una vez expuestos los fenómenos antes descriptos, Block hace referencia a un determinado modelo elaborado por Daniel Schacter (1989) acerca de cómo

puede constituirse la economía de la mente, esto es, un modelo de cómo diferentes "módulos" con funciones específicas (la memoria, la respuesta conductual, la percepción, el registro de rostros, la espacialidad, la consciencia fenoménica, etc.) podrían vincularse entre sí. Este modelo, al que Schacter denomina DICE (*Dissociable Interactions and Conscious Experience*) tiene relevancia para la presente discusión en virtud de algunas de las ideas sobre las que se funda:

> 3.- En algunos casos de impedimentos neuropsicológicos, módulos específicos de procesamiento y memoria son selectivamente desconectados del sistema consciente, lo que resulta en un déficit de la experiencia consciente específico en cuanto al dominio. 4.- La información que no tiene acceso al sistema consciente puede de todos modos afectar los sistemas de respuesta verbal/motora. 5.- El sistema consciente funciona como una puerta de acceso al sistema de control ejecutivo. (Schacter, 1989, p. 363).

El interés de Block en este modelo es que permite ilustrar la diferencia entre consciencia-P y consciencia-A, al tiempo que permite elaborar conjeturas interesantes acerca de las causas y los mecanismos de fenómenos descriptos en el marco de la visión ciega, el petit mal, y también la prosopagnosia.

En el modelo de Schacter, el módulo encargado de la consciencia fenoménica (Schacter le llama "*conscious awareness system*") es un intermediario entre los módulos especializados en cierto tipo de conocimiento (léxico, conceptual, facial, espacial, etc.) y el sistema de control ejecutivo, en el que puede tener lugar el procesamiento racional de la información. Si la vía de comunicación cerebral entre uno de estos módulos especializados (por ejemplo, el del conocimiento facial en el caso de la prosopagnosia) y el módulo de la consciencia fenoménica se encuentra dañada, entonces esa información no puede tornarse consciente y, por tanto, no puede estar disponible para su consideración *racional* a la hora de guiar la conducta o el discurso. Sin embargo, este módulo especializado en el conocimiento facial tiene sus propias vías de comunicación con los sistemas de respuesta verbal y motora. Podría especularse que esta

configuración es la que habilita a los pacientes con prosopagnosia a “acertar” a la hora de elegir fotografías a pesar de no tener registro consciente de la pertenencia de los rostros que observan. También podría, a partir de este modelo, explicarse por qué los sujetos que tienen más éxito en esta clase de pruebas son aquellos catalogados como más “pasivos” (Weiskrantz, 2012), en el entendido de que este vínculo entre los módulos especializados de conocimiento y los sistemas de respuesta verbal puede verse opacado cuando el sistema de control ejecutivo “se hace cargo” y trata de guiar racionalmente las respuestas a través de algún tipo de estrategia. Análogamente, algunos sujetos que presentan alexia (particularmente la alexia posterior u occipital, que impide leer adecuadamente palabras “de un vistazo” pero permite leerlas una letra a la vez) pueden “adivinar” correctamente algunas palabras que se les presentan fugazmente, pero esta capacidad se reduce si la palabra se exhibe durante unos instantes más (Coslett & Saffran, 1994). Nuevamente, parecería que el sistema de control ejecutivo tiene tiempo para recolectar más información y esto impide la puesta en funcionamiento de otros sistemas periféricos no conscientes que podrían entrar en acción.

El modelo de Schacter presupone la idea de que si un contenido mental no es consciente en términos fenoménicos, no puede estar disponible para su uso racional en el momento de guiar la conducta, pero al mismo tiempo permite que una información pueda influir en las respuestas del sujeto a pesar de no ser consciente en términos fenoménicos.

### 2.3.3.3.- Por qué estos razonamientos ignoran la distinción P-A

Dejando de lado por el momento la discusión que establece Block con otros conceptos de consciencia (como la auto-consciencia o el monitoreo interno),

es posible visualizar el problema que existe con los razonamientos que se han elaborado a partir de la visión ciega y del petit mal.

Recordemos el razonamiento que establece Searle: en tanto la pérdida de la consciencia implica la ausencia de creatividad y flexibilidad en la conducta, la función de la consciencia debe ser aportar esta creatividad y flexibilidad. Tanto Searle como Van Gulick (1989) hacen referencia a la consciencia-P, en tanto se refieren a ella con términos como "experiencia consciente en el sentido fenoménico". Sin embargo, no ofrecen ninguna razón para creer que en los casos de petit mal la consciencia-P se vea suprimida. Es cierto que la conducta de estos sujetos exhibe una menor -o nula- flexibilidad y creatividad, pero nada indica que su experiencia fenoménica, o los aspectos fenoménicos de su experiencia, desaparecieron. Por el contrario, es imposible dar cuenta de su comportamiento sin dar por hecho que siguen existiendo en ellos estados mentales P-conscientes.

> Por ejemplo, Searle, citando a Penfield, describe al caminante epiléptico como alguien que "zigzaguea su camino" a través de la multitud. ¿Acaso no *ve* los obstáculos que evita? Suponga que llega a su casa girando a la derecha en una pared roja. ¿Acaso no hay algo que es para él como ver una pared roja – y no es esto diferente a lo que es para él ver una pared verde? Searle no da ninguna razón para creer que la respuesta es no. (Block, 1995, pp. 239-240, cursivas en el original)

Lo que parece verse afectado en estos sujetos son los procesos cognitivos o de pensamiento, incluida la consciencia-A, pero no necesariamente la consciencia-P. Para que el razonamiento de Searle y de Van Gulick funcione, debería probarse que la consciencia-P ha desaparecido. Lo que ocurre es que se está apelando al concepto mestizo de consciencia, en el que se confunden y se yuxtaponen los dos sentidos que Block intenta distinguir. De esta manera, el razonamiento parte de la premisa de que la consciencia-P está ausente (sin evidencia de ello) y de que la creatividad está ausente, para concluir que aquella es la causa de ésta. El hecho de utilizar el concepto de consciencia sin distinguir entre consciencia-P y

consciencia-A hace que una función obvia de la consciencia-A se transfiera ilegítimamente a la consciencia-P.

En lo que respecta al razonamiento que se traza a partir del caso de la visión ciega, recordemos que tanto Marcel (1986), como Flanagan (1993) y Van Gulick (1989) afirman que un paciente con visión ciega jamás iniciará voluntariamente una acción hacia un estímulo que se le presente en el área ciega de su campo visual. El ejemplo paradigmático de ello, parafraseando a Marcel, es que un paciente con visión ciega sediento no tomará el vaso de agua que se le presenta en el área ciega de su campo visual. Infieren, a partir de ello, que una de las funciones de la consciencia es tornar disponible la información que proviene de los sentidos para su uso en el control racional de la conducta. Sin embargo, una vez que tomamos en consideración la distinción entre consciencia-P y consciencia-A, podemos notar que ambas están ausentes en el caso de los pacientes con visión ciega, pero que eso no implica afirmar que una función de la consciencia-A (a saber: hacer disponible la representación de su contenido para guiar la conducta) puede atribuirse necesariamente también a la consciencia-P. En el caso de la visión ciega, y habiendo trazado la distinción, podrían evaluarse otras opciones: la ausencia de consciencia-P se debe a la ausencia de consciencia-A; la ausencia de consciencia-A se debe a la ausencia de consciencia-P; la ausencia de ambos tipos de consciencia se debe a un tercer factor. El hecho de que la consciencia-P se presente en la mayoría de los casos *junto* a la consciencia-A es lo que explica la confusión que denuncia Block y, por otra parte, la existencia misma del concepto como concepto mestizo.

### 2.3.3.4.- Consciencia-A sin consciencia-P y viceversa

A la hora de pensar casos de consciencia-A en ausencia de consciencia-P, la discusión ha gravitado fuertemente en torno a muy poderosos experimentos mentales. Podemos imaginar, por ejemplo, un sistema informático que pudiera reproducir *computacionalmente* el funcionamiento de un cerebro humano pero sobre una infraestructura de silicio que por alguna razón no *soportara* la consciencia-P: una súper inteligencia artificial que replicara cómputo por cómputo el proceso de cálculo de información que hace un cerebro humano en un momento dado, pero sin dejar de ser un programa informático, una mera máquina, no un sujeto de experiencias fenoménicas en primera persona. Block afirma que estos casos son conceptualmente posibles. Como veremos más adelante, se trata de un caso concreto de toda una miríada de experimentos mentales vinculados con los llamados *zombis filosóficos*, protagonistas ineludibles de la discusión contemporánea sobre la consciencia fenoménica.

Otro ejemplo de consciencia-A sin consciencia-P podría ser el hipotético caso de la "súper visión ciega". Como se ha dicho, los pacientes que presentan visión ciega pueden "acertar" correctamente cuando se les ofrecen opciones binarias para describir el estímulo que se les presentó en el área ciega de su campo visual (horizontal o vertical, por ejemplo) pero no poseen consciencia-P de tales estímulos, ni tampoco consciencia-A, dado que distinguimos entre el procesamiento racional y la "intuición" con la que el paciente "adivina", o por lo menos hasta el momento en el que surge un nuevo y diferente estado mental: el de escuchar y *creer que son verdaderas* las propias intuiciones. Podríamos imaginar, sin embargo, un caso alternativo de visión ciega: la "súper visión ciega". Un paciente con súper visión ciega exhibiría las mismas propiedades que un paciente con visión ciega, pero podría inducirse a sí mismo, voluntariamente, a "adivinar" la naturaleza de un estímulo que se le presentara en su área ciega, a diferencia de los pacientes de visión ciega, que permanecen ajenos a cualquier estímulo hasta que se les induce externamente a "adivinar". Este paciente con súper visión ciega podría espontáneamente afirmar que *sabe* que hay una X en su campo visual ciego

a pesar de que no la *ve*. La idea de que hay una X simplemente *aparece* en su mente sin ninguna experiencia sensorial que la desencadene. Puede comparar la experiencia de *saber* que hay una X en su campo visual ciego, con *ver* que hay una X en su campo visual normal.

> Hay algo que es como experimentar esto último, pero no aquello, dice. Es la diferencia entre *sólo saberlo* y saberlo a través de una experiencia visual. Tomemos su palabra como sincera; he aquí el punto: el contenido de que hay una X en su campo visual es A-consciente pero no P-consciente (Block, 1995, p. 233).

Una vez que surge este pensamiento, este sujeto tiene un pensamiento que es al mismo tiempo A-consciente y P-consciente, pero lo interesante no es aquí el pensamiento, sino el estado de su sistema perceptivo que da origen a su pensamiento, que en este caso es A-consciente pero no P-consciente.

Ante la necesidad de pensar situaciones en las que podría -al menos teóricamente- presentarse consciencia-P sin consciencia-A, Block refiere a nuestra comprensión sobre cómo funciona la corteza visual, que puede dividirse en subsistemas dorsales y subsistemas ventrales, vinculándose mucho más lo que ocurre en los subsistemas ventrales con la consciencia-P que lo que ocurre en los subsistemas dorsales. Esta constatación habilita la posibilidad, al menos conceptual, de que los circuitos neuronales vinculados con el razonamiento y el control de la acción puedan verse dañados, sin que exista afectación de la consciencia-P[7]. La posibilidad de que la consciencia-P y la consciencia-A estén asentadas en regiones o sistemas cerebrales que puedan tener algún grado de disociación, constituye un elemento importante en la discusión sobre cómo enfocar la consciencia fenoménica a la hora de querer explicar su surgimiento desde una perspectiva evolutiva.

7 Hay un interesantísimo contrapunto a esta posibilidad en (Dennett, 2015), en el que el autor especula con una sección de tejido nervioso mantenida con vida en una caja de Petri que puede tener consciencia del rojo. Analizaremos un ejemplo similar más adelante, en el marco de la hipótesis de Cleeremans y Tallon-Baudry sobre la función de la consciencia fenoménica.

Otra razón para creer en la posibilidad de la consciencia-P sin consciencia-A es la que Block presenta por vía de ejemplo, y que parece relacionado con un comentario que hace Searle (1992, p. 144) respecto del carácter supuestamente *inconsciente* de la sensación de la tela de nuestra camisa rozando nuestro cuello, hasta que tomamos consciencia de tal sensación (Searle apela a este ejemplo para mostrar la confusión entre la distinción consciente/inconsciente y la distinción centro/periferia respecto de la atención). En el ejemplo que propone Block, hemos estado teniendo una conversación intensa a lo largo de la mañana, y descubrimos al mediodía que del otro lado de la ventana una ensordecedora máquina excavadora ha estado excavando la calle con inusitado estrépito desde hace horas. En ese instante nos damos cuenta de que hemos estado escuchando ese sonido toda la mañana, pero recién al mediodía tenemos consciencia de estarlo escuchando. Es decir que fuimos P-conscientes del sonido durante toda la mañana, pero recién al mediodía fuimos, además de P-conscientes, A-conscientes del mismo.

### 2.3.3.5.- Imbricación P-A

En lo que respecta al objeto de este trabajo, los comentarios finales de Block resultan de un profundo interés. Si la ausencia de consciencia-P determina la ausencia de consciencia-A, entonces Marcel, Flanagan y Van Gulick tienen razón *de alguna manera*, en el entendido de que el aspecto fenoménico de la consciencia quizás sea el que posibilita que la información contenida en él sea representada en términos de premisas en un razonamiento o elementos involucrados en el control de la conducta, pero de un modo *indirecto*, a través de la consciencia-A. Si bien los razonamientos que Block cuestiona están fundamentalmente errados al deslizar ilegítimamente una función de la consciencia-A hacia la consciencia-P, no es menos cierto que en la enorme

mayoría de las circunstancias, ambos aspectos de la consciencia se presentan simultáneamente. Refiriéndose a lo que ello podría significar, afirma que:

> Esto sugiere una íntima relación entre la consciencia-A y la consciencia-P. Quizás hay algo en la consciencia-P que engrasa los engranajes de la accesibilidad. Quizás la consciencia-P es como el líquido en una computadora hidráulica, el medio a través del cual la consciencia-A opera. Alternativamente, quizás la consciencia-P es el portal hacia los mecanismos de acceso como en el modelo de Schacter, en cuyo caso la consciencia-P tendría la función que Marcel y los otros mencionan. O quizás la consciencia-P y la consciencia-A *equivalen a casi la misma cosa empíricamente a pesar de que difieren conceptualmente, en cuyo caso la consciencia-P también tendría la mencionada función*. Quizás las dos están tan entrelazadas que no tiene sentido empírico la idea de una sin la otra. (Block, 1995, p. 242, cursivas nuestras).

Pero hay al menos dos formas en las que ambos tipos de consciencia pueden estar así de entrelazadas. En el primer caso -a saber, que la consciencia-P sea literalmente idéntica a una forma de procesamiento de la información- la consciencia-P tendrá la misma función que ese procesamiento cognitivo. En el segundo caso, cabe la posibilidad de que la consciencia-P sea un subproducto de un determinado tipo de procesamiento de la información, un epifenómeno superveniente a éste (Jaegwon Kim, 1993), en cuyo caso, presumiblemente, la consciencia-P no tiene ninguna función, posibilidad que se abordará en la siguiente sección.

Como puede verse, esta posibilidad -la de que la consciencia fenoménica sea un subproducto de, por ejemplo, ciertos tipos de procesamiento de la información- es particularmente relevante para el análisis de la hipótesis BAV. La hipótesis de que la consciencia fenoménica es un subproducto de otro rasgo o capacidad presupone la posibilidad de distinguir entre la consciencia-P y la consciencia-A, por lo que habrá de tenerse en mente esta distinción a la hora de evaluar los argumentos que pueden construirse al respecto de esta hipótesis.

## 2.4.- El concepto de *función*

Al reconocer las distinciones analíticas que propone Block entre la consciencia fenoménica y la consciencia de acceso, nos debemos aproximar a una discusión conceptual de relevancia para el objeto de este trabajo, a saber: la de definir correctamente a qué nos referimos con el término *función*. Después de todo, son algunos razonamientos que atribuyen funciones a la consciencia fenoménica los que estarían viciados por obviar la distinción entre consciencia-P y consciencia-A. El concepto de función es, además, central al debate sobre el carácter adaptativo o no un determinado rasgo fenotípico. Como se ha dicho, si un rasgo tiene una función adaptativa clara, es difícil sostener que se trata de una enjuta o de un accidente evolutivo, y se vuelven más plausibles las hipótesis que explican su surgimiento por acción de la selección natural. El término *función* requiere, pues, una aclaración.

En el caso de los objetos manufacturados, parece sencillo comprender la referencia del término: aquel efecto para cuya realización fue diseñado el objeto. Sin embargo, a la hora de establecer la función de un rasgo biológico, se torna necesario revisar esta definición. Al margen de las posiciones creacionistas o del llamado *diseño inteligente*, no cabe afirmar que un determinado rasgo biológico fue *hecho para* tener determinado efecto. Nadie *construyó* las alas de las aves *para* que pudieran volar. Sin embargo, parece sensato afirmar que, en cierto sentido, esa es la función que cumplen, por lo que a nuestra definición intuitiva de función deberemos quitarle la creación deliberada.

A su turno, en el caso de un rasgo como la presencia de alas (que genera una clara ventaja adaptativa para los organismos que lo portan, dadas ciertas condiciones ecológicas vinculadas con el modo de vida de determinadas aves), parece claro que el hecho de que las alas les permitan a las aves -voladoras- volar, es precisamente *la razón* por la cual la presencia de alas fue objeto de la selección

natural y, por tanto, es *la razón* de su presencia en las aves voladoras. Tomando esto en consideración, podríamos arribar a una noción de función que podría definirse de la siguiente manera: la función de un rasgo biológico es el efecto que dicho rasgo tiene en la aptitud adaptativa del organismo que lo porta y que, por tanto, explica su presencia en determinada línea filogenética por efecto de la selección natural.

### 2.4.1.- Teleo-funciones

Siguiendo la distinción que realiza Peter Godfrey-Smith (1996) entre dos concepciones fundamentales acerca de lo que puede ser entendido como una *función* en el campo de la biología contemporánea, diremos que la definición de función que se presenta en el párrafo anterior corresponde a la noción de función que presenta Larry Wright (1973) y que podríamos llamar *funciones teleonómicas* o *teleo-funciones*. Desde esta perspectiva, de los muchos efectos que puede tener un rasgo, el que señalaremos como su función es aquel que tenga una mayor *relevancia explicativa* en el marco de un proyecto explicativo sobre su existencia.

Existen tres razones de peso por las cuales es complejo establecer la teleo-función de la consciencia (Seth, 2008). En primer lugar, porque el cerebro humano exhibe una inusitada complejidad, y en un sistema complejo es muy difícil detectar relaciones de causalidad entre el efecto de una parte y la ventaja adaptativa que presenta el sistema en su conjunto. En segundo lugar, porque la función que puede tener un rasgo en el presente no necesariamente es la función por la cual ha sido seleccionado en el pasado. En tercer lugar, como se ha dicho y como veremos en mayor profundidad, no todos los rasgos biológicos son el resultado de la selección natural (la *deriva genética* o las *constricciones estructurales* podrían también explicar su existencia) por lo que no todos los

rasgos biológicos deben tener una teleo-función. Estas dos últimas razones serán de particular relevancia para la discusión de la hipótesis BAV.

Una dificultad adicional consiste en reconocer que existe una distancia entre elaborar una historia sobre cómo un determinado rasgo *pudo haber surgido* por mecanismos de selección natural, y encontrar evidencia empírica que verifique que *efectivamente* esta historia es veraz. Este es uno de los principales problemas del *programa adaptacionista* que denuncian Gould y Lewontin (1979) a partir de la consideración de las enjutas.

### 2.4.2.- Funciones de Cummins

Adicionalmente, existe otro enfoque acerca de cómo definir una *función* que a menudo se ha presentado como contrapuesto al de Wright pero que podría ser considerado complementario (Godfrey-Smith, 1996): se trata del concepto de función que presenta Robert Cummins (1975) y al que Godfrey-Smith llama *función de Cummins*. Desde esta perspectiva "adscribir una función a algo es adscribirle una capacidad que se distingue por su rol en el análisis de alguna capacidad del sistema del cual forma parte" (Cummins, 1975, p. 765). Una de las ventajas respecto de la perspectiva teleo-funcional es que la identificación de la función de una *parte* de un sistema no tiene por qué implicar que *el sistema en su conjunto* exhibe la misma función. Es más, la función del sistema en su conjunto podría perfectamente identificarse apelando al concepto de teleo-función. En el ejemplo clásico de establecer que la función del corazón es bombear sangre, diremos que esta es la función del corazón, desde una perspectiva teleo-funcional, porque es lo que explica la existencia *del corazón* en términos de su origen evolutivo. Desde una perspectiva de funciones de Cummins, diremos que su función es bombear sangre porque este es el efecto causal relevante para explicar la capacidad que exhibe *el sistema circulatorio en su conjunto* de hacer circular el oxígeno a través del organismo, entre otras. Esta discusión sobre teleo-funciones y

funciones de Cummins, y su relevancia para explicar la función de la consciencia (si la hubiere) deberá ser tenida en cuenta, en parte porque:

> (…) proponer una función de Cummins para la consciencia no excluye un rol teleológico, simplemente habilita la posibilidad de que tal rol teleológico pueda no existir. Por tanto, desde una perspectiva de la función de Cummins, podemos evaluar la *saliencia explicativa* de la consciencia en relación con el comportamiento del organismo en su conjunto sin necesariamente sostener la afirmación de que la consciencia fue seleccionada por la selección natural precisamente en virtud de estos efectos explicativamente salientes. (Seth, 2008, p. 283, cursivas nuestras).

En otras palabras, la saliencia explicativa de la consciencia fenoménica podría entenderse como la forma en que ésta constituye un elemento ineludible para explicar la capacidad de los organismos conscientes de llevar adelante determinados tipos de conducta, sin que esto implique necesariamente postular que la consciencia fenoménica existe porque fue seleccionada positivamente en el transcurso de la evolución. Como veremos más adelante, esta distinción entre la pregunta sobre la función de un rasgo y la razón de su existencia es fundamental para sopesar la hipótesis BAV en el marco de una explicación evolutiva de la consciencia fenoménica.

# 3.- Funciones de la consciencia

Cualquier aproximación racional al problema de la consciencia, y particularmente al de la consciencia fenoménica, consiste en la formulación de algunas interrogantes que, si bien íntimamente entrelazadas, pueden distinguirse entre sí. Estas interrogantes pueden organizarse en tres grandes áreas. La primera de ellas consiste en preguntarse acerca de la naturaleza de la consciencia fenoménica, cómo podemos describirla, qué elementos la componen, en síntesis: qué es. La segunda interrogante, vinculada con la necesidad de dar una explicación racional para la existencia de la consciencia fenoménica, consiste en preguntarse acerca de su origen: por qué surgió en nuestro planeta esto que llamamos "consciencia fenoménica", si es un efecto de la realidad física o si consiste en una propiedad no enteramente física pero también fundamental, de qué manera encaja la consciencia fenoménica en las explicaciones científicas con las que hemos dado cuenta de la enorme mayoría de los fenómenos naturales, si es un producto de la evolución o no, y si así es, de qué modo. A partir de esta segunda interrogante surge inmediatamente una tercera: cuál es la función que cumple la consciencia fenoménica, *para qué sirve*, si es que sirve *para algo*. Como se ha anticipado, esta última pregunta presupone que la consciencia fenoménica tiene una función, intuición que no debe darse por sentada. Otras interrogantes estrechamente vinculadas con esas tres podrían ser las de preguntarnos acerca del carácter intencional de la consciencia fenoménica (su capacidad para referir a algo), y acerca de los límites o condiciones epistemológicas que se encuentran frente a un objeto de estudio tan particular, entre muchas otras.

En las secciones anteriores se han bosquejado algunas respuestas a la primera de las interrogantes planteadas: cómo puede describirse la consciencia y

cómo, a partir de una determinada descripción, podemos establecer una distinción entre la consciencia de acceso y la consciencia fenoménica. Tanto la segunda interrogante como la tercera planteadas (el origen de la consciencia y su función) se abordarán teniendo en mente casi de forma exclusiva a la consciencia fenoménica, en atención a la hipótesis objeto de investigación.

Antes de abordar lo que la teoría de la evolución puede aportar a la dilucidación del origen de la consciencia fenoménica, será necesario organizar el listado de las posibles funciones que en la literatura se le han asignado.

Si bien nuestra experiencia cotidiana parecería mostrarnos que la consciencia fenoménica juega un rol fundamental en nuestra vidas, en el control de nuestras acciones, y en las relaciones que entablamos con el mundo y las demás personas que habitan en él, no es una tarea sencilla determinar con exactitud cuál es la función que ésta puede cumplir, y se han elaborado complejos argumentos filosóficos que sostienen la idea de que la consciencia fenoménica no cumple ninguna función, o que incluso no puede -por razones físicas o incluso metafísicas- tener efectos causales de ningún tipo en la realidad. Será necesario, para poder glosar algunas de las funciones que se han sugerido para la consciencia fenoménica, tener presente nuestra discusión sobre el carácter polisémico de términos como *consciencia* y *función*. Como veremos, para algunas de las funciones propuestas será de relevancia la distinción que tomamos de Block (1995) entre consciencia fenoménica (*consciencia-P*) y consciencia de acceso (*consciencia-A*). Para otras, la relevancia de esta distinción será más problemática. Análogamente, algunas de las funciones propuestas pueden ser catalogadas como teleo-funciones, mientras que otras como funciones de Cummins.

Siguiendo el análisis que al respecto realiza Seth (2008), antes de preguntarnos cuál es la función de la consciencia fenoménica es de orden preguntarnos acerca de aquellas posiciones para las cuales la consciencia fenoménica no cumple ninguna función. Las dos principales vertientes en filosofía

de la mente que representan esta línea de pensamiento son el *inesencialismo* de la consciencia fenoménica y el *epifenomenalismo*.

## 3.1.- Inesencialismo / Epifenomenalismo

### 3.1.1.- Inesencialismo

El inesencialismo de la consciencia es la contraintuitiva perspectiva según la cual "para cualquier actividad inteligente *i*, llevada a cabo en cualquier dominio cognitivo *d*, aun si *nosotros* hacemos *i* con acompañamientos conscientes, *i* puede -en principio- realizarse sin esos acompañamientos conscientes" (Flanagan, 1993, p. 12, cursivas en el original)[8]. En otros términos, las funciones que habitualmente asignamos a nuestra actividad mental consisten básicamente en acciones y procesos que, si bien se presentan en los seres humanos como procesos conscientes, pueden en principio realizarse sin esta condición. Numerosas actividades que hasta hace unas décadas hubiéramos creído potestad exclusiva de la mente humana, hoy pueden desarrollarse por sistemas informáticos sin que tengamos que suponer que en ellos hay *necesariamente* un "acompañamiento consciente".

Un ejemplo a este respecto puede ser el modo en el que funciona un sistema informático de navegación por GPS en el tránsito de una ciudad que elabora sugerencias en tiempo real. Este sistema, *a grosso modo*: a) "observa" la posición y velocidad de un conjunto de vehículos transitando en diferentes lugares de la ciudad; b) "lee" los mensajes enviados por las personas sobre accidentes, obras viales, enlentecimientos, etc.; c) "construye" una suerte de mapa en el que

8 Veremos que esta posición se relaciona con lo que Robinson y otros (2015) atribuyen a una lectura *débil* de los relatos adaptacionistas sobre la consciencia.

“visualiza” todas estas contingencias simultáneamente asignándoles diversos grados de confiabilidad a los reportes que recibe en función de su consistencia con el resto de los datos disponibles; d) “analiza” la solicitud hecha por el usuario de establecer la mejor ruta entre el punto *A* y el *B*, e) “pondera” las diferentes rutas que podrían plantearse en articulación con el mapa que construyó y, finalmente, f) indica varias rutas alternativas posibles con sus correspondientes longitudes y duraciones. Sin embargo, no es necesario sostener -aunque podría sostenerse- que esta “observación”, “lectura”, “construcción”, “análisis” o “ponderación” son procesos *conscientes* en ningún sentido relevante.

Sin embargo, algunas *actividades inteligentes* parecerían no poderse llevar a cabo en ausencia de consciencia. Hay una discusión acerca de si algunas de ellas, como la introspección o la manifestación deliberada de una mentira, pueden ser contraejemplos que pongan en duda el inesencialismo de la consciencia. Sin embargo, no se trata de dificultades insalvables para el inesencialismo. En el caso de la introspección, parece ser una actividad inteligente *esencialmente* consciente en la medida en que constituiría un contrasentido hablar de una introspección inconsciente: es una actividad que exige *acompañamientos conscientes* en virtud *de su definición*, y no de sus relaciones causales. En el caso de la mentira, podría definírsela como la intención de engañar, sea ésta consciente o no, y pueden observarse en la naturaleza numerosos casos de organismos que “engañan” a otros (por ejemplo, cuando aparentan un tamaño mayor del que tienen, o fingen estar muertos) sin que tengamos la obligación lógica de suponer en ellos una consciencia *de estar engañando*.

Si tomamos la distinción entre consciencia-A y consciencia-P, parece claro que el inesencialismo cobra interés en tanto refiera a la consciencia fenoménica. Por el contrario, si reinterpretamos la afirmación de Flanagan teniendo en mente la consciencia de acceso, podríamos pensar en algunas actividades inteligentes *i* que directamente no podrían llevarse adelante (por nosotros ni por ningún otro

sistema de procesamiento de la información) sin “acompañamientos conscientes” en términos de consciencia de acceso. Si un estado es A-consciente en la medida en que una representación del mismo está diponible para ser utilizada como premisa en un razonamiento y/o para guiar la conducta, entonces casi todas las actividades conscientes que podríamos llamar *inteligentes* implican estados A-conscientes, lo que implicaría la falsedad del inesencialismo así entendido.

Para lograr una mejor apreciación del interés en esta posición, es importante traer a colación el concepto de zombi filosófico (Flanagan & Polger, 1995).

### 3.1.1.1.- Un paréntesis sobre los zombis

Llamamos zombi filosófico (o *p-zombie*, por *philosophical zombie*) a una creatura hipotética, física y conductualmente indistinguible de un ser humano normal, pero que carece en absoluto de consciencia fenoménica. Tomando la expresión de Nagel (1974), no hay nada que sea como ser un zombi: un zombi está “oscuro por dentro”. Parte del interés filosófico en este concepto es que forma parte de numerosos argumentos, particularmente los vinculados con diferentes posiciones metafísicas respecto de lo mental.

Existen algunas formas alternativas de servirse de este concepto y, por tanto, varias formas de definirlo. Podríamos hablar de a) zombis como réplicas conductuales de los seres humanos que carecen de consciencia fenoménica, b) zombis como réplicas conductuales *y físicas* de los seres humanos que carecen de consciencia fenoménica, e incluso c) zombis como réplicas conductuales de los seres humanos *en todo lo concerniente a sus características adaptativas*, que

carecen de consciencia fenoménica. En la mayoría de las ocasiones, el término "zombi filosófico" se utiliza en el segundo de los sentidos planteados.

De acuerdo con esta clasificación, un zombi en el sentido *a* puede estar hecho de plástico y metal, con microprocesadores de silicio, pero responder del mismo modo que nosotros a los estímulos del mundo exterior. Los avances de la inteligencia artificial, eventualmente combinados con tecnología adecuada para recibir y procesar estímulos del mundo exterior, facilitarían la tarea de imaginar esta clase de zombis. Los zombis tipo *b* constituyen una versión más *exigente* de los zombis, en tanto aquí se postula su total equivalencia *física* con un ser humano normal, esto es: una creatura que no solo es indistinguible en lo que respecta a la conducta, sino que también es indistinguible en lo que respecta a la constitución orgánica y biológica de su ser, por decirlo de una forma: una réplica *átomo por átomo*, con exactamente las mismas relaciones funcionales, pero sin ningún vestigio de consciencia fenoménica.

La posibilidad real de los zombis, así como la mera posibilidad de concebirlos, ha sido objeto de un interesantísimo debate cuyo grado de sofisticación ha alcanzado niveles difíciles de describir en esta breve mención. Baste por ahora mencionar a los zombis como una realización de la hipótesis inesencialista.

### 3.1.2.- Epifenomenalismo

Como se ha dicho, otra importante perspectiva teórica según la cual la consciencia no cumple ninguna función es el epifenomenalismo. De acuerdo con esta perspectiva -a la que se denomina en ocasiones como *la sospecha epifenomenalista*- la consciencia existe, pero no juega ningún papel en los

procesos neuronales o cognitivos del cerebro humano, en el sentido de que las propiedades fenoménicas no tienen poder causal alguno. Del mismo modo que la sombra acompaña el movimiento de un cuerpo sin afectarlo en absoluto, la consciencia acompaña procesos cognitivos sin afectarlos en absoluto. Parafraseando una célebre cita del *bulldog de Darwin*, Thomas Huxley *(1874)*, la consciencia es a los procesos internos de un organismo, lo que el silbato de vapor al funcionamiento de la locotomora.

Trayendo nuevamente a escena nuestra distinción entre consciencia fenoménica y consciencia de acceso, parece claro que la hipótesis epifenomenalista puede englobar tanto a la consciencia-A como a la consciencia-P. De hecho, desde una perspectiva epifenomenalista, todo lo mental puede perfectamente ser causalmente inerte. Nuestra vida mental, a la que tanta importancia le atribuimos, no solamente es -desde esta perspectiva- la consecuencia de procesos físicos que ocurren en nuestro cerebro, sino que además nuestros estados mentales (A-conscientes o P-conscientes) no tienen efectos causales sobre ninguna otra cosa. Todos los efectos causales que atribuimos a nuestros estados mentales deberían ser, de hecho, las consecuencias de los estados físicos que a su turno le dieron origen a tales estados mentales.

Tomemos como ejemplo lo que ocurre cuando al salir de casa surge en mi mente la duda acerca de si guardé mis llaves en el bolsillo izquierdo o en el bolsillo derecho de mi pantalón. Podemos imaginar que ante esta interrogante razono que la manera más efectiva y eficiente de obtener una respuesta es tanteando mis bolsillos. Al hacerlo, noto en el bolsillo izquierdo un objeto que produce en mí una sensación táctil como la esperable ante un manojo de llaves y también una sensación auditiva que es coherente con esa hipótesis. Concluyo, a partir de estas percepciones, que las llaves están en ese bolsillo y las cambio de lado en atención al peligroso orificio que ese bolsillo tiene.

Todo lo que ocurrió en mi mente, desde el momento en que surge la duda hasta la acción de cambiar las llaves de lugar, incluidos los razonamientos elaborados y las experiencias cualitativamente conscientes de palpar y oír, deben tener -es difícil sostener lo contrario- un correlato en procesos neurofisiológicos que estaban ocurriendo paralelamente en mi cerebro: una sucesión de estados físicos que se corresponden en una línea temporal con todos estos procesos mentales. También es sensato pensar que todos estos procesos cerebrales (que podríamos reducir a impulsos electroquímicos enviados desde una neurona a otras en una red de interconexiones de sobrecogedora complejidad) son de naturaleza física y que, por tanto, responden a causas físicas. Pero si así es, es sensato concluir que el carácter consciente de estos estados mentales es causalmente irrelevante para la sucesión de estados físicos que comenzaron con el estado de mi cerebro cuando dudé sobre la ubicación de las llaves, hasta el estado físico de mi cerebro cuando ordené a los músculos de mi brazo que hicieran las contracciones necesarias para cambiar las llaves de lugar. En otros términos, podría elaborarse una cadena causal paso a paso desde un estado físico inicial a un estado físico final que diera cuenta de cada una de estas pequeñas conexiones causales sin que surgiera en ningún momento la necesidad de apelar al carácter consciente de ningún estado mental como *causa* de alguno de los estados físicos (aunque, desde luego, bien puede ser aquel *consecuencia* de éstos). Negar esta conclusión implicaría poner en duda la hipótesis de que todo hecho físico tiene una causa suficiente que también es física, en otros términos: implicaría poner en duda la *clausura causal* del mundo físico (Jaegwon Kim, 1992).

### 3.1.3.- Relaciones y distinciones entre estas perspectivas

El inesencialismo respecto de la consciencia y el epifenomenalismo no deben identificarse. Que el inesencialismo sea falso no implica que el

epifenomenalismo también lo sea, en la medida en que los procesos cognitivos podrían implicar *necesariamente* el surgimiento de la consciencia, sin que ésta tenga ningún poder causal. Análogamente, que el inesencialismo sea cierto tampoco implica que el epifenomenalismo lo sea, en la medida en que un determinado proceso cognitivo podría en principio desarrollarse sin la presencia de la consciencia, pero en aquellos casos concretos en los que ésta está presente, podría perfectamente tener poderes causales en el desarrollo de tales procesos.

De acuerdo con lo que plantea Seth (2008) tomando como referencia un trabajo de Antti Revonsuo (2005), no debe confundirse el epifenomenalismo antes descripto (al que podemos llamar epifenomenalismo *metafísico*, y es el sentido habitual en el que se utiliza el término "epifenomenalismo") con el epifenomenalismo *biológico*. El epifenomenalismo metafísico implica afirmar que la consciencia no tiene poderes causales en absoluto, y en la medida en que tener poderes causales puede ser una de las características distintivas de *lo real* (Alexander, 1920), la consciencia perfectamente podría no ser real. Esta postura está relacionada con algunas perspectivas llamadas *eliminativistas*.

El epifenomenalismo biológico, por otra parte, asume que la consciencia puede tener poderes causales, pero estos poderes causales no han incidido en absoluto en la aptitud adaptativa de los organismos que la poseen, por lo que no puede haber sido objeto de selección natural. Muchos rasgos de los seres vivos son claramente reales y tienen poderes causales, pero son irrelevantes desde el punto de vista del éxito reproductivo de aquellos organismos que los portan: no pueden haber sido objeto de la selección natural por esta razón. El epifenomenalismo biológico, como puede verse, está íntimamente relacionado con la hipótesis BAV respecto del origen evolutivo de la consciencia.

Tanto el epifenomenalismo como el inesencialismo respecto de la consciencia representan posiciones contraintuitivas, como se ha dicho, dada la importancia que parece tener la consciencia en nuestra vida mental, dado el hecho

de que la actividad neuronal parece ser la única actividad de carácter físico que le da origen, y dado el carácter aparentemente funcional que tiene la consciencia a la hora de generar estados mentales intrínsecamente valiosos como el placer que ,a su vez, parecen motivar la conducta (Cleeremans & Tallon-Baudry, 2021). Sin embargo, tanto la evidencia empírica como los desarrollos teóricos en contra del epifenomenalismo y del inesencialismo respecto de la consciencia continúan siendo polémicos, por lo que estas perspectivas (como casi todas las perspectivas que abrevan de un cierto espíritu escéptico) no se apartan con un simple ademán.

## 3.2.- Funcionalismo

Antes de avanzar en la presentación de algunas de las funciones que se le han atribuido a la consciencia, es importante establecer la diferencia entre una indagación racional acerca de la función de la consciencia y la adhesión al *funcionalismo* como corriente de pensamiento en el campo de la filosofía de la mente. A cuenta de un análisis más detallado, por "funcionalismo" nos referimos a la perspectiva en virtud de la cual todo estado mental *no es más* que una propiedad "de segundo orden", resultante de las relaciones entre estímulos y respuestas (ya sea que se entiendan estas relaciones en un sentido causal o en el sentido de operaciones computacionales), así como la miríada de disposiciones que tales estados mentales implican. Por ejemplo: "el estado mental de sentir dolor está completamente caracterizado por disposiciones a decir 'ay', a preguntarse si uno está mal, a tomar una aspirina, etc." (Seth, 2008, p. 283). Uno de los principales flancos a través de los cuales ha sido objeto de crítica el funcionalismo es, paradójicamente, una de las tesis que en un momento se presentó en su defensa: la llamada tesis de la *realización múltiple* (Putnam, 1972). Esta tesis consiste en afirmar que los mismos estados mentales que sobrevienen al

funcionamiento del cerebro humano, podrían ser realizados por múltiples substratos físicos, incluso no orgánicos. Es sensato que esta tesis haya sido presentada por los funcionalistas para cuestionar la *identidad psicofísica* (la idea de que los estados mentales *son* idénticos a procesos neurofisiológicos).

Por otra parte, se han presentado argumentos contra el funcionalismo en su versión computacional, como el célebre argumento de la habitación china de Searle (1980). Como veremos, podría perfectamente identificarse una función para la consciencia (ya sea una teleo-función o una función de Cummins) desde algunas perspectivas fisicalistas que son contrarias al funcionalismo.

## 3.3.- Retomando dos distinciones (creatura-estado y consciencia-autoconsciencia)

A la hora de discutir sobre la función de la consciencia, algunas de las distinciones entre diferentes usos del término que hemos referido en secciones anteriores serán de relevancia. Una de las distinciones que fueron establecidas es aquella entre la consciencia como característica de una creatura a la que llamamos "consciente", y la consciencia como característica de un estado mental al que llamamos "consciente". Aquí la diferencia entre niveles, al igual que la distinción entre funciones teleológicas y funciones de Cummins, adquiere una relevancia central: la función que cumple para un organismo su *capacidad de tener* estados conscientes no necesariamente es idéntica a la función que cumple el hecho de que *determinado* estado mental sea un estado consciente, como afirma David Rosenthal (2008) al evaluar esta distinción.

En la misma línea, es relevante para esta discusión la distinción que hemos trazado entre ser consciente de algo vinculado con el mundo externo (lo que

podríamos llamar la consciencia *primaria*) y ser consciente de que somos conscientes de algo (lo que podríamos llamar la auto-consciencia, consciencia de orden superior o HOC, por el inglés *"higher-order consciousness"*). La función de la consciencia primaria seguramente difiere de la función de la consciencia de orden superior, y esta distinción resulta de interés evolutivo porque permite elaborar una discusión sobre la evolución de la consciencia en animales no-lingüísticos, aun sin suponer en ellos ningún nivel de auto-consciencia. La relevancia de esta distinción entre consciencia primaria y consciencia de orden superior (que se relaciona, pero que no es idéntica, a la distinción entre consciencia-P y consciencia-A) es central en importantes teorías de la consciencia que apelan a estados mentales de orden superior (o teorías HOT, por el inglés *"higher-order thought"*) (Rosenthal, 2006).

## 3.4.- Algunas funciones propuestas

### 3.4.1.- La consciencia es el motor de la acción

Respecto de las funciones que pueden proponerse para la consciencia fenoménica, una de las que más se amolda a nuestro sentido común es que la consciencia fenoménica juega un rol fundamental en el inicio de la mayoría de las conductas motoras o verbales, esto es: la decisión *consciente* de llevar adelante una conducta determinada (fenoménicamente consciente, en la medida en que *siento* que decido actuar de tal manera) es la causa -en virtud de su aspecto volitivo- de la acción que luego se lleva adelante. En el ejemplo de las llaves, es el carácter consciente de mi decisión de mover las llaves de lugar lo que motiva e inicia la acción motora de moverlas.

Esta idea de sentido común presenta numerosos problemas. Algunos de ellos surgen de la investigación empírica acerca del modo en el que cierta actividad neuronal en la corteza motora (medida a través del "potencial de preparación" o *"readiness potential"*) se produce *antes* de que el individuo tenga el deseo consciente de moverse (Haggard, 2008; Libet, 1985). A primera vista, si los procesos cerebrales que dan origen a una conducta pueden desarrollarse antes de que el sujeto tenga la experiencia consciente de iniciar esa conducta, entonces no es un acto voluntario consciente lo que desencadena la acción. Desde algunas perspectivas se ha cuestionado esta interpretación de los resultados experimentales, planteando que sólo desde una posición cartesiana que acepte a la consciencia como un "primer motor inmóvil" de la mente, o desde una posición que sostenga que los procesos mentales conscientes sólo pueden ser provocados por otros estados mentales conscientes, puede observarse con sorpresa que existan procesos inconscientes que desencadenen lo que conscientemente reconocemos como nuestra voluntad de actuar (Flanagan, 1993).

Sin embargo, la discusión al respecto se ha complejizado aún más a partir de una serie de experimentos comenzados por Daniel Wegner y Thalia Wheatley, (1999) con los cuales se ha podido encontrar evidencia empírica de que los estados mentales conscientes que interpretamos subjetivamente como el desencadenante volitivo de una acción pueden ser -en condiciones experimentales- tanto *inducidos* en situaciones donde no fue el sujeto el iniciador consciente de la acción, como *suprimidos* en situaciones en los que sí lo fue. Esto ha dado origen a lo que estos autores llaman la "causación mental aparente", en virtud de la cual tanto la consciencia volitiva como la acción desarrollada tienen en ambos casos orígenes inconscientes que las desencadenan, pero ciertos factores que relacionan a la voluntad de realizar la acción con la acción en sí (su precedencia y proximidad en el tiempo, su adecuación a la acción realizada, la ausencia de una causa externa notoria) generan en los sujetos la *ilusión* de que el estado mental consciente fue la causa de la acción. Como puede verse, este

interesante argumento no sólo cuestiona a la voluntad de la acción como candidata a ser una de las funciones de la consciencia, sino que además se relaciona íntimamente con el problema de la sobredeterminación causal (Jaewgon Kim, 2000) que ha sido uno de los principales argumentos a favor del epifenomenalismo.

### 3.4.2.- La consciencia fenoménica habilita un determinado tipo de razonamiento

Una posible función de la consciencia fenoménica que también puede resultar intuitiva es que la consciencia fenoménica cumpla un rol en el razonamiento, entendiendo éste como la evaluación racional de la información de la que se dispone con el objetivo de guiar racionalmente la conducta. En términos adaptativos, parece una ventaja para algunos organismos no responder a su entorno de manera únicamente instintiva sino sopesar racionalmente sus acciones. Introspectivamente, parecería que sólo es posible razonar acerca de qué curso de acción seguir si uno tiene consciencia no solamente de la información involucrada en el razonamiento sino del proceso mismo de evaluación. Desde esta perspectiva, se torna importante la conceptualización que hace Block (1995) de estados A-conscientes, dado que se los define como aquellos cuyas representaciones están disponibles para ser utilizadas como premisas en un razonamiento, para guiar racionalmente la conducta o el discurso. Por tanto, es a la consciencia de acceso y no a la consciencia fenoménica a quien cabría asignarle esta función.

Sin embargo, la evidencia empírica puede sembrar algunas dudas respecto de esta intuitiva idea. Por ejemplo, los experimentos de Richard Nisbett y Timothy Wilson (1977) parecen mostrar que las personas no son totalmente conscientes de los procesos de razonamiento que condujeron a una determinada decisión, y

consultadas luego sobre estos procesos, construyen relatos sobre los mismos que no se corresponden con la realidad, sino con modelos idealizados de cómo habría de razonarse correctamente. En la misma línea, experimentos realizados más recientemente (Johansson et al., 2006) indican que las personas pueden construir relatos acerca de cómo razonaron para tomar determinada decisión cuando, de hecho, tomaron una decisión diferente de la que están "justificando".

En los experimentos mencionados anteriormente, se pone en duda la relación entre el carácter consciente de un determinado proceso cognitivo y la acción desencadenada -aparentemente- por éste. Sin embargo, existe otro ángulo desde el cual puede ponerse en duda esta supuesta función de la consciencia, y consiste en indagar si la evaluación racional de la conducta es siempre o necesariamente una ventaja. Algunos trabajos (Dijksterhuis et al., 2006) arriban a la contraintuitiva conclusión de que, en lo que respecta a decisiones complejas, son mejores (tanto objetiva como subjetivamente) las opciones que se toman sin realizar un razonamiento consciente y minucioso sobre los aspectos positivos y negativos de cada opción disponible. Asimismo, lo que podríamos llamar un "exceso de raciocinio" tampoco resulta ventajoso en términos adaptativos para el sujeto, como parecen indicar algunos casos (Damasio, 1999) en los que la excesiva atención a la evaluación racional de ciertas decisiones inhabilita al sujeto para poder tomarlas en lo que consideraríamos un tiempo "prudente".

### 3.4.3.- El consenso de la integración

Otras funciones que se le han adjudicado a la consciencia, y que cuentan con importante respaldo empírico, son las que se engloban dentro de lo que Ezequiel Morsella (2005) llama "el consenso de la integración": una constelación de teorías que comparten como denominador común el lugar de la consciencia

como integradora de distintos tipos de información que de otro modo se encontrarían compartimentados en diferentes áreas y subsistemas del cerebro. Entre las teorías que comprenden este consenso, podemos mencionar la teoría del espacio de trabajo global o GW (por el inglés *Global Worskspace*), la teoría de la interacción supra-modular, la teoría de la selección de grupos neuronales o TNGS (por el inglés *Theory of Neural Group Selection*), dentro de la cual se encuentran la llamada hipótesis del núcleo dinámico o DCH (por el inglés *Dynamic Core Hypothesis*), y la teoría de la información integral o IITC (por el inglés *Information Integration Theory of Consciousness*).

Al margen de las especificidades de cada uno de estos enfoques teóricos, es relevante para la consideración de la hipótesis BAV la idea general de que la consciencia fenoménica permite unificar información proveniente de distintas "fuentes" y generar una suerte de modelo central del cual el sujeto fenoménicamente consciente es el centro. Como veremos, esta perspectiva permite establecer una línea de razonamiento contraria a la hipótesis BAV (Nichols & Grantham, 2000), línea que a su vez puede ser controvertida por quienes la defienden (Robinson et al., 2015).

### 3.4.4.- La consciencia fenoménica le otorga un carácter intrínsecamente valioso a ciertos estados mentales

Por último, una de las posibles funciones de la consciencia que se ha mencionado anteriormente será de particular relevancia, ya que se puede formular como una réplica directa y explícita a la hipótesis BAV. Se trata de la posición presentada por Axel Cleeremans y Catherine Tallon-Baudry (2021) según la cual la función de la consciencia fenoménica es la de otorgarle a ciertos estados mentales un carácter intrínsecamente valioso. De acuerdo con esta perspectiva, la

distinción entre consciencia-A y consciencia-P, si bien ha enmarcado en gran medida las sucesivas investigaciones sobre la función de la consciencia, parte de un supuesto debatible: el de asumir que el carácter fenoménico de la consciencia *puede disociarse* de sus aspectos funcionales. Para Cleeremans y Tallon-Baudry esta disociación es inadecuada, en la medida en que el carácter fenoménico de ciertos estados mentales puede constituir, *en sí mismo*, parte de su función. Según los autores, el hecho de que un estado mental sea consciente en sentido fenoménico es precisamente lo que hace que la experiencia en primera persona asociada con él pueda ser *intrínsecamente valiosa* y, por tanto, es lo que hace que los organismos emprendan acciones para aproximarse a estados de cosas que impliquen experiencias fenoménicas de valor "positivo", así como acciones que los alejen de estados de cosas que impliquen experiencias fenoménicas de valor "negativo".

Esta perspectiva será analizada con más detalle más adelante, pero es relevante incluirla como una posibilidad destacada a la hora de asignarle funciones a la consciencia fenoménica, en parte porque el cuestionamiento a la distinción entre consciencia-P y consciencia-A nos aproxima a una distinción conceptual íntimamente ligada con ésta: la distinción que propone Chalmers entre el *problema fácil* y el *problema dífícil* de la consciencia.

## 3.5.- El problema *fácil* y el problema *difícil*

Habiendo enumerado algunas de las funciones propuestas para la consciencia, es momento de dar un paso atrás para poder ampliar el alcance de nuestra mirada y evaluar en qué medida estas posibles respuestas a la función de la consciencia pueden enmarcarse en la investigación general sobre el fenómeno

de la consciencia, y cuáles son los límites o las fronteras a las que dicha investigación se enfrenta.

En su influyente artículo *"Facing up to the problem of consciousness" (1995)*, David Chalmers analiza el modo en el que habitualmente se ha enfocado el tema de la consciencia en las investigaciones tanto filosóficas como científicas al respecto. De los muchos problemas a los que estas investigaciones han intentado dar respuesta, la enorme mayoría de ellos pueden ser enmarcados en lo que Chalmers llama *el problema fácil*. Se trata de problemas para los cuales, si bien en algunos casos no existe una respuesta clara aún, *podrían* (al menos en principio) ser resueltos a través de la investigación y la experimentación empírica. Problemas como el de explicar el modo en el que los organismos categorizan la información proveniente de los estímulos externos y reaccionan a ellos, el de explicar el modo en el que la atención puede "focalizarse", el de explicar el modo en el que la información de diferentes fuentes puede integrarse en un único sistema cognitivo, el de explicar el modo en el que puede deliberadamente controlarse la conducta, entre otros, constituyen problemas que son susceptibles de ser abordados exitosamente a través de programas de investigación enmarcados en los métodos estándar de las ciencias cognitivas. Es perfectamente sensato esperar que el avance de nuestra comprensión de cómo se procesa la información en el cerebro y cómo se establecen redes neuronales para ello, finalmente lleva a la elaboración de respuestas sólidas a todas estas preguntas.

> No existe una verdadera discusión respecto de si *estos* fenómenos pueden ser explicados científicamente. Todos ellos son directamente vulnerables a la explicación en términos de mecanismos computacionales o neuronales. (…) A pesar de que aún no tenemos nada cercano a una explicación completa de estos fenómenos, tenemos una idea clara de cómo podríamos arreglárnosla para explicarlos. Esta es la razón por la que los llamo problemas fáciles. Desde luego, "fácil" es un término relativo. Dar cuenta adecuadamente de los detalles probablemente insumirá uno o dos siglos de dificultoso trabajo empírico. Aun así, tenemos todas las razones para creer que los métodos de la ciencia cognitiva y de la neurociencia tendrán éxito. (Chalmers, 1995, pp. 2-3, cursivas en el original)

Como puede verse, muchas de las funciones de la consciencia que hemos enumerado están directamente asociadas con esta clase de problemas. Por ejemplo, para dar cuenta de la focalización de la atención podría elaborarse una explicación que mostrara la forma en la que algunos procesos neuronales tienen prioridad sobre otros en determinadas situaciones. Para explicar la manera en la que el contenido de ciertos estados mentales es accesible en términos de su reportabilidad, podría elaborarse una explicación del modo en el que ciertas redes neuronales conectan la información de ciertas áreas del cerebro con otras áreas encargadas del control de la acción y la expresión verbal. Para explicar la manera en la que la consciencia implica integración de la información proveniente de diferentes fuentes, podría elaborarse una explicación del modo en el que algunas áreas cerebrales ocupan un lugar en la "economía informacional" del cerebro en su conjunto que les permite realizar tal integración. En todos estos casos, una explicación de estos fenómenos a través de la indagación empírica y la experimentación parece compleja pero, perfectamente posible.

Sin embargo, estos problemas no agotan las preguntas que surgen cuando pensamos en el fenómeno de la consciencia. Particularmente, la pregunta que compete a la presente investigación es precisamente una pregunta que parece quedar al margen del problema fácil. Se trata de la pregunta acerca de por qué y cómo es que ciertos procesos mentales se ven acompañados de un carácter fenoménico; por qué razón los procesos neurofisiológicos que se desencadenan al probar una copa de vino entrañan, adicionalmente, la *experiencia* de tal sabor o tal textura; cómo es que el procesamiento de la información que, en el caso de ciertas máquinas, parece poder desarrollarse sin ningún carácter fenoménico ni ninguna experiencia en primera persona, en el caso de los seres humanos y presumiblemente de muchos otros organismos, *sí involucra* una consciencia de tipo fenoménico; por qué es que estos procesos neuronales de tráfico y procesamiento de información no ocurren "a oscuras". Este es el problema difícil de la consciencia.

No se trata de un problema difícil en el sentido de que aún no tenemos respuesta para él, o de que tal respuesta implicará complejísimas indagaciones. Si así fuera, se disolvería la distinción entre el problema fácil y el problema difícil. Se trata de un problema difícil en el sentido de que el enfoque que resultaría apropiado para resolver el problema fácil es insuficiente para abordar el problema difícil. Este enfoque es el que consiste en descubrir cuál es el mecanismo en virtud del cual se realiza cierta función. Presentado el mecanismo mediante el cual un sistema realiza determinada función, se agota en gran medida la explicación acerca del fenómeno en cuestión, al menos en lo que refiere a cómo se ejecuta.

En el caso del problema difícil, la explicación de los mecanismos en virtud de los cuales se realizan ciertas funciones mentales no agota la explicación de la consciencia fenoménica. Recordemos el ejemplo de las llaves en el bolsillo: la explicación minuciosa de toda la cadena causal entre el surgimiento de la duda y la ejecución de la acción, podría -al menos en principio- relatarse sin que en ningún paso aparezca como obvia la necesidad de postular un papel para la consciencia fenoménica. En palabras de Chalmers:

> Lo que hace que el problema difícil sea difícil y casi único es que va *más allá* de los problemas acerca de la realización de funciones. Para ver esto, note que incluso una vez que hayamos explicado la realización de todas las funciones cognitivas y conductuales en los alrededores de la experiencia -la discriminación perceptual, la categorización, el acceso interno, el reporte verbal- quizás subsista aún una ulterior pregunta no respondida: *¿por qué la realización de estas funciones está acompañada de la experiencia?* Una explicación simple de las funciones deja esta pregunta abierta. (Chalmers, 1995, p. 5, cursivas en el original)

## 3.6.- La brecha explicativa

El señalamiento del problema difícil de la consciencia está íntimamente ligado con una suerte de hiato epistemológico en el que la distinción entre consciencia-P y consciencia-A vuelve a escena. Se trata de lo que Joseph Levine (1983) llamó "la brecha explicativa". Si fuera cierto que los estados cerebrales son idénticos a los estados mentales (es decir: si fuera verdadera la identidad psicofísica), el aspecto físico de esta identificación no parece poder explicar el aspecto psicológico de la misma.

Tomando el ejemplo de Levine, imaginemos que confirmamos empíricamente que existe un correlato entre la experiencia del dolor y la activación de unas determinadas vías nerviosas (las "fibras C") y sus estructuras cerebrales conexas. Imaginemos también que podemos explicar perfectamente en términos neurofisiológicos cómo y por qué se activan estas fibras y cómo se desencadenan ciertos procesos sinápticos y neuroquímicos a partir de su activación. Desde una perspectiva materialista, deberíamos afirmar que hemos *explicado* la experiencia consciente del dolor mediante la explicación de estos procesos neurofisiológicos. Sin embargo, parece también claro que ninguna explicación sobre cómo se activan las fibras C -explicación que se compondrá en gran medida de transmisiones electroquímicas entre conjuntos de neuronas- contendrá nada parecido a una explicación de por qué el dolor se siente como se siente.

Tal como se ha dicho, esta brecha explicativa está íntimamente ligada con el problema difícil de la consciencia: cómo es posible que procesos que son -a primera vista- puramente físicos, como la actividad electroquímica que tiene lugar en el cerebro, puede dar origen a la consciencia fenoménica. Es importante remarcar la apreciación que hace Block (1995, p. 231) al referirse a la brecha explicativa: esta brecha sólo se verifica cuando intentamos explicar científicamente el aspecto fenoménico de la consciencia; otros aspectos -como el aspecto cognitivo o el aspecto funcional- son objeto de programas de

investigación de los que cabe sensatamente esperar que puedan explicar adecuadamente los fenómenos de los que tratan. Aun con dificultades, existen notorios avances en la investigación sobre cómo el cerebro humano procesa la información, cómo se relacionan diferentes partes y aspectos del mismo para llevar adelante este procesamiento, y cómo algunas de estas relaciones pueden ser modeladas a través de programas informáticos. Se trata, como podemos ver, de los problemas fáciles de la consciencia.

Sin embargo, no parece haber ninguna perspectiva teórica que resulte incuestionablemente promisoria a la hora de explicar el aspecto fenoménico de la consciencia. Tal como fue aventurado por Chalmers, las décadas han pasado y los enormes desarrollos en lo que refiere a los problemas fáciles de la consciencia no han logrado echar luz sobre el problema difícil (Horgan, 2023).

Tanto la distinción entre el problema fácil y el problema difícil, como la brecha explicativa respecto de la consciencia fenoménica, parecen ser dos aspectos del mismo núcleo problemático. Ambos abrevan del argumento de la concebibilidad de Saul Kripke (1980). Se trata de un influyente argumento contra el fisicalismo que se sostiene sobre la idea de que para que exista una verdadera identidad entre dos términos (como la identidad psicofísica), esta debe ser *necesariamente* verdadera: debe ser cierta en todos los mundos posibles. Sin embargo, es concebible la separación entre el término físico y el mental en tanto puede concebirse una activación de las fibras C que no entrañe la experiencia del dolor y también a la inversa. La concebibilidad de esta separación es razón suficiente para negar la *necesidad* de la identidad propuesta.

Hasta aquí, hemos enmarcado la discusión objeto de este trabajo trazando algunas coordenadas sobre los conceptos de consciencia fenoménica, enjuta y función, y hemos podido realizar un recorrido panorámico por algunas funciones posibles que cabría asignarle a la consciencia fenoménica. En este recorrido se indicaron también los problemas generales de la pregunta por la función así como

algunas perspectivas escépticas respecto de la posibilidad de asignarle una función, cualquiera sea ésta. Se ha abordado también el modo en el que la distinción entre consciencia-P y consciencia-A, la distinción entre el problema fácil y el problema difícil, así como la brecha explicativa, de algún modo implican restricciones, fronteras, o incluso límites para las investigaciones a este respecto. Estamos en condiciones, entonces, de abordar con mayor profundidad la hipótesis BAV y analizar los argumentos que pueden construirse en torno a ella.

# 4.- Volviendo sobre la hipótesis BAV

La distinción que trazamos entre consciencia-A y consciencia-P es fundamental en la delimitación de la hipótesis BAV, ya que en el caso de la consciencia-A, podríamos afirmar que en la propia formulación del concepto está incluida la explicitación de su función.

> No nos concierne directamente la consciencia de acceso, esto es: la disponibilidad de información al sistema cognitivo para el cumplimiento de sus funciones (Block, 1995). La consciencia de acceso obviamente es importante y puede que esté estrechamente conectada con la consciencia fenoménica, pero no existe un problema especial a la hora de encontrar el valor adaptativo de la consciencia de acceso porque, por definición, la consciencia de acceso contribuye a las funciones del sistema cognitivo. En contraste, si la consciencia fenoménica contribuye o no con las funciones del sistema cognitivo, es algo que aún no se ha determinado. (Robinson et al., 2015, p. 365)

Teniendo en mente la distinción entre los problemas fáciles y el problema difícil, es sensato afirmar que aún no se ha determinado una función para la consciencia fenoménica, a la luz de las diferentes funciones que se han propuesto. En casi todas las funciones que se han enumerado, parece poderse trazar la distinción entre consciencia-P y consciencia-A, lo cual dejaría indemne la pregunta por al función de la consciencia fenoménica en sí misma. Por otra parte, si la consciencia-A tiene una función tan clara[9], entonces la discusión de Block sobre si la consciencia-P podría ser un aspecto de la consciencia-A que se

9 Parecen aquí referirse a una función en el sentido de función de Cummins, lo cual supone un problema, dado que adscribir una función de Cummins a un determinado rasgo no allana el camino hacia la comprensión de su origen evolutivo, como sí lo haría una teleo-función. Adicionalmente, en otro pasaje del artículo de Robinson y sus colegas se define el concepto de "función biológica" en un sentido claramente teleo-funcional. Como afirma Godfrey-Smith (1996), no tiene por qué plantearse -como se ha hecho a menudo- una contraposición entre ambos conceptos, lo cual no obsta para que haya que analizar con más detalle si este inadvertido doble rasero afecta en algo la solidez de la distinción que pretenden plantear.

encuentra íntimamente ligado a ésta, podría servir para respaldar la hipótesis de que la consciencia fenoménica es un subproducto de otro rasgo (en este caso la consciencia de acceso) que sí tiene una función y presenta una ventaja adaptativa para su portador.

Respecto de la formulación de la hipótesis BAV, parece correcto lo que afirman Robinson y sus colegas respecto de que su enfoque es el primero en formular esta hipótesis de un modo independiente del modelo de consciencia que se adopte; en otros términos, otros autores han planteado ideas similares pero como consecuencia de sus puntos de vista *específicos* respecto de lo que la consciencia es. Al respecto, podemos señalar algunos trabajos también referidos aquí de David Rosenthal (2008) y particularmente de Thomas Polger y Owen Flanagan (2002), en el que se presenta una visión muy cercana a la hipótesis BAV a la que los autores denominan “epifenomenalismo etiológico”.

## 4.1.- Relatos adaptacionistas sobre la consciencia fenoménica

En el artículo de Robinson y sus colegas se da cuenta de un conjunto de lo que llaman “relatos adaptacionistas” sobre el origen de la consciencia. Muchos de ellos fueron referidos en la sección correspondiente a las funciones de la consciencia. Las diferencias entre cada uno de ellos, para los objetivos del artículo, son irrelevantes, en la medida en que comparten -en tanto asumen que se trata de una adaptación- que la consciencia confiere a los organismos una determinada ventaja adaptativa que ha sido seleccionada positivamente a lo largo de las generaciones por los mecanismos propios de la selección natural.

Estos relatos adaptacionistas pueden interpretarse desde dos grandes perspectivas. En una lectura *fuerte*, afirman que existen procesos mentales relevantes que *no pueden* llevarse adelante sin la presencia de la consciencia fenoménica: ésta constituye una condición necesaria para ellos y, por tanto, ningún ser que carezca de consciencia fenoménica podrá desarrollar tales procesos mentales que -en virtud del presupuesto adaptacionista- confieren una ventaja adaptativa al organismo. Uno de los relatos adaptacionistas más sólidos que confiere una función clara a la consciencia fenoménica en un sentido *fuerte* es el que presentan Axel Cleeremans y Catherine Tallon-Baudry, ya mencionado y que se analizará en detalle más adelante.

Existe otro modo de interpretar los relatos adaptacionistas, un modo *débil* que consiste en afirmar que la consciencia es una condición necesaria para el desarrollo de ciertos procesos mentales que no podrían lograr sus objetivos de modo inconsciente, pero esta necesidad no está vinculada con los procesos en sí, sino con la forma en la que estos procesos se realizan *en nosotros*. En otros términos, son algunas características de nuestro ser biológico, o el modo particular en el que los organismos han evolucionado en el planeta Tierra, lo que hace que la consciencia sea una condición necesaria para ciertos procesos mentales. Estas lecturas están relacionadas con las posiciones que hemos definido bajo el rótulo de *inesencialismo* respecto de la consciencia fenoménica.

Un curioso corolario de esta perspectiva es que habilita la posibilidad de que en seres diferentes a nosotros (vida alienígena no basada en carbono, computadoras, sistemas compuestos por habitaciones cerradas con filósofos recluidos dentro siguiendo instrucciones (Searle, 1980), ciudadanos intercambiando señales de radio (Block, 1978), etc.) puedan desarrollarse los mismos procesos mentales cuyo valor adaptativo se postula, pero en ausencia de la consciencia fenoménica. En esta lectura débil de los relatos adaptacionistas, si bien se afirma la existencia de determinada función para la consciencia (aunque

más no sea de una forma *contingente*), es sencillo deslizarse hacia la *deflación* de esta función, muy contraria a nuestra idea intuitiva de la importancia de la consciencia fenoménica. Si la consciencia fenoménica es importante para nosotros porque contingentemente hemos evolucionado *de esta forma* en la que lo hicimos, perfectamente podrían haber evolucionado junto a nosotros otros seres carentes de consciencia fenoménica que desarrollaran con equivalente habilidad las funciones cognitivas que desarrollamos nosotros. Si esta "bifurcación" en la filogenia se hubiera producido una vez comenzado el proceso de hominización (y también si hubiera ocurrido antes), podrían existir zombis entre nosotros. Una lectura débil de los relatos adaptacionistas habilitaría tal conclusión. Como veremos, es problemática la consideración del escenario hipotético de una competencia evolutiva entre nuestros ancestros conscientes y sus contrapartes zombis (Theurer & Polger, 2018).

## 4.2.- Criterios para que un rasgo sea considerado una enjuta o un accidente evolutivo

A la hora de defender la plausibilidad de la hipótesis BAV, Robinson y sus colegas intentan mostrar cómo esta hipótesis cumple con tres criterios no arbitrarios que pueden utilizarse para sostener la posibilidad del carácter de enjuta o de accidente evolutivo de un rasgo fenotípico. Para cumplir con los tres criterios, el rasgo fenotípico en cuestión: a) no puede representar una desventaja adaptativa neta para los organismos que lo exhiben; b) no debe presentar una ventaja adaptativa y, simultáneamente, ser objeto de una explicación plausible en términos de selección natural; y c) debe existir otro rasgo -presumiblemente adaptativo- del cual pueda plausiblemente ser un subproducto.

### 4.2.1.- La consciencia fenoménica no presenta una desventaja neta

Si un rasgo fenotípico *S* es el subproducto de un rasgo adaptativo *A*, o bien se trata de un subproducto *necesario* o de un subproducto *accidental*. Llamamos subproducto *necesario* a aquel rasgo *S* que es una consecuencia ineludible de la presencia de *A.* Por ejemplo, imaginemos que el latido del corazón es una consecuencia ineludible del hecho de que el corazón bombee sangre, en la medida en que dicho bombeo *no puede* realizarse de manera silenciosa. Imaginemos también que el latido cardíaco en una presa puede ser en sí mismo desventajoso en la medida en que puede alertar a un depredador sobre su presencia. Si el corazón bombeante se ha mantenido presente a lo largo de mucho tiempo en la historia evolutiva de la especie, presumiblemente es porque este órgano presenta una ventaja adaptativa muy superior a la desventaja que podría acarrear el sonido generado. Así, si un rasgo *S* presenta una desventaja, esta debe ser menor en términos relativos que la ventaja del rasgo *A*. Si fuera mayor la desventaja del subproducto que la ventaja que confiere el rasgo adaptativo, entonces este rasgo no sería adaptativo en términos netos, constituiría un perjuicio adaptativo para los individuos que lo porten, y finalmente desaparecería de la línea filogenética de la especie producto de una dinámica normal de selección natural negativa. En otros términos, si el latido del corazón amenazara tanto la vida de los organismos que portan corazones en una especie, el corazón en esa especie no habría evolucionado hasta estar presente en la actualidad, en la medida en que la mayoría de los individuos que tuvieran rudimentos de corazón habrían perecido antes, se habrían reproducido menos, y por tanto no hubieran pasado sus genes “para corazón” a la siguiente generación, con la misma tasa que aquellos individuos que hubieran mantenido su sistema de circulación tal como era previo a la innovación cardíaca, o que quienes hubieran desarrollado un sistema de circulación alternativo que prescindiera de corazón y que fuera, por tanto, “silencioso”.

Pero también cabe la posibilidad de que el subproducto S del rasgo adaptativo A sea un subproducto *accidental*. Llamemos subproducto *accidental* a aquel rasgo S que, si bien ha surgido como consecuencia de A, no se trata de una consecuencia necesaria, en la medida en que el rasgo adaptativo A podría estar presente aun en ausencia del subproducto S. El latido del corazón en la especie que estamos imaginando podría constituir un subproducto *de una clase determinada de corazón*, que podría coexistir al interior de la especie con otra variante de corazón que bombea sangre con exactamente la misma eficacia y eficiencia, pero de manera silenciosa. En este caso, si existieran en la especie individuos con corazones "ruidosos" y con corazones "silenciosos" (y si, siguiendo nuestro razonamiento, el tipo de corazón desarrollado fuera la consecuencia de información genética heredable a la siguiente generación), entonces la selección natural presionaría positivamente en la dirección de corazones silenciosos y negativamente en la dirección de corazones ruidosos, al punto tal de que éstos últimos finalmente dejarían de existir.

Ya sea que el subproducto analizado sea necesario o accidental, en ningún caso puede suponer una desventaja adaptativa tan importante que supere a la ventaja adaptativa que confiere el rasgo del cual es un subproducto.

Al parecer, ninguna evidencia permite intuir el carácter desventajoso de la consciencia fenoménica. Salvo -podría argüirse- a partir del hecho de que la consciencia sobre los propios procesos cognitivos podría constituir una desventaja, tal como se ha indicado a partir de los experimentos de Dijksterhuis y sus colegas (2006). Recordemos que estos experimentos indican que a la hora de tomar decisiones complejas, el ser consciente del devenir de los propios procesos deliberativos -no solamente deliberar sobre cuál es el mejor curso de acción, sino observar desde una mirada metacognitiva la forma y el modo en el que estamos emprendiendo esos procesos deliberativos- no solamente no mejora los procesos deliberativos en sí, sino que por el contrario tiende a generar decisiones que son

peores (tanto objetiva como subjetivamente) que las que se hubieran tomado sin este grado de “autoconsciencia”.

Sin embargo, la distinción entre consciencia-P y consciencia-A permitiría mostrar que esta desventaja refiere exclusivamente a la consciencia de acceso, y se trata de una desventaja muy limitada a ciertas circunstancias, que parece superarse con creces por las ventajas adaptativas que este tipo de consciencia confiere en los más variados escenarios.

Asimismo, aun si la consciencia fenoménica implicara en algún sentido y en algunos casos una desventaja adaptativa, esto no implicaría rechazar la hipótesis BAV, siempre y cuando se pudiera confirmar que la consciencia es un subproducto necesario de otro rasgo cuya ventaja adaptativa es superior a esta desventaja.

### 4.2.2.- La consciencia fenoménica no es, a la vez, adaptativamente ventajosa y explicable a través de mecanismos de selección natural

El segundo criterio es que para ser un posible candidato a ser una enjuta o un accidente, un rasgo no debe, simultáneamente, otorgar una ventaja adaptativa obvia a su portador y ser pasible de una explicación verosímil acerca de su origen por mecanismos de selección natural. Si la consciencia fenoménica implicara una ventaja adaptativa y pudiera ser objeto de una explicación mediante selección natural, sería simplemente una adaptación en el sentido clásico del término, como la enorme mayoría de los rasgos fenotípicos relevantes. Son estas las razones por las que las alas de las aves voladoras no podrían ser consideradas una enjuta o un accidente evolutivo de manera sensata. Por un lado, la presencia de alas en las aves voladoras les confiere una ventaja adaptativa notoria. Por otro lado, existe

una hipótesis adaptacionista plausible y respaldada por sólida evidencia empírica que permite reconstruir la historia filogenética entre algunos dinosaurios (particularmente los *celurosaurios*) y las aves modernas, pasando por etapas intermedias como las del género *Archaeopteryx,* relato en el que podemos reconstruir, a partir del registro fósil y la comparación morfológica y genética, numerosas etapas intermedias en las que los huesos se ahuecan, las escamas devienen progresivamente en plumas, y la construcción esquelética general adquiere formas y dimensiones propicias para un desplazamiento arbóreo primero y aéreo después (Ostrom, 1976). En la medida en que estas características están determinadas en gran medida por la información genética que se hereda a la siguiente generación, la selección natural es el mecanismo más adecuado para explicar estos cambios. No hay razón, por tanto, para considerar las alas de las aves voladoras como una enjuta o un accidente evolutivo, sino que deben considerarse como un rasgo adaptativo en toda regla, en tanto provee a sus portadoras de ventajas adaptativas claras y existe un relato plausible de su origen a través de mecanismos de selección natural.

Como hemos podido ver a partir del análisis de las posibles funciones de la consciencia fenoménica, no existe una ventaja obvia (no controversial) que este rasgo otorgue a sus portadores, como tampoco existe un relato adaptacionista sólido que permita dar cuenta de su origen mediante procesos selectivos.

No obstante, es interesante detenerse aquí en una pregunta que podemos formularnos a partir de una apreciación que realiza Chalmers sobre lo que una explicación evolutiva podría y no podría lograr. Para ello, recordemos el concepto de zombi que hemos denominado “zombi tipo b”: una réplica *conductual y física* de un ser humano pero carente de consciencia fenoménica.

> Desafortunadamente, esta idea [de que detectando la función de la consciencia fenoménica podremos entender la razón evolutiva por la que surgió en el mundo] sobrevalora lo que una explicación evolutiva podría proporcionarnos. El proceso de la selección natural no puede

> distinguir entre mi persona y la de mi gemelo zombi. La evolución selecciona propiedades según su papel funcional, y mi gemelo zombi realiza todas las funciones que yo realizo tan bien como yo; en particular deja diseminadas tantas copias de sus genes como yo. Se deduce que la evolución sola no puede explicar por qué evolucionaron creaturas conscientes y no zombis. (Chalmers, 1999, pp. 163-164)

Como hemos visto, la apreciación de Chalmers que cierra la cita esconde un razonamiento polémico. A primera vista, en el caso de que la consciencia fenoménica fuera un subproducto necesario de algún otro rasgo adaptativo (e incluso si fuera un subproducto accidental), la evolución sí podría explicar perfectamente su surgimiento, sólo que no a través de mecanismos de selección natural. En otros términos, si la consciencia fenoménica fuera un subproducto de, pongamos por caso, un cierto grado de integración de la información en el cerebro, la evolución podría explicar perfectamente el surgimiento de sistemas integradores de información, en virtud de su ventaja adaptativa y a través del mecanismo clásico de la selección natural.

La clave radica en distinguir con precisión qué es lo que una explicación evolutiva está explicando. Si un rasgo adaptativo *A* implica la presencia de un subproducto *S*, existen al menos dos formas en las que una explicación evolutiva podría echar luz sobre el surgimiento de S. En primer lugar, una explicación clásica en términos de selección natural podría mostrar por qué el rasgo *A* es adaptativo y por qué la selección natural pudo presionar positivamente su expansión en la filogenia de la especie. Algunos otros elementos vinculados con las constricciones constructivas o con el modo en el que se desarrolla en la embriogénesis el rasgo *A* podrían perfectamente explicar el hecho de que *A* tenga como subproducto a *S*. Sería ésta una explicación si se quiere "completa" de la presencia de *A* y *S*. Por ejemplo, podemos explicar a través del mecanismo de la selección natural la presencia de corazones en virtud de su ventaja adaptativa y de la heredabilidad de este rasgo a través de la información genética. Asimismo, la física permite explicar por qué el funcionamiento del corazón implica la

generación de cierto sonido, mostrando el modo en el que la sangre atravesando rápidamente canales estrechos de cierta manera y con cierta frecuencia produce inevitablemente un determinado sonido. Se explica así la presencia del rasgo adaptativo *A* (a través de mecanismos de selección natural) y de su subproducto *S* (a través de las leyes de la física).

Sin embargo, podría existir una explicación evolutiva que podríamos llamar "incompleta". Supongamos que podemos determinar con claridad por qué el rasgo adaptativo *A* fue seleccionado positivamente, y supongamos que podemos también tener una relativa certeza de que el subproducto *S* es efectivamente un subproducto de *A*, por ejemplo, porque podemos confirmar empíricamente que quienes portan el rasgo *A* también portan el subproducto *S*, o porque podemos confirmar experimentalmente que suprimiendo de manera artificial el rasgo *A* se suprime también el suproducto *S*. Puede ocurrir, no obstante, que no tengamos claro cómo exactamente es que el rasgo *A* da origen al subproducto *S*.

Imaginemos que en una determinada familia de pinzones la forma puntiaguda del pico se ve indefectiblemente acompañada de una determinada mancha azul en el plumaje de la cola. Imaginemos también que la forma puntiaguda del pico implica una ventaja adaptativa clara (por ejemplo, por el tipo de fruto que compone la dieta de estos pinzones) y que su surgimiento y permanencia en la familia puede rastrearse a través del registro fósil o a través de la comparación genética. Observamos que todos los ejemplares que por alguna razón nacen con un pico no puntiagudo, carecen también de la típica mancha azul. A través de complejos procedimientos, detectamos cuál es el gen o el grupo de genes que determina la forma puntiaguda del pico, y logramos experimentalmente incubar pinzones en los que estos genes han sido suprimidos, sólo para encontrarnos con el esperable resultado de que estos pinzones nacen también sin ninguna mancha azul en su plumaje. Imaginemos también que podemos experimentalmente mostrar que la mancha azul, en sí misma, es adaptativamente

neutra: no implica ni una ventaja ni una desventaja para los individuos que la portan. Nuestra explicación evolutiva de la mancha azul está, pues, casi completa. Sabemos que es un subproducto del pico puntiagudo y sabemos por qué el pico puntiagudo se mantuvo en la filogenia de esta familia en virtud de su ventaja adaptativa. Lo que falta en esta explicación para que sea tan completa como en el ejemplo anterior, es precisamente una explicación (presumiblemente genética) de por qué la forma del pico se asocia biológicamente a la presencia o no de la mancha azul en la cola. Desde luego, la hipótesis más sólida es que los genes que regulan el rasgo adaptativo regulan también la presencia del subproducto. Sin embargo, hasta que no tengamos una cadena causal clara que parta de los genes que sabemos regulan la forma del pico, y nos conduzca a través de rutas bioquímicas a la generación del pigmento azul en determinadas plumas, nuestra explicación estará incompleta. Más aún: el tipo de explicación que puede servir para explicar el rasgo *A* (la selección natural) puede no servir para explicar la presencia del subproducto *S*. Esto es precisamente lo que ocurre, según Chalmers, con las explicaciones evolutivas sobre la consciencia fenoménica[10].

Aun asumiendo la hipótesis BAV sobre la consciencia fenoménica, esto de ninguna manera agota la explicación de su existencia, como sí podría pasar con otros rasgos fenotípicos como el sonido que produce el latido del corazón. Tomemos por ejemplo la idea de que la consciencia fenoménica es un subproducto de un alto nivel de integración de la información. Podríamos afirmar que un alto nivel de integración de la información en un sistema como el cerebro humano es un rasgo adaptativo y, en tanto responde a la organización interna de las redes neuronales cuya arquitectura está en gran medida determinada genéticamente, es perfectamente un rasgo que puede explicarse a través de mecanismos de selección natural. Sin embargo, esta parte de la explicación solamente explica por qué existen cerebros con un alto grado de integración de la

10 Como veremos más adelante y tal como se anunció al introducir el concepto de zombi, existen fuertes razones para dudar de los razonamientos que parten de la hipótesis de una competencia evolutiva entre seres conscientes y sus gemelos zombis (Theurer & Polger, 2018).

información, pero no explica por qué un alto grado de integración de la información da origen a la consciencia fenoménica. Es más: las leyes o principios que determinan que una determinada configuración de un sistema físico dé origen (necesaria o accidentalmente) a la consciencia fenoménica, perfectamente pueden ser leyes independientes de las que regulan la evolución por selección natural, del mismo modo que la física de los fluidos que explica el sonido del latido, es *epistemológicamente* diferente de las leyes estadísticas y biológicas que permiten explicar el carácter adaptativo de los corazones.

En palabras de Chalmers:

> (…) nótese que el verdadero problema de la consciencia es explicar los principios en virtud de los cuales ésta surge en los sistemas físicos. Es de suponer que estos principios -sean verdades conceptuales, necesidades metafísicas o leyes naturales- son constantes en el espacio y el tiempo: si una réplica física de mí hubiese surgido a la existencia hace un millón de años, habría sido exactamente tan consciente como yo. Los principios conectores mismos son por lo tanto independientes del proceso evolutivo. Aunque la evolución puede ser muy útil para explicar por qué evolucionaron sistemas físicos particulares, es irrelevante para la explicación de los principios puente en virtud de los cuales algunos de estos sistemas son conscientes. (Chalmers, 1999, p. 164)

Desde luego, esta posición está íntimamente vinculada con el cuestionamiento al fisicalismo que implica la hipótesis de los zombis, lo que ameritará un abordaje particular cuando evaluamos la relación entre la hipótesis BAV y diferentes posiciones metafísicas respecto de lo mental.

Baste por ahora indicar que la consciencia fenoménica no presenta una ventaja adaptativa clara (no controversial) para sus portadores ni existe un relato adaptacionista convincente sobre su surgimiento, condiciones que -de cumplirse- la descartarían de plano como una posible enjuta o subproducto.

### 4.2.3.- Debe haber otro rasgo del cual es un subproducto

En tercer y último lugar, para que la consciencia fenoménica sea una buena candidata a ser una enjuta o subproducto, es relevante que contemos con al menos una hipótesis de cuál puede ser el rasgo del cual la consciencia fenoménica es un subproducto. Retomando nuestro ejemplo de los picos de los pinzones, aún si ignoramos los detalles de la relación causal entre la forma del pico y la mancha azul, otros indicios apuntaban a que este rasgo es un subproducto de aquel. En este ejemplo elaborado arbitrariamente, la conexión seguramente sea de corte genético, en tanto se trata de aspectos del organismo considerablemente diferentes el uno respecto del otro. En el caso de la consciencia fenoménica, por el contrario, los rasgos que pueden ser buenos candidatos a ser el rasgo *A* del cual surge el subproducto *S* están, de acuerdo a algunas teorías, íntimamente ligados.

En este caso, dos perspectivas parecen aportar buenos argumentos para cumplir con este criterio: la consciencia fenoménica puede ser un subproducto de sistemas físicos con un altísimo grado de complejidad (Edelman, 1989); o bien la consciencia puede ser un subproducto de sistemas con una altísima integración de la información (Tononi & Koch, 2015).

En ambos casos, estamos ante un problema que podemos visualizar a partir de algunos conceptos que ya hemos abordado. Por un lado, estos son buenos candidatos a ser el rasgo adaptativo *A* del cual la consciencia fenoménica puede ser un subproducto, en la medida en que desde ambas perspectivas se ofrecen argumentos para mostrar que la consciencia fenoménica *acompaña* los sistemas físicos con un alto grado de complejidad o de integración de la información. Sin embargo, podría argüirse a este respecto que se trata de una explicación incompleta, en tanto la detección de una correlación no agota la comprensión de la conexión causal entre estos dos aspectos. Aun si confirmáramos que un sistema con un altísimo grado de integración de la información es fenoménicamente

consciente, esto no implica necesariamente entender por qué esto es así. Estamos ante el límite que impone la distinción entre los problemas fáciles y el problema difícil y que puede aplicarse a este caso también: no es claro el modo concreto en el que un sistema con un altísimo grado de integración de la información tenga que ser (o pueda ser) fenoménicamente consciente.

No obstante, en lo que respecta a la hipótesis BAV, es relevante el hecho de que existan *buenos candidatos* a ser el rasgo adaptativo A del cual la consciencia fenoménica podría ser un subproducto. Si eventualmente lográramos confirmar que la consciencia fenoménica es *efectivamente* el subproducto de un alto grado de complejidad o de un alto grado de integración de la información, entonces la hipótesis BAV dejaría de ser una hipótesis y se convertiría en una "verdad" científica a la manera de la relación entre el funcionamiento cardíaco y el sonido que produce. Hasta tanto eso no ocurra, el contar con buenos candidatos es un argumento a favor de la hipótesis, del mismo modo que la ausencia total de candidatos constituiría un argumento en contra.

## 4.3.- La hipótesis BAV y su relación con la metafísica de lo mental

Como hemos visto, la hipótesis BAV parece poderse aplicar de forma plausible a la consciencia fenoménica, en tanto ésta cumple con varios criterios no arbitrarios que en general tomamos en consideración a la hora de determinar si un determinado rasgo fenotípico podría o no ser una enjuta o un accidente evolutivo.

Adicionalmente, existe otro enfoque desde el cual puede sustentarse tal hipótesis, un enfoque que nos aleja ligeramente de discusiones propiamente evolutivas, y nos acerca a consideraciones que indefectiblemente atraviesan la casi

totalidad de las discusiones sobre la consciencia en general y sobre la consciencia fenoménica en particular. Se trata del profuso y sofisticado debate en torno al carácter metafísico de lo mental.

Desde la forma en la que René Descartes traza las líneas generales del problema mente-cuerpo hasta la actualidad, el debate sobre la metafísica de lo mental no ha hecho más que adquirir grados crecientes de complejidad y sofisticación, en gran medida abrevando, reformulando o rebatiendo las coordenadas que la formulación cartesiana del problema delimitó hace ya casi cuatro siglos (Lassevich, 2020). No es necesario, sin embargo, ahondar en las insondables profundidades de este debate para exponer la forma en la que la hipótesis BAV se relaciona lógicamente con diferentes posiciones respecto de la metafísica de lo mental. El objetivo de las consideraciones que siguen no es arribar a ninguna conclusión irrebatible en relación con estas posiciones, sino mostrar a través de ellas y desde una nueva perspectiva, la plausibilidad de la hipótesis.

### 4.3.1.- El epifenomenalismo implica la hipótesis BAV

Hemos definido el epifenomenalismo metafísico como aquella posición en virtud de la cual los estados mentales conscientes no tienen poderes causales de ningún tipo, es decir, son causalmente inertes. La hipótesis BAV no necesariamente implica aceptar el carácter causalmente inerte de la consciencia fenoménica, sino solamente afirmar que la consciencia fenoménica, en caso de tener efectos causales, no los tiene sobre la conducta del organismo en términos de una mayor adaptación evolutiva que pudiera ser objeto de la selección natural. Como hemos visto, se le puede llamar a esta posición *epifenomenalismo biológico*. Si bien es posible pensar en muchos rasgos biológicos que sean neutros

respecto de la ventaja adaptativa, es díficil pensar en alguno que sea causalmente inerte: la superficie de la cóclea con su color amarillento, absolutamente inerte respecto de la adaptación del organismo, tiene sin embargo efectos causales, en la medida en que refracta la luz de una determinada manera si la examinamos a través de un procedimiento quirúrgico. Por tanto, aceptar la hipótesis BAV no nos compromete necesariamente con una posición epifenomenalista, al menos en el sentido metafísico del término.

Sin embargo, aceptar una posición epifenomenalista sí nos compromete con la hipótesis BAV. Si el epifenomenalismo está en lo correcto, entonces la consciencia fenoménica es causalmente inerte y no tiene efectos causales sobre ningún otro aspecto de la realidad, incluidos otros procesos cerebrales o el comportamiento de los organismos. Si la consciencia fenoménica no tiene ningún efecto causal sobre el comportamiento, entonces no puede haber sido objeto de la selección natural, en tanto la selección natural opera solamente sobre aquellos rasgos que implican -directa o indirectamente- una variación en la morfología, el funcionamiento o el comportamiento de los organismos que redunde, a su turno, en una diferencia en la cantidad de copias de sus genes que se heredan a la siguiente generación.

Si la consciencia fenoménica es causalmente inerte y, por tanto, no puede haber sido objeto de la selección natural, entonces la explicación de su presencia en los organismos desde un punto de vista evolutivo tiene que deberse a que se trata de un subproducto causalmente inerte o un accidente evolutivo causalmente inerte. En consecuencia, el epifenomenalismo implica la hipótesis BAV, siempre y cuando mantengamos que los rasgos biológicos o bien son adaptaciones, o bien son enjutas, o bien son accidentes evolutivos.

Por el contrario, aceptar la hipótesis BAV no nos compromete necesariamente con una posición epifenomenalista, en la medida en que puede afirmarse que la consciencia fenoménica no tiene efectos causales en el

comportamiento adaptativo del organismo, sin por ello comprometernos con la afirmación de que es causalmente inerte en términos absolutos. Es más, la hipótesis BAV puede contemplar también la posibilidad de que la consciencia fenoménica tenga efectivamente efectos sobre el comportamiento adaptativo en un sentido ligeramente desventajoso, pero que esta desventaja se vea superada por la ventaja que le otorga al individuo el rasgo de base que le da origen a la consciencia fenoménica, cualquiera sea éste.

### 4.3.2.- El dualismo de propiedades implica la hipótesis BAV

El dualismo de propiedades es la posición metafísica en virtud de la cual las propiedades que constituyen la consciencia son propiedades no físicas, esto es: ontológicamente diferentes de las propiedades físicas. Esta posición se distingue del dualismo ontológico *clásico* en la medida en que no se compromete con la existencia de dos "substancias" ontológicamente distintas, sino con una única materia que puede poseer propiedades de distinto tipo.

El dualismo de propiedades es compatible con el epifenomenalismo, en tanto las propiedades mentales bien podrían ser causalmente inertes; pero también es compatible con la negación del epifenomenalismo, en tanto estas propiedades podrían tener efectos causales. Sin embargo, dados los argumentos de exclusión causal, en general quienes adhieren al dualismo de propiedades aceptan también el epifenomenalismo que parece implicar.

Con argumentos *de exclusión causal* nos referimos principalmente a los argumentos metafísicos que parten de la clausura causal del mundo físico: afirman que todo hecho físico tiene una causa física suficiente y rechazan la sobredeterminación causal sistemática. Si esto es así, entonces, dada una

propiedad física *E* que tiene una causa física suficiente *C*, no puede haber ninguna otra causa no física *C'* que sea también causa de *E*. De otro modo, tendríamos que aceptar la sobredeterminación sistemática del mundo físico, en el que los fenómenos físicos tienen una causa física que les es suficiente y, asimismo, una causa no física que también les es suficiente (Jaewgon Kim, 2000). Hemos tocado tangencialmente este argumento al discutir si una de las funciones de la consciencia fenoménica puede ser la de oficiar como causa de nuestras decisiones voluntarias. En este caso, el problema de la sobredeterminación es compatible con la hipótesis de que nuestros estados mentales conscientes no son la causa real de nuestras acciones, sino tan solo las causas aparentes, producidos a su turno por hechos físicos que son la causa tanto del estado mental consciente como de la acción.

Volvamos al ejemplo de las llaves y, a efectos de simplificar, consideremos como efecto *E* la acción motora de cambiar las llaves de bolsillo. Dada la hipótesis de la clausura causal del mundo, la causa física suficiente $C_f$ debe ser el estado cerebral que le dio origen físico-químico a ese impulso nervioso que recorrió mis nervios motores. Este estado cerebral, a su vez, está acompañado de una percepción consciente de la voluntad de cambiar las llaves de lugar: esta voluntad tiene aspectos fenoménicos en tanto se siente de determinada manera. Desde la perspectiva del dualismo de propiedades, podría argüirse que esta voluntad es la causa no física suficiente $C_{nf}$ de la acción motora. Sin embargo, implica una sobredeterminación sistemática de todas nuestras acciones el hecho de que sean efectos con una doble causa suficiente $C_f$ y $C_{nf}$. Por tanto, el dualismo de propiedades que asuma la clausura causal del mundo físico debe aceptar que las propiedades no físicas son causalmente inertes, so pena de tener que aceptar una muy implausible sobredeterminación causal sistemática. Como hemos mencionado, existe evidencia empírica que respalda esta línea de razonamiento (Wegner & Wheatley, 1999).

En virtud de este argumento, aceptar la hipótesis del dualismo de propiedades nos conduce a aceptar que las propiedades no físicas -en tanto propiedades ontológicamente diferentes de las propiedades físicas- deben ser causalmente inertes. Si esto es así, entonces las propiedades no físicas no pueden ser objeto de la selección natural, en la medida en que ésta opera únicamente sobre aquellas propiedades que tienen un efecto físico sobre la realidad, particularmente sobre la cantidad de copias de determinados genes que se heredan a la siguiente generación de organismos. Por tanto, el dualismo de propiedades que -en virtud del argumento de exclusión causal- acepta que las propiedades vinculadas con la consciencia fenoménica son causalmente inertes, también implica desde un punto de vista lógico a la hipótesis BAV. Por las mismas razones que hemos aplicado al epifenomenalismo en general, esto no es así a la inversa: aceptar la hipótesis BAV no nos compromete necesariamente con el dualismo de propiedades.

### 4.3.3.- El fisicalismo no reduccionista implica la hipótesis BAV

El argumento de exclusión causal también puede aplicarse a la hora de visualizar la relación lógica entre el fisicalismo no reduccionista y la hipótesis BAV. Con *fisicalismo no reduccionista* nos referimos a un conjunto de posiciones metafísicas respecto de lo mental en virtud de las cuales la realidad es enteramente física, pero las propiedades y los fenómenos mentales, entre ellos la consciencia fenoménica, son propiedades de segundo orden que *emergen* o *sobrevienen* sobre las propiedades y fenómenos estrictamente físicos del cerebro, sin por eso reducirse *completamente* a ellos.

Cuando analizamos, por ejemplo, una partida de ajedrez, nos vemos en la necesidad de apelar a ciertas reglas y conceptos que nos permiten explicar las

jugadas o fenómenos que se verifican sobre el tablero. Estas propiedades emergentes surgen sobre la base de propiedades enteramente físicas, como la composición molecular de las piezas, el rozamiento sobre el tablero, las velocidades, aceleraciones y direcciones de los desplazamientos, entre otras. Sin embargo, parece claro que las propiedades *fundamentales* de la física (la masa, la carga, etc.) no son suficientes para dar cuenta de manera completa de los fenómenos de segundo orden que implica la partida de ajedrez. Desde esta perspectiva, sin necesidad de postular la existencia de propiedades no físicas en las partículas atómicas y subatómicas que componen las piezas, podemos sí afirmar que el sistema "partida de ajedrez" exhibe ciertas propiedades emergentes cuya complejidad no puede ser reducida *completamente* a las propiedades físicas fundamentales de las partículas que lo componen.

No podríamos estirar el ejemplo más allá de este límite sin dificultad, ya que las reglas que rigen el ajedrez son convencionales y, por tanto, arbitrarias. Si pensamos en fenómenos mentales o incluso en otros fenómenos biológicos como la homeostasis (la conservación dinámica de los valores fisiológicos dentro de un cierto margen a pesar de los cambios exteriores al organismo), podemos visualizar cómo se puede mantener en pie la tesis fisicalista desde esta vertiente no reduccionista, en la medida en que las leyes o principios puente que ligan las propiedades fundamentales a las propiedades de segundo orden (la consciencia fenoménica o la homeostasis) pueden perfectamente ser leyes o principios enteramente naturales.

De este modo, el fisicalismo no reduccionista también puede ser analizado a la luz del argumento de exclusión causal: en la medida en que las propiedades de segundo orden son *ontológicamente diferentes* de las propiedades fundamentales, puede advertirse el mismo problema de la sobredeterminación sistemática salvo que aceptemos que estas propiedades de segundo orden son causalmente inertes, del mismo modo que lo son las propiedades no físicas en la perspectiva del

dualismo de propiedades. Así, si la consciencia fenoménica es una propiedad de segundo orden que emerge o superviene sobre propiedades físicas vinculadas con la neurofisiología del cerebro, entonces -por el argumento de exclusión causal- se trata de un epifenómeno causalmente inerte y, por tanto, no pudo haber sido objeto de la selección natural.

Nuevamente, podemos ver que el fisicalismo no reduccionista implica la hipótesis BAV, sin necesidad de que la aceptación de la hipótesis BAV nos comprometa con esta posición metafísica en particular.

### 4.3.4.- Algunas intuiciones funcionalistas implican la hipótesis BAV

Como hemos mencionado, el funcionalismo es la posición en virtud de la cual los estados mentales son esencialmente estados funcionales, es decir: pueden ser individualizados y descriptos en términos de las relaciones funcionales que presentan con estímulos, respuestas y con otros estados mentales.

Existen algunas formulaciones del funcionalismo que definen a las relaciones funcionales que individualizan ciertos estados mentales como relaciones que presentan un valor adaptativo para el organismo que las presenta (Sober, 1990). En estos casos, dada la propia definición, la hipótesis BAV se ve excluida. Sin embargo, el hecho de que algunas relaciones funcionales que definen estados mentales puedan presentar un valor adaptativo, no implica necesariamente que *todas* las relaciones funcionales que definen a *todos* los estados mentales tengan un valor adaptativo. De este modo, cabe la posibilidad de que aquellas relaciones funcionales que definen los estados mentales fenoménicamente conscientes sean precisamente aquellas que no presentan un

valor adaptativo, por lo cual la hipótesis BAV podría sostenerse aun desde estas perspectivas.

Sin necesidad de comprometernos con una vertiente concreta en el interior de las perspectivas funcionalistas, sí es posible recoger lo que Robinson y sus colegas denominan “intuiciones funcionalistas”, de las cuales hemos hablado al indicar el problema de la realización múltiple. Desde estas intuiciones, puede imaginarse una entidad funcionalmente equivalente a un determinado organismo consciente que, sobre un substrato material diferente, ejecute las mismas funciones que su contraparte. Imaginemos, por ejemplo, que podemos definir los estados fenoménicamente conscientes de un organismo como relaciones funcionales entre los estímulos que recibe del entorno, otros estados mentales, y las conductas que ejecuta como respuesta a estas circunstancias. Así, definidos los estados mentales conscientes en términos fenoménicos, podríamos también imaginar una entidad (biológica o no) que responda exactamente de la misma forma ante exactamente los mismos estímulos, pero sin que tengamos por qué postular en ella ningún atisbo de consciencia fenoménica. Este ejemplo se corresponde con los zombis que hemos enumerado como “zombi tipo a” y “zombi tipo c”. De este modo cualquier función llevada adelante por un organismo consciente podría ser llevada adelante por una réplica funcional pero no consciente de dicho organismo. En la medida en que la selección natural puede operar solamente sobre aquellos aspectos de los organismos que comportan una diferencia en términos morfológicos o conductuales, pero es ciega a los procesos internos disímiles que pueden dar origen a respuestas conductuales idénticas, se puede deducir la hipótesis BAV a partir de estas intuiciones funcionalistas.

Como veremos, esta relación entre ciertas intuiciones funcionalistas y la hipótesis BAV puede ser fuertemente controvertida.

### 4.3.5.- El fisicalismo reduccionista es compatible con la hipótesis BAV

Hemos enumerado hasta ahora algunas perspectivas metafísicas sobre lo mental que parecen implicar la hipótesis BAV. Otra importante perspectiva en el debate metafísico sobre la consciencia en general y sobre la consciencia fenoménica en particular es el fisicalismo reduccionista, también llamado materialismo reduccionista, en virtud del cual todos los estados, fenómenos y propiedades mentales pueden reducirse a estados físicos -particularmente neurofisiológicos- y sus propiedades físicas. Las llamada *teorías de la identidad* (en el marco de una posición fisicalista en general) pueden ser entendidas como un paso más en esta dirección, en la medida en que postulan la total identidad entre estados mentales y estados físicos: un estado mental determinado *es una y la misma cosa* que un estado cerebral determinado.

El fisicalismo reduccionista no necesariamente implica desde un punto de vista lógico la hipótesis BAV; sin embargo, es completamente compatible con ella. Postular la identidad entre estados mentales y estados cerebrales o físicos no implica necesariamente afirmar que todos estos estados cerebrales tienen una función adaptativa que pueda haberlos hecho objeto de la selección natural. Esto es así también para los estados mentales fenoménicamente conscientes, considerados éstos como un tipo de estado mental particular. Por tanto, aun aceptando la teoría de la identidad, esto no contradice la posibilidad de que los estados cerebrales que identificamos con los estados mentales fenoménicamente conscientes constituyan accidentes evolutivos o enjutas, lo cual deja en pie la hipótesis BAV respecto de la consciencia fenoménica.

> En síntesis, la hipótesis BAV es compatible con todas las perspectivas metafísicas sobre la consciencia [fenoménica] y se desprende de manera plausible de varias perspectivas metafísicas populares: el epifenomenalismo, el dualismo de propiedades, algunas versiones del fisicalismo no reduccionista y algunas versiones del funcionalismo. Si la

hipótesis BAV es falsa, cualquier perspectiva metafísica que implique esta hipótesis está en serios problemas. Por tanto, la hipótesis BAV debería ser tomada en serio. (Robinson et al., 2015, p. 381)

# 5.- Eludiendo la determinación de una función

Al abordar la plausabilidad de la hipótesis BAV, se enumeraron algunas condiciones que debería reunir un determinado rasgo para poder ser candidato a ser considerado una enjuta o un accidente evolutivo: a) que este rasgo no presente una desventaja adaptativa neta, b) que no presente simultáneamente una ventaja adaptativa clara y una explicación plausible de su origen en términos de selección natural, y c) que exista una hipótesis plausible sobre cuál es el rasgo primario del cual es un subproducto. Si nos centramos en la segunda de estas condiciones, podemos pensar en aquello que hemos llamado "relatos adaptacionistas": narrativas más o menos verosímiles sobre cómo pudo haberse moldeado una determinada característica de una especie como consecuencia de la variabilidad genética y de la presión selectiva de la selección natural sobre algunos de los alelos que conforman esta variabilidad. Sin embargo, los relatos adaptacionistas -como parte de lo que Gould y Lewontin llamaban el "programa adaptacionista"- constituyen una estrategia de investigación que puede ser cuestionada, y cuyo cuestionamiento es relevante a la hora de considerar la hipótesis BAV.

> Desafortunadamente, un procedimiento común entre los evolucionistas no permite tal rechazo definible [al programa adaptacionista] por dos razones. En primer lugar, el rechazo de un relato adaptacionista usualmente conduce a su remplazo por otro, antes que a la sospecha de que se requiere otro tipo de explicación. Dado que el espectro de los relatos adaptacionistas es tan amplio como fértiles son nuestras mentes, siempre pueden ser postulados nuevos relatos. Y si un relato no está disponible inmediatamente, siempre se puede alegar ignorancia temporal y confiar en que aparecerá en el futuro (…). En segundo lugar, los criterios para aceptar un relato adaptacionista son tan laxos que muchos de ellos se aceptan sin una confirmación adecuada. A menudo, los evolucionstas utilizan la *consistencia* con la selección natural como el único criterio y consideran su trabajo culminado cuando construyen un relato plausible. Pero siempre pueden contarse relatos plausibles. La clave para la investigación histórica

> reside en divisar criterios para identificar las explicaciones adecuadas entre el conjunto de caminos plausibles para cualquier resultado moderno. (Gould & Lewontin, 1979, pp. 153-154)

De modo que los relatos adaptacionistas sobre el surgimiento de la consciencia fenoménica, con independencia de cual sea su contenido concreto, deben poder constituirse en algo más que meros relatos verosímiles para poder ser considerados como una explicación adecuada en términos de selección natural. Para ello, deben existir evidencias que respalden el relato.

Consideremos por ejemplo el pelaje blanco en los osos polares. Puede intuitivamente construirse un relato adaptacionista que pretenda explicar el surgimiento y expansión de este rasgo en la especie. En este relato se podría identificar una especie como especie ancestral entre los osos polares y, pongamos por caso, los osos pardos de zonas boreales. Se podría imaginar que algunos de esos osos, en virtud de su expansión geográfica, fueron habitando zonas cada vez más nevadas. Puede asumirse que el color del pelaje es un rasgo genéticamente controlado y, por tanto, sujeto a la variabilidad genética producto de mutaciones espontáneas, lo que hace que dentro de la especie ancestral existan osos con diferentes tonalidades de pelaje, algunos más oscuros y otros más claros. Puede imaginarse que los osos portadores de un pelaje más claro fueron más exitosos en cuanto a su supervivencia y reproducción en zonas más nevadas como consecuencia de su capacidad de camuflarse ante sus posibles presas. Puede asumirse que la selección natural favoreció durante cientos de miles de años a aquellos individuos con pelajes más claros respecto de los más oscuros, y cada vez que por mutación genética surgían individuos más claros que el promedio, eran éstos quienes estadísticamente dejaban más copias de sus genes en la siguiente generación, siempre que se encontraran en zonas nevadas. Y finalmente, puede imaginarse que la divergencia geográfica se vio acompañada de un proceso de divergencia genética (al que llamamos *especiación peripátrida* si no hay barreras geográficas sólidas) que condujo finalmente al establecimiento de dos

especies independientes que no pueden cruzarse entre sí. Hemos construido, de este modo, un relato adaptacionista perfectamente plausible para el pelaje blanco de los osos polares.

Sin embargo, este relato no sería más que una mera narrativa plausible si no fuera porque se encuentra acompañado de otro cúmulo de evidencias sistemáticamente consistentes con él. Los ensayos que permiten establecer la ventaja comparativa que ofrece el pelaje blanco a sus portadores, así como la evidencia sobre el carácter genéticamente controlado y por tanto heredable del color del pelaje, se suman a la secuenciación genética que permite visualizar el proceso de especiación a lo largo del tiempo, su consistencia con los registros fósiles y la anatomía comparada entre especies cercanamente emparentadas entre sí, y concomitantemente el modo en el que estos registros fósiles y las especies actuales se distribuyen geográficamente de manera consistente con este proceso de especiación hipotetizado, constituyen evidencias sólidas que respaldan el relato adaptacionista y lo convierten en algo más que sólo una historia plausible.

No obstante, al virar nuestra mirada desde rasgos como el color del pelaje a rasgos como la presencia de la consciencia fenoménica, el escenario cambia radicalmente, y la reconstrucción que habría que realizar para poder dar cuenta de su origen evolutivo presenta una naturaleza esquiva.

> Es concebible que puedan haber evolucionado creaturas tan capaces como nosotros y más o menos indistinguibles conductualmente de nosotros que puedan arreglárselas sobre la única base de la sensibilidad informacional y procesos inconscientes de alto nivel para la resolución de problemas (por ejemplo, a la manera en la que una computadora juega al ajedrez) sin nunca experimentar absolutamente nada. Nosotros, sin embargo, no evolucionamos de esa manera, y por tanto la consciencia surgió a la existencia. La naturaleza del tejido cerebral nos garantiza que no habrá ningún registro fósil de la consciencia fenoménica. El proyecto de reconstruir los orígenes causales de la consciencia tal como la conocemos nos involucra en lo que Dennett (1986) adecuadamente describe como “arqueología de software”. Necesitamos reconstruir un relato creíble a partir de la estructura y función del cerebro actual y a partir de la evidencia

> tangible acerca de cómo nuestros ancestros no humanos y homínidos vivían. La historia verdadera y completa quizás nos eluda eternamente. (Flanagan, 1993, p. 45)

Sin embargo, existe una forma de tomar un rodeo alrededor de estas dificultades. Este rodeo consiste en dejar momentáneamente de lado la pregunta por la función de la consciencia fenoménica y enfocarnos en la pregunta por si la consciencia fenoménica es o no una adaptación. Parece claro que responder a la primera pregunta indicando una función de la consciencia fenoménica implica una respuesta afirmativa a la segunda pregunta, en tanto es difícil pensar en un rasgo que posea una función en términos de mejora de la adaptabilidad y sin embargo no sea una adaptación. Esto es así tanto si pensamos en una función en el sentido teleonómico como en el sentido de Cummins. Sin embargo, es posible abordar directamente la pregunta acerca de si la consciencia fenoménica es o no una adaptación, sin comprometerse necesariamente con ninguna hipótesis acerca de cuál puede ser la función que cumple. Esta operación es de sumo interés para el análisis de la hipótesis BAV, ya que si pudiera establecerse con claridad que la consciencia fenoménica es una adaptación sin establecer cuál puede ser su función, esto implicaría refutar la hipótesis sin tener que responder a los numerosos problemas que suscita cada una de las funciones que se han propuesto para la consciencia fenoménica. Esta línea de análisis es la que siguen Shaun Nichols y Todd Grantham (2000).

En su artículo, Nichols y Grantham se proponen enfocar este asunto desde una perspectiva estrictamente biológica, esto es, a través de herramientas y criterios que suelen utilizarse en biología evolutiva para considerar las adaptaciones, y no a partir de consideraciones metafísicas sobre la consciencia fenoménica. Este enfoque, no obstante, tiene a juicio de los autores consecuencias de carácter metafísico, por ejemplo, estableciendo argumentos que socavan las bases del epifenomenalismo de un modo si se quiere indirecto.

En lo que respecta al campo de la psicología evolucionista[11], los autores relevan dos métodos centrales a la hora de determinar si un rasgo mental es o no una adaptación, y evalúan cómo estos métodos pueden ponerse en práctica para enfocar esta pregunta respecto de la consciencia fenoménica en particular: los argumentos de tipo *comparativo* y los argumentos que se enfocan en la *complejidad*.

## 5.1.- Argumentos comparativos

El método comparativo es un método al que puede apelarse en muchos casos para establecer el carácter adaptativo de un determinado rasgo. En general se funda en dos clases de comparaciones. En primer lugar, pueden establecerse comparaciones entre especies cercanamente emparentadas entre sí pero que difieren en determinados rasgos morfológicos o conductuales, particularmente cuando estas diferencias se corresponden con diferencias ecológicas, esto es: relativas a la dieta, a los depredadores potenciales, al tipo de hábitat, etc. Si dos especies cercanas entre sí difieren en un rasgo morfológico (la forma del pico) y esta diferencia se corresponde también con una diferencia en la dieta (el tipo de fruto del que se alimentan), esto puede constituir un argumento muy sólido a favor de la hipótesis de que la forma del pico es una adaptación. En segundo lugar, pueden establecerse comparaciones entre especies lejanamente emparentadas pero que comparten un determinado nicho ecológico. De este modo, si especies lejanamente emparentadas como el oso polar y el zorro ártico -animales cazadores que habitan zonas heladas o nevadas- comparten el rasgo de tener un pelaje blanco

11 Aquí la traducción puede inducir a confusión. En la tradición hispanohablante, llamamos "*psicología evolutiva"* al análisis de cómo se desarrolla la psicología de un individuo a lo largo de su vida a medida que transita por diferentes estadios de desarrollo. Por otra parte, llamamos "*psicología evolucionista"*, al análisis de cómo se ha desarrollado la psicología humana a lo largo de la historia evolutiva de la especie, desde el marco de análisis de la teoría de la evolución.

y una capa gruesa de tejido adiposo, es sensato pensar que estas características constituyen adaptaciones que surgieron en cada una de sus respectivas líneas filogenéticas al colonizar paulatinamente zonas geográficas similares. El primer tipo de comparaciones -las realizadas entre especies cercanamente emparentadas que difieren en un determinado rasgo- es el que ofrece razones más sólidas para pensar en el carácter adaptativo del mismo.

Sin embargo, a la hora de establecer el carácter adaptativo de la consciencia fenoménica, el método comparativo parece no ser de utilidad. Desde la perspectiva de Nichols y Grantham, las dificultades son básicamente dos. En primer lugar, no existen métodos fiables para establecer fuera de toda duda razonable si un determinado organismo posee o no consciencia fenoménica. Pero aun asumiendo -como suele asumirse- que por lo menos todos los vertebrados la poseen, esto dificulta todavía más el uso de argumentos comparativos.

> Los argumentos comparativos son efectivos únicamente cuando podemos comparar especies emparentadas que difieren respecto del rasgo de interés. Si todos los vertebrados tienen consciencia fenoménica, es posible que ésta haya evolucionado una única vez y haya sido conservada filogenéticamente. En suma, no podemos distinguir entre especies que tienen consciencia [fenoménica] y no la tienen de una manera precisa; pero es precisamente esta clase de distinción (entre especies emparentadas que tienen o no tienen consciencia [fenoménica]) la que sería necesaria para poder aplicar el método comparativo. Por tanto, el método comparativo es actualmente inútil a la hora de determinar si la consciencia fenoménica es o no una adaptación. (Nichols & Grantham, 2000, p. 653)

## 5.2.- Argumentos por la complejidad

### 5.2.1.- Complejidad funcional

Un segundo tipo de argumentos que suelen utilizarse en biología para establecer el carácter adaptativo de un rasgo son los argumentos que apelan a la complejidad. Desde una perspectiva biológica ampliamente aceptada, si un rasgo exhibe una gran complejidad (esto es: se compone de elementos diferentes que se articulan entre sí para desarrollar una función que ninguna de las partes podría cumplir aisladamente) esto constituye en algún sentido una evidencia de su carácter adaptativo. El ejemplo que protagoniza la discusión al respecto (y que se remonta a los argumentos que se ofrecían con anterioridad a la teoría de la evolución de Darwin para sostener la existencia de un dios creador) es el del ojo de los vertebrados. La omnipresencia de este ejemplo no es casual en la medida en que el ojo se compone de una serie de elementos que operan articuladamente para lograr una determinada función de una manera extremadamente sofisticada: una córnea que lo protege de elementos externos, un iris que se dilata o se contrae para regular la cantidad de luz que ingresa, un cristalino que enfoca la luz a la manera de un lente y que puede modificar el punto de foco, una retina que se compone de diferentes tipos celulares, cada uno de los cuales es sensible a un determinado rango de longitud de onda de la luz, etc. Ante un órgano de tan elaborada complejidad, es muy inverosímil sugerir que su existencia se debe a mecanismos aleatorios de deriva genética o que se trata de un subproducto o accidente evolutivo; antes bien, la hipótesis del carácter adaptativo de este órgano se torna la más plausible de todas, y puede reconstruirse a lo largo de la filogenia de los vertebrados una historia evolutiva en la que se reconocen varios estadios intermedios entre la primitiva mancha ocular de los euglénidos y el sofisticado ojo humano.

> (…) la selección natural es la única explicación científica para la complejidad adaptativa. La "complejidad adaptativa" describe cualquier sistema compuesto por muchas partes que interactúan entre sí donde los detalles de la estructura y el arreglo de estas partes sugiere un diseño para cumplir alguna función. (…) ningún proceso físico más que la selección natural puede explicar la evolución de un órgano como el ojo. La razón para ello es que las estructuras que pueden hacer lo que el ojo puede hacer son organizaciones de la materia cuya probabilidad es extremadamente baja. Por un margen inimaginablemente amplio, la mayoría de los objetos definidos por el espacio de la organización biológicamente posible de la materia no pueden poner en foco una imagen, modular la cantidad de luz que ingresa, responder a la presencia de límites de bordes y de profundidad, etc. Las chances de que la deriva genética, por ejemplo, pueda tener como resultado la fijación en una población de exactamente aquellos genes que podrían dar origen a un objeto así son infinitesimalmente pequeñas, y un evento así sería virtualmente un milagro. (Pinker & Bloom, 1990, pp. 709-710)

Esta clase de argumentos por la complejidad son interesantes a la hora de evaluar el carácter adaptativo de un rasgo en la medida en que pueden funcionar aun si no podemos reconstruir la historia evolutiva del mismo. En el caso de la consciencia fenoménica -en el que el registro fósil o el análisis genético no parecen muy promisorios para tal reconstrucción- podría argüirse que la complejidad que este rasgo exhibe en los términos en los que la misma ha sido definida, es un fundamento sólido para establecer su carácter adaptativo. Sin embargo, se presenta un problema: la complejidad que ha sido definida en este argumento es una complejidad que presupone una determinada función bien establecida. Se afirma que el ojo de los vertebrados es complejo en la medida en que, tal como se indicaba en la cita precedente, constituye un "sistema compuesto por muchas partes que interactúan entre sí donde los detalles de la estructura y el arreglo de estas partes sugiere un diseño para cumplir *alguna función*". No obstante, la función de la consciencia fenoménica es precisamente lo que aún no ha podido ser establecido con claridad. No podría, por tanto, apelarse a la complejidad adaptativa de la consciencia fenoménica sin establecerse previamente una función determinada en el marco de la cual tal complejidad pueda afirmarse.

> En la medida en que los argumentos por la complejidad descansan sobre el tipo de complejidad funcional ejemplificada por el ojo, estos argumentos se topan con el problema del epifenomenalismo. Al caracterizar la complejidad del ojo, asumimos que el ojo funciona para proveer información visual acerca del entorno. Es esta comprensión de fondo de la función del ojo la que sostiene la afirmación de que los varios componentes del ojo están "bien organizados" para facilitar la visión. Desplegar un argumento por la complejidad en el caso de la consciencia fenoménica requeriría un conocimiento previo sobre la función de la consciencia fenoménica, y por tanto constituiría una petición de principio contra el epifenomenalismo. (Nichols & Grantham, 2000, pp. 655-656)

## 5.2.2.- Complejidad anatómica

Existe, no obstante, otra clase de argumentos por la complejidad que parecen eludir la especificación de una función y, por tanto, eluden también el problema que plantea el epifenomenalismo. Nichols y Grantham analizan el caso de la línea lateral en muchas especies de peces, así como el de las ampollas de Lorenzini en los tiburones. En ambos casos, se trata de estructuras sumamente complejas desde el punto vista anatómico: implican la disposición regular en el espacio de determinados órganos, la exhibición de una cierta simetría bilateral, la conexión de estas estructuras con el sistema nervioso del organismo de formas muy específicas, entre otras. Sin embargo, durante muchas décadas posteriores a la descripción anatómica exhaustiva de estas estructuras, su función permaneció absolutamente desconocida para la biología. A pesar de ello, esta ausencia de una función bien establecida no parece ser suficiente para dudar del carácter adaptativo de estos rasgos.

> Una razón para tomar la complejidad anatómica como evidencia de adaptación es que la construcción de un rasgo complejo requiere energía. Como resultado, los rasgos complejos que no mejoran la aptitud o bien se deteriorarán pasivamente o bien sufrirán una presión selectiva negativa. (…) La precisa distribución espacial de sus elementos es una de las

> características llamativas del sistema de la línea lateral. Presumiblemente, implica una cantidad significativa de energía y requiere mecanismos genéticos y embriológicos especiales la creación de estas estructuras para asegurar que estén conectadas de la manera correcta. Debido a que estas estructuras (y sus relaciones espaciales) se forman de manera consistente aun siendo energéticamente costosas, uno sospecha que desempeñan una función significativa en términos de la mejora de la aptitud. Si no desempeñaran ninguna función, uno esperaría que se deterioraran. (Nichols & Grantham, 2000, p. 657)

De este modo, la complejidad anatómica de un rasgo puede constituir evidencia de su carácter adaptativo aun sin conocer cuál es la función que podría desempeñar. Sin embargo, surge un primer inconveniente al intentar utilizar esta clase de argumentos a la hora de considerar el carácter adaptativo de la consciencia fenoménica: se trata de un caso en el que no podemos hablar estrictamente de una complejidad de carácter *anatómico*. Adquiere entonces relevancia la analogía que podamos trazar entre el modo en el que entendemos funciona la consciencia fenoménica y estructuras anatómicas complejas como las que por vía de ejemplo se han mencionado. Tal analogía puede fundarse en dos aspectos: en primer lugar, la unidad que parece exhibir la consciencia fenoménica; en segundo lugar, la presunta existencia de diferentes canales de entrada a lo que podría llamarse el “sistema de la consciencia fenoménica”.

#### 5.2.2.1.- Carácter unificado

Cuando hablamos del carácter unificado de la experiencia consciente, nos referimos a los dos sentidos que distingue Searle (1992, pp. 135-136) : la unidad “vertical” y la unidad “horizontal” de la experiencia consciente. En el primer caso, la unidad se verifica en la yuxtaposición de los diferentes aspectos que conforman tal experiencia: en un momento dado un sujeto puede ser consciente del sonido ambiente que ingresa por la ventana, del sonido de las palabras de su interlocutor,

del aroma de la habitación en la que se encuentra, del roce de su calzado y de un dolor lumbar, todo ello de manera unificada. En el segundo caso, nos referimos a la unidad que parece existir entre fragmentos de experiencia consciente que se suceden cercanamente en el tiempo. Tomando el ejemplo que propone Searle, quien lee esta frase *tiene presente* el inicio de la misma aun cuando ya se encuentre leyendo sus últimas palabras; lo mismo ocurre cuando alguien tiene la experiencia consciente de una breve melodía, en la medida en que si bien las notas que la componen pueden percibirse de una forma discreta, también existe la experiencia *de la melodía* que da cuenta de una unificación entre notas que difieren en el tiempo de ejecución. Esta persistencia de las experiencias conscientes muy recientes que dan sentido a un determinado lapso experiencial implica una unidad en este sentido *horizontal*.

Si bien existen perspectivas que ponen en duda el carácter real de tal unidad y postulan -al menos en algún sentido- su carácter ilusorio (Dennett, 1995), así como también existen otras que dudan de la existencia de un único centro unificado que dé cuenta de la totalidad de la consciencia fenoménica (Flanagan, 1993), Nichols y Grantham afirman que incluso si existieran múltiples esferas de la consciencia fenoménica en la mente humana, el hecho de que al menos uno de los “sistemas” de la consciencia fenoménica exhiba esta capacidad de unificar en una única experiencia la información que le llega por múltiples canales, esto es suficiente para la solidez del argumento.

### 5.2.2.2.- Múltiples canales de ingreso

Otro de los aspectos en los que puede fundarse la analogía entre la consciencia fenoménica y rasgos anatómicamente complejos como la línea lateral, radica en el hecho de que ambos parecen componerse de un conjunto diverso de

canales de ingreso de la información, independientes entre sí, que conducen ciertos tipos de información desde los órganos sensoriales hacia un espacio o estructura central.

El hecho de que estos canales sean independientes entre sí puede desprenderse del análisis de múltiples condiciones neuropsicológicas disociativas en las que un determinado tipo de información perceptual -aun estando accesible en términos de consciencia de acceso- no es experimentada por los sujetos en su aspecto fenoménico, sumado al hecho de que la afectación de un tipo de información a la hora de alcanzar la consciencia fenoménica no parece necesariamente afectar la ruta de otros tipos de información que sí lo logran. Fenómenos que han sido mencionados como el de la visión ciega, la alexia o la prosopagnosia, son ejemplos de condiciones clínicas en los que un determinado tipo de estímulo no está presente en la consciencia fenoménica (aunque en algunos casos sí en la consciencia de acceso). La lista, no obstante, puede expandirse. Sin adentrarnos en los detalles (Nichols & Grantham, 2000, pp. 661-663), a estos ejemplos pueden sumarse otros: ciertos casos análogos al blindsight como el *blindtouch* o "tacto ciego" (en el que los pacientes no reportan tener consciencia fenoménica de un determinado estímulo táctil pero pueden "adivinar" correctamente ciertas características del mismo), o incluso un análogo en el campo de la audición (el *deaf hearing* o "audición sorda"), entre otros. Incluso dentro del campo de la percepción visual, parecen existir canales de ingreso de la información independientes entre sí, como lo sugiere el caso de la prosopagnosia o la *simultagnosia*, en el que los pacientes reportan no poder percibir simultáneamente más de un objeto, pero en situaciones experimentales específicas parecen poder hacer uso de información relativa a más de uno.

Todos estos ejemplos pueden contribuir a mostrar dos aspectos fundamentales de la estructura del "sistema" de la consciencia fenoménica. En primer lugar, se trata de un sistema al que llegan múltiples canales de entrada,

algunos de ellos en apariencia independientes de otros canales análogos que hacen disponible la información perceptual para su procesamiento. Esto puede fundamentarse sobre el hecho de que la ausencia de consciencia fenoménica no siempre implica una ausencia de consciencia de acceso, por lo que debería existir una vía que conduce la información de un determinado sentido hacia la consciencia fenoménica, y esta vía debe ser -en cierto modo difícil de precisar- distinguible de la que conduce esta información para su procesamiento en tanto consciencia de acceso. En segundo lugar, estos canales de comunicación entre la información perceptual y la consciencia fenoménica parecen ser independientes entre sí, tanto si nos referimos a áreas perceptivas diferentes como a la interna de un determinado campo perceptual como el visual. En lo que refiere a áreas perceptivas diferentes, nótese que un paciente con alexia tiene plena consciencia fenoménica de los colores, así como la visión ciega no impide en absoluto la consciencia fenoménica sobre los estímulos táctiles. En lo que refiere a canales independientes dentro de una misma área perceptiva como por ejemplo la visual, recordemos que un paciente con prosopagnosia tiene consciencia fenoménica de formas y colores, aun aquellas vinculadas con los rasgos faciales, a pesar de no poder *reconstruir* esa información a efectos de reconocer un rostro. Análogamente, algunos pacientes con acromatopsia, aun reportando experimentar fenoménicamente un mundo en blanco y negro, pueden "adivinar" correctamente respecto de cuestiones cromáticas en su campo visual, sin que este déficit en cuanto a su consciencia fenoménica de los colores afecte en nada su capacidad para reconocer rostros.

Cabría preguntarse si estos múltiples canales a los que nos hemos referido, en tanto están caracterizados desde una perspectiva *funcional*, pueden formar parte o no de un argumento basado en la complejidad *anatómica* de un supuesto sistema de la consciencia fenoménica. Nichols y Grantham asumen que, si bien no puede caerse en la simplificación de creer que existe un único nervio independiente para cada uno de estos canales que conduce a una estructura

anatómica intracerebral discreta que es el asiento de la consciencia fenoménica, sí se trata de canales que -en la medida en que las condiciones clínicas disociativas así lo indican- deben ser físicamente discretos, al menos en algunos de sus componentes.

> Los canales y órganos de la línea lateral son estructuras anatómicamente visibles. En contraste, la evidencia para múltiples "canales de entrada" hacia la consciencia fenoménica es menos directa. Dado que varios tipos diferentes de información pueden estar presentes en la mente y sin embargo no estar disponibles para la consciencia fenoménica (aun cuando otras formas de entrada sensorial alcanzan con éxito la consciencia fenoménica), es razonable inferir que los "canales" que llevan estas entradas sensoriales son físicamente distintos. Aunque estos canales de entrada funcionalmente caracterizados no son directamente visibles, deben ser realizados por algún tipo de (costosa) estructura neuronal". (Nichols & Grantham, 2000, pp. 663-664)

En síntesis, si bien no podemos establecer la complejidad funcional de la consciencia fenoménica sin antes establecer cuál es su función, si podríamos establecer su complejidad anatómica, en un sentido particular del término: la consciencia fenoménica exhibe un carácter complejo en tanto parece presentar múltiples canales de entrada independientes entre sí y estructurados de tal modo que implican una unificación de la experiencia consciente, complejidad que debe corresponderse con estructuras anatómicas en el cerebro que deben ser también complejas en un sentido análogo. Esta complejidad anatómica que, en cualquier otro rasgo de un organismo viviente, atribuiríamos sin dificultades a su carácter adaptativo, aun si no tuviéramos todavía evidencias firmes respecto de la función que tal rasgo desempeña, nos permitiría desde esta perspectiva establecer el carácter adaptativo de la consciencia fenoménica, aun sin estipular su función. De este modo, la sospecha epifenomenalista parecería quedar descartada indirectamente: incluso si no sabemos qué función desempeña la consciencia fenoménica, el hecho de que sea una adaptación es suficiente para establecer que debe cumplir *alguna función* que tenga un efecto en la aptitud del organismo para la supervivencia y la reproducción. Sólo los rasgos que tienen efectos causales en

este dominio pueden teóricamente ser objeto de la selección natural y, por tanto, ser considerados adaptaciones.

## 5.3.- Problemas con el argumento de la complejidad

### 5.3.1.- Respuestas epifenomenalistas al argumento de la complejidad

Sin embargo, Nichols y Grantham se proponen abrir la puerta para dar algunos pasos más en el sentido de la determinación de una función para la consciencia fenoménica, y proponen para ello la consideración de un *principio de la complejidad adaptativa*, en virtud del cual las características que hacen a un rasgo complejo deben estar de alguna forma vinculadas con la función que este rasgo cumple. Sin intenciones de adentrarnos en tal hipótesis, es interesante notar que, para poder presentarla, resulta importante analizar previamente si efectivamente estos argumentos por la complejidad son inmunes a otra clase de cuestionamientos epifenomenalistas. Considerar este aspecto es relevante porque conecta la distinción antes presentada entre consciencia-A y consciencia-P, con la valoración sobre si estos argumentos por la complejidad son suficientes para descartar o no la hipótesis BAV.

En primer lugar, desde una perspectiva epifenomenalista podría sostenerse que esta complejidad *podría* ser aparente mas no real. Esta sospecha, no obstante, no parece poder ir más allá de la formulación de una mera posibilidad. El argumento de Nichols y Grantham parte de la afirmación de que la consciencia fenoménica es compleja en el sentido requerido, y sobre esta complejidad funda la hipótesis de que se trata de una adaptación. Es responsabilidad del

epifenomenalista, así las cosas, presentar evidencia que eche por tierra la afirmación de que la consciencia fenoménica es compleja en la dimensión antes planteada, en la medida en que existe evidencia a favor de esta complejidad.

Un segundo argumento crítico que desde el epifenomenalismo podría plantearse es que esta complejidad que exhibe la consciencia fenoménica es, en última instancia, *heredada* del "sistema" de la consciencia de acceso. Si así fuera, la complejidad de la consciencia fenoménica no sería razón suficiente para defender su carácter adaptativo.

> Para defender la afirmación de que el sistema de la consciencia fenoménica hereda su complejidad de algún otro sistema, se debería mostrar que a) un segundo sistema (presumiblemente complejo) existe, y b) la complejidad de la consciencia fenoménica es una polizona. Por tanto, el partidario de tal afirmación debe mostrar que el sistema de la consciencia de acceso es estructuralmente distinto al sistema de la consciencia fenoménica que hemos estado discutiendo. (…) afirmar que la consciencia fenoménica no es realmente compleja sino que hereda su complejidad del órgano que tiene la función de la consciencia de acceso, no provee una hipótesis alternativa (competidora) salvo que se tenga evidencia independiente de que la función de la consciencia de acceso es desempeñada por otro órgano diferente a la consciencia fenoménica. Si, por otra parte, la función de la consciencia de acceso es desempeñada parcialmente por el sistema de la consciencia fenoménica, entonces la consciencia fenoménica continúa siendo una adaptación: es una de las estructuras que ha sido moldeada por la selección natural para cumplir la función de la consciencia de acceso. (Nichols & Grantham, 2000, pp. 665-666)

A partir de esta cita puede intuirse que expresiones tales como "complejidad heredada" y "polizón" pueden relacionarse con las nociones de enjuta o de accidente evolutivo. Robinson y sus colegas apuntan directamente a este vínculo para afirmar que los argumentos por la complejidad de Nichols y Grantham, si bien impresionantes -en la medida en que afirman el carácter adaptativo de la consciencia fenoménica sin tener que pagar el precio de comprometerse con una función determinada de la consciencia y, por tanto, defenderla de sus críticos- no logran echar por tierra la hipótesis BAV. En otras

palabras, si bien la complejidad de la consciencia fenoménica parece ser un indicador de su carácter adaptativo, no se trata de un argumento suficiente para descartar la hipótesis BAV, precisamente porque esta complejidad podría perfectamente ser un subproducto del carácter complejo de otro rasgo como, por ejemplo, el "sistema" para la consciencia de acceso.

## 5.3.2.- Otros cuestionamientos a los argumentos por la complejidad

### 5.3.2.1.- La distinción entre dos "sistemas"

No es imprescindible comprometerse con una posición epifenomenalista para poner en duda los argumentos por la complejidad. Recordemos que si el epifenomenalismo es cierto, entonces la consciencia fenoménica no puede ser una adaptación. Sin embargo, si concluyéramos que el epifenomenalismo es falso -y, por tanto, la consciencia fenoménica tuviera efectos causales- esto no implicaría necesariamente que su origen evolutivo se deba a la presión de la selección natural. Esto es importante porque los argumentos por la complejidad que se han presentado están enfocados en cuestionar la posición epifenomenalista, cuando de hecho podría aceptarse el carácter complejo del sistema que unifica la información proveniente de diferentes canales perceptuales, e incluso aceptar a partir de allí que la unificación de esta información provee una ventaja adaptativa a los organismos (lo cual parece sumamente plausible) sin que ello implique negar la posibilidad de que esta *experiencia unificada en su dimensión fenoménica* sea un subproducto o un correlato accidental de tal proceso de unificación. Un primer cuestionamiento a estos argumentos por la complejidad puede articularse en la siguiente dirección:

> Un sistema no consciente podría perfectamente unificar diferentes corrientes sensoriales, y aun así la consciencia fenoménica podría ser un subproducto o un accidente evolutivo que ocurre después de la que unidad es inconscientemente producida. Por tanto, incluso concediendo a Nichols y Grantham que nuestra experiencia está unificada en la forma en la que ellos lo afirman y que el sistema que logra tal unidad es complejo en la forma en la que Nichols y Grantham lo afirman, podría resultar que el sistema que logra la unidad es inconsciente mientras que la experiencia unificada es o bien su subproducto o bien un correlato accidental. Nichols y Grantham no dicen nada que descarte esta posibilidad. (Robinson et al., 2015, p. 370)

Nichols y Grantham parecen partir de la base de que existe algo así como un "sistema" u "órgano" para la consciencia fenoménica. Dada la estructura de la experiencia consciente, este sistema debe ser complejo en el sentido antes descripto y, por tanto, debe ser una adaptación. Un primer problema con este argumento es que acepta sin más la idea de un sistema para la consciencia fenoménica que hemos de distinguir de un sistema para la consciencia de acceso. La pregunta que debemos formularnos es si es evidente que existe tal sistema para la consciencia fenoménica.

Si la consciencia fenoménica es efectivamente una enjuta o un accidente evolutivo, un subproducto contingente o necesario que no constituye en sí misma una adaptación sino el correlato de otro rasgo que sí lo es, entonces perfectamente puede no existir un "sistema para la consciencia fenoménica". El cerebro podría estar configurado de tal modo que lograra unificar la información proveniente de diferentes canales, unificar también la información proveniente de diferentes momentos en un lapso temporal, y lograr mediante esta unificación respuestas apropiadas de parte de los organismos en virtud de los estímulos multiformes y cambiantes que se presentan constantemente en el entorno. Tal configuración cerebral sería notoriamente ventajosa para un organismo, en la medida en que la unificación de esta información permite respuestas que no podrían lograrse si esta información estuviera compartimentada.

Si tomamos, a modo de ejemplo, la situación en la que un gato persigue frenéticamente a un ratón que se escabulle entre la maleza, podemos visualizar claramente la ventaja que provee la unificación de la información. La respuesta motora del cerebro del gato, para poder ser efectiva, debe responder simultáneamente a un conjunto enorme de datos que se presentan por diversos canales y en diferentes instantes: la figura, el color, el tamaño y el aroma de esta entidad que confirman que se trata de un ratón, el movimiento del ratón en tal dirección, la velocidad a la que se mueve, el hecho de que hace unos segundos se encontraba detrás de tal trozo de corteza, el obstáculo que supone este orificio demasiado pequeño, la presencia en la cercanías de otro gato que parece haber sido alertado de la existencia del mismo ratón, etc. Sin embargo, todo lo que ocurre en este ejemplo puede reducirse a la entrada de datos a través de múltiples canales, la unificación de esos datos a través de un sistema complejo que presumiblemente contiene criterios de ponderación y ciertos algoritmos en cuanto a su procesamiento, y una respuesta motora (dinámica y cambiante según se modifica la información de entrada) que se ajusta en mayor o menor medida a esta unificación de la información. Nada de lo que se ha descripto hasta ahora en el ejemplo implica necesariamente distinguir entre dos “sistemas”, en la medida en que podría existir un único sistema que unifique la información en la construcción de un único modelo virtual de la realidad, que la respuesta del cerebro se ajuste a este modelo, y que el carácter fenoménicamente consciente de tal modelo sea un subproducto -accidental o necesario- de este proceso inconsciente de unificación.

Aceptar el carácter complejo del sistema que unifica la experiencia no nos conduce necesariamente a aceptar la existencia de otro sistema que explique el carácter fenoménicamente consciente de dicha experiencia. Análogamente, aceptar que el sistema circulatorio es complejo y que esta complejidad es señal de adaptación, no necesariamente nos conduce a pensar que existe también un sistema para el sonido del latido cardíaco, y que la complejidad de tal sonido es también indicadora de adaptación.

### 5.3.2.2.- El problema de la energía

Existe un problema con los argumentos por la complejidad que, a nuestro juicio, resulta aun más importante que esta distinción entre dos sistemas y que a su vez se relaciona con ella: se trata de la forma en la que se establece el carácter energéticamente costoso de la puesta en funcionamiento y mantenimiento de rasgos biológicamente complejos. Al respecto, si bien la idea ha sido referida tangencialmente al considerar los argumentos por la complejidad, nótese la forma explícita en la que se plantea en dos pasajes del trabajo de Nichols y Grantham, el primero referido a los rasgos complejos en general, el segundo a la complejidad de la consciencia fenoménica en particular:

> Una razón para tomar la complejidad anatómica como evidencia de adaptación es que la construcción de un rasgo complejo requiere energía. Como resultado, los rasgos complejos que no proveen una ventaja en términos de aptitud o bien se deteriorarán pasivamente o bien serán seleccionados negativamente. (Nichols & Grantham, 2000, p. 657)

> Tal como requiere energía desarrollar y mantener una red de canales y órganos de la línea lateral en los peces, debe requerir energía crear y mantener un sistema de múltiples canales que provea entradas hacia la consciencia fenoménica. Si esta clase de estructura no cumpliera ninguna función, se deterioraría pasivamente o sería seleccionada negativamente. (Nichols & Grantham, 2000, p. 663)

Tal como se ha dicho, un problema con esta clase de argumentos es que presupone la existencia de un "sistema" para la consciencia fenoménica. La presunción de que existe tal sistema es lo que permite hablar de la ventaja o desventaja del mismo en tanto rasgo fenotípico diferenciado. Sin embargo, una vez que se pone en duda la existencia de un sistema para la consciencia fenoménica y, por tanto, se toma en consideración la posibilidad de que la consciencia fenoménica sea un subproducto de otro rasgo fenotípico, se vuelve imprescindible retomar la distinción entre subproductos accidentales y subproductos necesarios.

Recordemos que un subproducto accidental es aquel que se presenta como consecuencia de un rasgo fenotípico ventajoso, pero que no se trata de una consecuencia ineludible: el rasgo ventajoso podría estar presente sin que el subproducto también lo esté. En el caso de los subproductos necesarios, se trata de aquellos que necesariamente acompañan al rasgo ventajoso por razones físicas o por constricciones de desarrollo, y por tanto no puede presentarse el rasgo ventajoso sin su subproducto.

Un ejemplo de un subproducto de un rasgo ventajoso que se encuentra unido a él por constricciones vinculadas con el desarrollo y la filogenia, es el increíblemente extraño recorrido que realiza el nervio laríngeo en las jirafas. Los ancestros lejanos de los vertebrados que en primer lugar desarrollaron un nervio que inervaba las branquias fueron creaturas similares a los peces actuales. En esas morfologías, el recorrido más directo y simple entre el cerebro y las branquias implicaba que el nervio pasara por detrás de las principales arterias del corazón. A lo largo de la evolución, conforme estos animales fueron dando origen a vertebrados con cuello, el progresivo alejamiento entre la cabeza y el corazón ocasionó una elongación del nervio laríngeo. En las jirafas actuales, este nervio -cuya salida del tronco encefálico y su entrada en la laringe guardan menos de 10 cm. de distancia- realiza un recorrido de casi 5 m. descendiendo a lo largo de todo el cuello, rodeando la aorta, y volviendo a subir hasta la cabeza del animal.

Se trata de un elegante argumento para refutar las posiciones creacionistas que se presentan bajo el argumento del diseño, en tanto el recorrido de este nervio implica un “diseño” absolutamente ineficiente que no se condice con la solución al problema que podría dar el “diseñador” con la imaginación más pobre (Wedel, 2012). Se trata, también, de un ejemplo que revela un problema en los argumentos vinculados con la energía que presentaban Nichols y Grantham en las citas precedentes. La energía que implica la creación y el mantenimiento orgánico de un nervio tan extraordinariamente largo resulta a primera vista ineficiente

considerando la estrecha distancia que separa el origen del nervio del órgano al que se dirige. Sin embargo, se trata de un subproducto desventajoso pero necesario de un rasgo fenotípico cuya ventaja es incuestionable: poder inervar la laringe a efectos de controlar centralmente la fonación y la articulación entre la deglución y la respiración. Esta es la razón por la que el complejo recorrido del nervio, aun siendo energéticamente ineficiente, se conserva en las especies contemporáneas: si bien puede existir una presión selectiva negativa respecto del recorrido, ésta se ve superada ampliamente por la presión selectiva positiva de la que es objeto la existencia misma del nervio.

La disanalogía de este ejemplo podría buscarse en el hecho de que la complejidad de este recorrido no es la complejidad en el sentido en el que la planteaban Nichols y Grantham. No obstante, esta diferencia no invalida el argumento: una característica compleja (en cualquier sentido en el que tomemos este término) puede perfectamente ser energéticamente ineficiente y, sin embargo, surgir y mantenerse en la filogenia de una especie, siempre y cuando se trate de un suproducto necesario de otro rasgo fenotípico cuya ventaja supere a la desventaja que implica esta ineficiencia. Incluso podría tratarse de un subproducto accidental (no necesario), pero que por razones de la historia evolutiva de una especie y de las constricciones estructurales que esta historia determina, se halle invariablemente presente junto con el rasgo del cual es un subproducto.

# 6.- Sobre el enfoque de la pregunta por la adaptación y un problema en el argumento del funcionalismo a la hipótesis BAV

Considerar la hipótesis BAV implica tener en cuenta dos preguntas: la pregunta por el origen evolutivo de la consciencia fenoménica (cómo y por qué evolucionó este rasgo) y la pregunta por la función de la misma (para qué sirve o qué ventaja adaptativa confiere). Ambas preguntas guardan entre sí una estrecha relación, y hay quienes creen que respondiendo a la primera lograremos responder simultáneamente a la segunda, en la medida en que no se podría explicar cómo surgió un determinado rasgo en la filogenia de una especie sin determinar cuál es su función. Sin embargo, tal como hemos visto, la pregunta por el carácter adaptativo de la consciencia no es exactamente lo mismo que preguntarse por su función, y la respuesta a la primera pregunta no necesariamente conduce a una respuesta a la segunda.

Esto es así por al menos tres razones que Kari Theurer y Thomas Polger (2018) exponen como preámbulo a un interesante cuestionamiento que realizan a uno de los aspectos del trabajo de Robinson y sus colegas. En primer lugar, podría establecerse el carácter adaptativo de un rasgo simplemente porque el aumento de su frecuencia en la población no puede explicarse solamente por la deriva genética, por lo que debe existir una presión selectiva positiva respecto del mismo. De este modo, podría confirmarse el carácter adaptativo de un rasgo sin decir nada sobre su función. En segundo lugar, la función que un rasgo pudo tener en el pasado y en virtud de la cual se fijó en la población, no necesariamente es la función que cumple en el presente; es decir, el rasgo pudo haber sido “cooptado”

para la realización de una función diferente[12] o, incluso, podría tratarse de un rasgo vestigial: funcional en el pasado cuando surgió como adaptación, pero carente de función en el presente. En tercer lugar, podría ocurrir que un rasgo efectivamente tuviera efectos causales, pero los efectos causales que tiene no son aquello por lo que fue seleccionado positivamente, e incluso podría no haber sido seleccionado en absoluto, como supone la hipótesis BAV respecto de la consciencia fenoménica. Estas tres posibilidades reafirman la idea de que responder a la pregunta por el carácter adaptativo de la consciencia fenoménica y responder a la pregunta por su función no solamente no son equivalentes, sino que tampoco puede trazarse en todos los casos una línea lógica entre las respuestas que podamos hallar a la primera pregunta, y las respuestas que podemos hipotetizar respecto de la segunda.

Sin embargo, desde la perspectiva de Theurer y Polger, si bien responder a la pregunta por el carácter adaptativo no responde automáticamente a la pregunta por la función, es posible establecer una línea de razonamiento en este sentido. Para ello, si logramos establecer evidencia de que la consciencia fenoménica fue objeto de presiones selectivas positivas, y si podemos comprender el modo en el que la presencia de este tipo de consciencia aumentaba la aptitud de sus portadores, esto quizás pueda echar luz sobre la función que la consciencia fenoménica tenía en aquel entonces, lo cual a su turno podría aclarar la función que la consciencia fenoménica cumple en los organismos actuales, incluido el ser humano. El modo en el que los autores enfocan este desafío es interesante, en la medida en que reconocen las dificultades de evaluar el carácter adaptativo de una característica tan particular como la consciencia fenoménica, dado que no existe

12 Llamamos a este fenómeno “exaptación”, término acuñado por Stephen Jay Gould y Elisabeth Vrba (1982) para referirse a estructuras que surgieron en una filogenia por su valor adaptativo en virtud de la realización de una determinada función, pero que -una vez establecidas- fueron alterándose y perfeccionándose en atención a una nueva función diferente a la original. Un ejemplo clásico de exaptación es el de las plumas de las aves, surgidas como modificaciones de las protoplumas de los celurosaurios. Estas protoplumas tenían una función de regulación de la temperatura, pero paulatinamente fueron cambiando su forma y dimensión como consecuencia de su uso como estructura voladora en el planeo de algunos terápodos arborícolas.

aún una teoría general de consenso sobre la consciencia fenoménica o sobre sus correlatos neuronales. Reconocida esta dificultad, el foco vira desde la consciencia fenoménica tomada como un todo y se dirige hacia ciertos *componentes* o *marcas* de la consciencia fenoménica que podrían entenderse como *precursores* de la misma, cuya historia evolutiva podría rastrearse mucho más allá del surgimiento de los primeros mamíferos.

De esta manera, se presume que una verdadera comprensión de la consciencia fenoménica no puede surgir más que como el resultado de ensamblar pequeños trozos de comprensión parcial sobre cada uno de los aspectos (o de los componentes de estos aspectos) que en su conjunto configuran este tipo de consciencia. Como en muchos otros órdenes del conocimiento, la comprensión del conjunto se presenta como un horizonte en el que se articula la comprensión de las partes. Esta concepción es ampliamente aceptada cuando tomamos en consideración cualquier rasgo biológico, o cuando descartamos la posibilidad de considerar como rasgos lo que deberían entenderse antes bien como constelaciones articuladas de rasgos.

> Un enfoque alternativo consiste en postergar la gran teorización sobre la consciencia, o incluso la formulación de un modelo evidencial total, y en su lugar concentrarse individualmente solo en los marcadores de la consciencia tal como los hallamos en el presente. Compare la cuestión de la evolución de, digamos, las aves. ¿Cuándo y cómo comenzó la evolución del rasgo *ser un ave* (i.e., la *avitud*)? Desde luego, tal como con el complejo rasgo de la consciencia, no esperamos descubrir que la *avitud* es el tipo de rasgo que se prende o se apaga debido a un gen particular o a un cromosoma particular (si es que hay algo así en absoluto). Por el contrario, la *avitud* implica la presencia coordinada de un cierto arreglo de rasgos. Podríamos escoger algunos como marcadores de la *avitud*, e investigar cuándo estuvieron por primera vez co-presentes en alguna clase de creatura, y qué cosas podían hacer esas creaturas en los que estaban por primera vez co-presentes, que otras creaturas no podían -creaturas que no tenían todos esos rasgos juntos, o quizás no tenían ninguno, dado el contingente ambiente evolutivo en el que pudo ocurrir-. Si existe un amplio acuerdo respecto del arreglo de características que son constitutivas o indicativas de la *avitud*, entonces la investigación será directa -lo que no

> es lo mismo que decir que es una pregunta fácil de responder, sino solamente que los problemas son los problemas usuales de hacer ciencia, especialmente ciencia histórica-. Alternativamente, podríamos diferir la cuestión de la *avitud* en sí misma, y en su lugar investigar sobre los rasgos particulares que podrían (o no) ser constitutivos o indicativos de la *avitud*. Esta sería una aproximación prometedora si existe un consenso acerca de algunas de las características asociadas a la *avitud*, pero no acerca de todas. Y ofrece una forma de avanzar, aun provisoriamente, incluso si no existe consenso sobre los rasgos asociados a la *avitud*. Después de todo, tendríamos que buscar ese tipo de evidencia incluso si tuviéramos una teoría o análisis de la *avitud*. (Theurer & Polger, 2018, p. 257)

En efecto, el enfoque parece prometedor, particularmente porque la consciencia fenoménica comparte al menos dos cirscuntancias que la asemejan a la *avitud.* En primer lugar, no parece ser un rasgo que podamos considerar como una característica fenotípica habitual y discreta, del tipo que depende de la activación o no de un proceso embriológico individual, sino más bien la composición de varias características, algunas de las cuales pueden distinguirse teórica y empíricamente entre sí, y que presumiblemente responden a procesos embriológicos relacionados pero separados. En segundo lugar, si bien no existe una teoría completa sobre la consciencia fenoménica, sí puede verificarse un cierto nivel de acuerdo (cuya magnitud, como un muchos casos, depende del grado de detalle con el que estemos observando la discusión) respecto de algunas características o aspectos que podrían ser considerados como indicativos o constitutivos de la consciencia fenoménica, al menos tal como la concebimos en el caso del ser humano. Uno de estos aspectos que los autores consideran como marcadores es el dolor, y uno de los componentes de este marcador es la nocicepción.

## 6.1.- La evolución de la nocicepción y la evolución de la consciencia fenoménica

Es difícil imaginar que podríamos finalmente elaborar una explicación científica de la consciencia fenoménica que no implique, también, una explicación científica del aspecto fenoménico del dolor. A su turno, parece insensato suponer que puede existir una teoría que dé cuenta del fenómeno del dolor en su aspecto de experiencia cualitativa, sin suponer también que esta teoría contendrá referencias a los mecanismos celulares que se presentan en la historia evolutiva como directamente asociados al dolor, los nociceptores. Este enfoque no necesariamente debe considerarse como reduccionista, en la medida que no se afirma que la comprensión de la consciencia fenoménica pueda reducirse a la comprensión de estos elementos individuales, sino que la comprensión de estos diminutos “ladrillos” de la consciencia puede contribuir a echar luz sobre una comprensión más general del asunto.

Con *nociceptores* nos referimos a las estructuras celulares presenten en los axones de ciertas neuronas sensitivas que reaccionan a determinados estímulos potencialmente nocivos (ya sean mecánicos, químicos o térmicos) enviando señales específicas hacia otras regiones del sistema nervioso y produciendo así, en muchos casos, una respuesta motora tendiente a evitar la fuente del daño y a proteger el área dañada. En los seres humanos y presumiblemente en los demás vertebrados, este proceso se experimenta bajo la forma distintiva de la sensación de dolor. La nocicepción es sólo una parte de ese complejo cúmulo de dimensiones y capacidades psicológicas y neuropsicológicas que componen el dolor, y el dolor es sólo uno de los aspectos que configuran esa miríada de procesos que forman parte de la consciencia fenoménica. Sin embargo, ensamblando las piezas pequeñas del puzzle pueden generarse los espacios para que encajen las piezas más grandes, y el ejemplo de la nocicepción permite

visualizar lo que se requiere para poder establecer, desde una perspectiva evolucionista, el carácter adaptativo o no de un determinado rasgo.

La evolución de la nocicepción es un aspecto complejo de nuestra comprensión actual sobre la evolución de los organismos en general, pero la cantidad de evidencia empírica permite sostener sin dificultades algunas afirmaciones a este respecto. En primer lugar, la nocicepción está muy ampliamente distribuida en casi todo el reino animal y algunas estructuras básicas de los nociceptores (como los canales iónicos) son comunes a todos los tipos de nociceptores conocidos. Esto parece indicar con firmeza que la nocicepción es una capacidad que surgió hace mucho tiempo en la historia evolutiva de las especies, y que se fijó en un ancestro muy antiguo, ancestro común a muchas de las especies actuales, lo cual explicaría su ubicuidad en el espectro zoológico. En segundo lugar, a pesar de estas estructuras subcelulares comunes, la variedad de nociceptores presentes en un mismo organismo, y considerándola también entre diferentes especies, es muy amplia. Existen diferentes tipos de nociceptores especializados en diferentes tipos de estímulos nocivos y que se articulan con otros aspectos fenotípicos del organismo o la especie de maneras muy diversas. Esto parece indicar también con mucha firmeza que la nocicepción ha sido objeto de presiones selectivas múltiples, lo cual contribuyó a lo largo de millones de años a su progresiva diferenciación y especialización.

A primera vista, un rasgo que intuitivamente confiere una ventaja adaptativa clara (como la de escapar de los estímulos nocivos o proteger subsecuentemente el tejido dañado), que a su vez está presente en amplísimos sectores de una línea filogenética y que presenta una diferenciación y especialización orientada a diferentes funciones, parece ser un excelente candidato a ser un rasgo indudablemente adaptativo. Sin embargo, por atractiva que resulte, no deja de ser una hipótesis que adquiere la forma de un relato plausible acerca de cómo *pudo* haber evolucionado este rasgo. Ninguna de las afirmaciones

vinculadas con la ubicuidad o con la especialización de los nociceptores permite descartar de plano la posibilidad de que este rasgo se haya fijado en una población ancestral como consecuencia de fenómenos diferentes al de la selección natural: podría perfectamente tratarse de un accidente evolutivo (producto de la deriva genética) o directamente de una enjuta. Por tanto, estas características de la nocicepción no permiten aseverar que se trate de una adaptación y tampoco, por ende, determinar cuál es su función o qué es lo que hace[13].

Tratando otro caso de un rasgo sobre cuyo carácter adaptativo existe una controversia, Elisabeth Lloyd (2005) sugiere cinco líneas de evidencia que resultaría necesario recabar para establecer que un determinado rasgo es efectivamente una adaptación.

a) El rasgo debe tener un efecto diferencial en la aptitud de su portador, esto es: debe existir una relación entre la presencia o ausencia del rasgo y las probabilidades de supervivencia o reproducción del organismo que lo porta. Adicionalmente, esta relación no solamente debe ser de carácter estadístico (pues podría tratarse de una correlación no causal), sino que debe existir alguna evidencia que muestre la conexión causal entre ambos factores. En el caso de la nocicepción, la capacidad de alertar sobre estímulos nocivos y evitarlos tiene una relación causal con la posibilidad de supervivencia y reproducción, y esta relación causal se ha establecido de manera independiente de la tesis respecto de su carácter adaptativo.

13 La utilización del término *especialización* en este párrafo debe entenderse en un sentido que podrá ser figurado o literal, dependiendo de qué punto de vista asumamos sobre el carácter adaptativo de un rasgo. Si la nocicepción se diversificó en múltiples formas y estructuras de nociceptores como consecuencia de diferentes presiones selectivas ante diferentes tipos de estímulos nocivos, el uso del término *especialización* puede darse en términos si se quiere literales: cada tipo de nociceptor se especializó en un tipo diferente de estímulo nocivo (térmico, mecánico, etc.) en tanto adaptó su forma y su estructura a una variedad de funciones. Si, por el contrario, la variabilidad entre las estructuras de los nociceptores es un producto de la deriva genética, un accidente evolutivo, un subproducto o una enjuta, entonces el término *especialización* tiene la misma intención metafórica que cuando en el contexto del paradigma evolucionista se afirma que las alas tienen un *diseño* que les permite a las aves volar.

Adicionalmente, este aumento de la aptitud que un rasgo confiere a su portador debe compararse con aquellos organismos que -en un marco de competencia ecológica- no posean tal rasgo. La comparación debe realizarse, según Theurer y Polger (2018, pp. 257-258), no con hipotéticos zombis que podrían desarrollar las mismas respuestas conductuales sin la presencia del rasgo, sino con los competidores evolutivos reales que formaron efectivamente parte del ecosistema en el que tal rasgo evolucionó. Esta distinción es de suma importancia para la consideración de la hipótesis BAV, y volveremos a ella más adelante.

Adicionalmente, la existencia de presiones selectivas a favor de la presencia y continuidad de un rasgo como la nocicepción resulta evidente, en la medida que a lo largo del tiempo evolutivo han existido siempre estímulos mecánicos, términos y químicos potencialmente dañinos para casi todas las formas de vida.

b) Para poder tratarse de una adaptación, el rasgo debe tener una base heredable, que en la enorme mayoría de los casos será una heredabilidad de índole genética. La supervivencia diferencial que un rasgo provee a su portador aumentará su frecuencia en la población siempre y cuando los genes que determinan (o condicionan) su existencia en un organismo se diseminen en el pool genético de la población en mayor medida que sus alelos alternativos. En el caso de la nocicepción esta base genética está bien establecida.

c) Además de la heredabilidad, debe existir en la población una variabilidad intraespecífica respecto del rasgo: organismos que lo presentan y otros que no, u organismos que lo presentan en un mayor grado de desarrollo, de eficacia o de eficiencia que otros. Esta variabilidad es lo que permite afirmar que un determinado rasgo fue seleccionado, en la medida en que fue seleccionado *entre otros* rasgos posibles, u otros grados de desarrollo del mismo. También la evidencia indica que la nocicepción evolucionó en un contexto donde no todos los

seres vivientes tenían esta capacidad, y hasta el presente -si bien extendida- la nocicepción no está presente en todos los organismos pluricelulares.

d) En el punto *a* se hizo referencia a la existencia de una conexión causal. Una línea de evidencia que tornaría más sólida cualquier hipótesis sobre el carácter adaptativo de un rasgo sería la que surgiría de una comprensión cabal de la naturaleza de esta mecanismo causal, esto es: del mecanismo en virtud del cual la presencia del rasgo aumenta la posibilidad de supervivencia de su portador. Parece fundarse esta distinción en la posibilidad de diferenciar entre el nivel en el que detectamos una correlación estadística entre dos variables, el nivel en el que detectamos una conexión causal entre ellas, y el momento en el que podemos describir adecuadamente el mecanismo que da lugar a esa conexión causal.

e) Por último, una quinta línea de evidencia consistiría en verificar o corroborar empíricamente el carácter adaptativo de un rasgo, observando la variación ya sea de las presiones selectivas o del rasgo en sí, tanto en el entorno natural como en el laboratorio.

Todas estas líneas de evidencia parecen sostener con firmeza la hipótesis de que la nocicepción es una adaptación. Esto resulta, en cierto sentido, una suerte de *sobrejustificación* de una hipótesis que a primera vista parecía verdadera de un modo muy directo. El objetivo de Theurer y Polger no es, sin embargo, abonar la tesis de que la nocicepción es adaptativa, sino mostrar cuál es el estándar al que debería llegar la evidencia que pemita mostrar que una pequeña parte (la nocicepción) de una pequeña parte (el dolor) de la consciencia fenoménica, es una adaptación o no. Como se ha dicho, desde la perspectiva de estos autores, comenzar a resolver los aspectos parciales del rompecabezas de la consciencia fenoménica es una aspiración sensata en un escenario en el que aún no existe un consenso sobre la naturaleza de la consciencia fenoménica en su conjunto. Retomando un ejemplo ya aludido, si no contamos con una teoría plausible sobre

la *avitud*, podemos comenzar explorando las hipótesis referidas al carácter adaptativo de las plumas, las alas, los picos, etc.

El enfoque que proponen Theurer y Polger a este respecto resulta de interés, particularmente, por el modo en el que desde esta perspectiva se responde al cuestionamiento de estar abordando los problemas fáciles de la consciencia mientras se elude el problema difícil.

> Dado que la nocicepción no es equivalente al dolor, y existe más de una forma de experimentar dolor, algunos filósofos se quejarán de que sólo dimos algunos pasos en la dirección de abordar uno de los problemas "fáciles" de la consciencia, dejando intocado el infame problema "difícil". Como evidencia, algunos filósofos tentados por esta línea de razonamiento podrían advertir que pueden imaginar organismos que podrían hacer todo lo que nosotros hacemos -incluida la evasión del daño- sin dolor. O que tales organismos podrían tener nocicepción y no dolor. O que organismos alienígenas o robóticos podrían lograr evitar estímulos dañinos sin dolor ni nocicepción. (…) De hecho no necesitamos apelar a nuestra imaginación o a filosóficos experimentos mentales para ver esto. Muchos organismos son capaces de la nocicepción pero no de experimentar dolor, porque tienen capacidades nociceptivas pero carecen de la sofisticada maquinaria cognitiva y neuronal necesaria para el dolor pura sangre de mamífero con el que estamos tan íntimamente familiarizados. Sin embargo, deberíamos recordar, en primer lugar, que si nuestro argumento es persuasivo, entonces deberemos admitir que difícilmente logremos algún avance en el problema del dolor sin abordar primero un paso clave en su evolución. Entender el dolor es, a su turno, un paso clave para entender la consciencia. No hemos explicado la consciencia y tampoco el dolor; pero no se puede arribar a ninguno de esos destinos sin hacer las paradas obligadas del camino. Entender la nocicepción es una de esas paradas. (Theurer & Polger, 2018, pp. 263-264)

## 6.2.- El problema con los zombis hipotéticos

En el transcurso de la presentación que acabamos de realizar del análisis que realizan Theurer y Polger, se afirmó que a la hora de comparar la aptitud que

confiere un determinado rasgo esta comparación debe realizarse con los competidores reales a los que -en un determinado entorno ecológico- se enfrentan los organismos que presentan tal rasgo, y no con un competidor hipotético al estilo de un zombi filosófico.

Recordemos que uno de los apoyos a la plausibilidad de la hipótesis BAV puede formularse apelando a este concepto, y así lo hacen Robinson y sus colegas cuando afirman que a partir de algunas intuiciones funcionalistas se desprende la hipótesis BAV. Tal como lo plantean los autores y lo referimos aquí, el razonamiento puede presentarse de la siguiente manera:

> 1.- Intuición funcionalista: Cualquier función biológica *f* realizada por un organismo consciente *O** pudo haber sido realizada por un organismo no consciente *O*, funcionalmente equivalente.
>
> 2.- Las propiedades que no hacen la diferencia respecto de una función biológica no pueden ser seleccionadas por la selección natural.
>
> 3.- Por tanto, la hipótesis BAV es correcta: la consciencia es o bien una enjunta o bien un accidente evolutivo. (Robinson et al., 2015, p. 380)

Como puede verse, la referencia a los zombis está implícita en la alusión a un hipotético organismo funcionalmente equivalente pero carente de consciencia fenoménica. De forma análoga y de manera más explícita razona Chalmers en un pasaje ya referido y cuya conclusión establecía:

> El proceso de la selección natural no puede distinguir entre mi persona y la de mi gemelo zombi. La evolución selecciona propiedades según su papel funcional, y mi gemelo zombi realiza todas las funciones que yo realizo tan bien como yo; en particular deja diseminadas tantas copias de sus genes como yo. Se deduce que la evolución sola no puede explicar por qué evolucionaron criaturas conscientes y no zombis. (Chalmers, 1999, pp. 163-164)

Sin embargo, y tal como lo exponen Theurer y Polger con analítica agudeza, existe en estos razonamientos una premisa inexpresada que, de hecho,

resulta falsa. Esta premisa implícita, a la que los autores llaman *"Lo reemplazable es epifenomenal"*, puede enunciarse en los siguientes términos: "Para cualquier rasgo R*, si R* pudiera ser reemplazado por un rasgo R funcionalmente equivalente, entonces R* no hace la diferencia[14]". (Theurer & Polger, 2018, p. 255) Dicho en otros términos, los únicos rasgos que pudieron haber evolucionado por selección natural son aquellos que cumplen con una función que no podría realizarse de ninguna otra manera. Como las funciones que pretendidamente podrían realizarse en virtud de la consciencia fenoménica también podrían realizarse de un modo no consciente, entonces, de acuerdo a esta premisa, la consciencia fenoménica no pudo haber sido seleccionada positivamente.

Expuesta de esta manera, la premisa no solamente impacta directamente como falsa, sino que además resulta implausible su aplicación para cualquier rasgo biológico que consideremos. Es cierto que si R y R* tienen exactamente los mismos efectos en términos adaptativos, ninguno de los dos podría ser seleccionado positivamente respecto del otro, *siempre y cuando ambos estén presentes en el mismo contexto evolutivo*. Sin embargo, esta última cláusula es de capital importancia: la afirmación sólo se sostiene si efectivamente ambos rasgos coexisten en el tiempo. Pero la comparación que debemos realizar no es entre los portadores de R* y R (salvo que sepamos que ambos rasgos estaban presentes en el mismo contexto ecológico), sino entre los portadores de R* y los competidores evolutivos reales y concretos que constituyeron el entorno dinámico de un determinado organismo a lo largo de un determinado tiempo evolutivo. En lo que respecta a la consciencia fenoménica, que podamos imaginar un contexto evolutivo en el que los organismos conscientes estuvieran en competencia por la supervivencia con sus gemelos zombis resulta irrelevante para establecer el carácter adaptativo de la consciencia fenoménica. La consciencia fenoménica pudo haber evolucionado en un contexto ecológico en el que tales zombis no

14 La expresión "hacer la diferencia" está invocada aquí en el sentido de ser "seleccionable", en cierto sentido "visible" para la selección natural.

existían, y por tanto la selección natural no operaría entre organismos conscientes y sus equivalentes funcionales no conscientes, sino entre organismos conscientes y organismos que no lo eran.

El carácter implausible de esta premisa inexpresada se torna más evidente cuando pensamos en rasgos diferentes a la consciencia fenoménica. Por ejemplo, podríamos imaginar osos polares que sobreviven al gélido clima de su entorno no a través de un gruesa capa de grasa y un tupido pelaje, sino a través de placas cartilaginosas que generen cámaras de aire con efecto termoaislante. Sin embargo, el mero hecho de poder *imaginar* estos osos polares alternativos provistos de cámaras de aire no es suficiente para poner en duda el carácter adaptativo de la piel de los osos polares actuales. Desde luego que la función de termorregulación *podría* realizarse de muchas otras maneras diferentes a cómo la realizan los osos polares, pero esto no implica que la forma en la que efectivamente realizan esta función no haya sido objeto de la selección natural. De allí la importancia de establecer las características reales del contexto evolutivo en el que un rasgo pudo haber evolucionado, en lugar de construir razonamientos a partir de como podría haber sido ese contexto si hubiera estado poblado por organismos funcionalmente equivalentes pero carentes de ese rasgo.

> Nótese que responder a la pregunta respecto de si un rasgo R* evolucionó por selección natural requiere información acerca de R* y también acerca de los otros rasgos que estaban presentes en su entorno evolutivo real. Por eso hemos hecho énfasis en que R* evoluciona en una una población real con compañeros ambientales reales e históricamente contingentes. R* no tiene por qué ser adaptativamente ventajoso en comparación con una población imaginaria de creaturas que puede hacer todo lo que hacen quienes portan R* pero sin R*. Es suficiente con que R* evolucione en una población que carece del beneficio adaptativo que de hecho es provisto por R*, cualquiera sea éste y sin perjuicio de su carácter accidental. (Theurer & Polger, 2018, p. 256)

El error en esta clase de razonamientos deriva, quizás, de la necesidad de intentar espantar al espectro del epifenomenalismo. Si la consciencia fenoménica

no tiene efectos causales, no pudo haber sido seleccionada por la selección natural, por lo que si nuestra intención es socavar las bases del epifenomenalismo podríamos apostar a buscar evidencia del carácter adaptativo de la consciencia fenoménica. Desafortunadamente, las ventajas que imaginamos puede conferir a un organismo este particular rasgo consisten en conductas que, al menos en principio, podrían ser desarrolladas en ausencia de tal consciencia. Podemos vernos empujados, pues, a la búsqueda de una capacidad o función para la cual la consciencia fenoménica resulte no solamente accesoria sino totalmente necesaria. Si encontráramos tal función, podríamos reconstruir en reversa el razonamiento, afirmar el carácter adaptativo de la consciencia fenoménica y, como consecuencia, confirmar su poder causal.

No obstante, al examinar la implausibilidad de afirmar que "lo reemplazable es epifenomenal", advertimos que no es necesario establecer ninguna función para la cual la consciencia fenoménica sea una condición necesaria si nuestra intención es mostrar su carácter adaptativo. La consciencia perfectamente podría ser una adaptación aun si pudiéramos imaginar equivalentes funcionales no conscientes de cada organismo consciente sobre el planeta. Así, la hipótesis BAV no encuentra ninguna clase de apoyo en el hecho de que las funciones que a primera vista puede realizar la consciencia fenoménica, puedan realizarse también por hipotéticos zombis idénticos en aptitud.

## 6.3.- Inesencialismo, epifenomenalismo, y una nueva especie zombi

Puede apreciarse aquí, nuevamente, la importancia de la distinción entre el *inesencialismo* de la consciencia fenoménica y el *epifenomenalismo*. En el primer caso, afirmar que cualquier actividad inteligente que puede realizarse de manera

consciente también puede realizarse en ausencia de consciencia fenoménica, es una afirmación perfectamente compatible con la hipótesis de que la consciencia fenoménica es una adaptación. No obstante, esto no es equivalente a afirmar que la consciencia fenoménica carece de poderes causales, ya que esto implicaría asumir una posición epifenomenalista desde la cual una explicación adaptacionista para la consciencia fenoménica resultaría lógicamente inconsistente. Tener presente esta distinción entre el inesencialismo y el epifenomenalismo impide que consideremos como un argumento a favor de la hipótesis BAV una circunstancia que, en virtud de su condición hipotética, de hecho no lo es, a saber: que los primeros organismos conscientes tuvieron que competir por la supervivencia en el mismo contexto evolutivo con organismos no conscientes, *pero funcionalmente equivalentes*.

Recordemos que cuando anteriormente se presentaron tres formas posibles de trazar el contorno del concepto de zombi (entre muchas otras formas posibles, dependiendo de lo que se busque en el contexto concreto de un razonamiento), mencionamos un tipo de zombi, el tipo c. Podemos ahora definir este tipo de zombi con mayor precisión de modo de observar a través de él esta distinción entre el inesencialismo y el epifenomenalismo. Llamamos zombi tipo c a "un sistema que replica toda la conducta biológicamente adaptativa pero no duplica la conducta biológicamente neutral o disfucional que se ve afectada por la consciencia, en el que la consciencia es o bien una enjuta o bien un accidente evolutivo (carente de función) que sin embargo afecta a la conducta de maneras biológicamente inútiles". (Robinson et al., 2015, p. 374) Si bien este tipo de zombis podrían hipotéticamente haber evolucionado de manera paralela a los seres conscientes sin que podamos advertirlo, el escenario hipotético de una comparación entre seres conscientes y sus competidores zombis es, por las razones que se indicaron, irrelevante a la hora de establecer el carácter adaptativo de la consciencia fenoménica.

# 7.- Una réplica a la hipótesis BAV que pretende disolver el problema difícil

La hipótesis BAV, como puede verse, está íntimamente ligada con la discusión respecto de si la consciencia fenoménica tiene o no una función. Como también hemos podido observar, ambas discusiones son -no obstante- diferentes, y la relación lógica entre las respuestas que formulemos para cada una de estas interrogantes es de carácter crítico e intrincado. Existe, a este respecto, una perspectiva que resulta de interés para nuestro análisis, en la medida en que pretende establecer una función adaptativa clara para la consciencia fenoménica que no solamente pone en duda la hipótesis BAV, sino que paralelamente contribuye a poner también en duda la distinción entre consciencia-P y consciencia-A, con las consecuencias que esta mirada implica en términos de la consideración filosófica del llamado *problema difícil*. Se trata de un trabajo de Axel Cleeremans y Catherine Tallon-Baudry (2021) en el que se intenta mostrar que la función de la consciencia fenoménica es, lisa y llanamente, la de producir experiencia subjetiva.

Gran parte del debate a este respecto, al menos en las últimas décadas, ha sido construido sobre la base de que el procesamiento de información compleja puede realizarse sin que necesariamente exista consciencia fenoménica. Numerosos avances en el campo de la inteligencia artificial han apoyado esta posición, y en muchos casos se han podido construir dispositivos con una sobrecogedora capacidad para procesar información compleja de maneras también muy complejas, sin que por ello tengamos la necesidad de atribuir a tales dispositivos nada parecido a la consciencia fenoménica. Este sustento empírico, fundamental para poder sostener posiciones inesencialistas respecto de la

consciencia, e importante también a la hora de sopesar la tesis epifenomenalista, ha contribuido a cimentar una determinada intuición: parece ser cuestión de tiempo para que dispositivos electrónicos que -al menos en apariencia- carecen de consciencia fenoménica, puedan realizar los mismos procesos de manejo de la información y elaboración de respuestas que puede realizar un ser humano. Sin embargo, en parte por el surgimiento de nuevas perspectivas teóricas sobre el vínculo entre procesamiento *consciente* y procesamiento *inconsciente*, y en parte por la dificultad para replicar resultados previos en el campo de la psicología experimental, en los últimos años ha habido una suerte de cambio de dirección en el movimiento del péndulo, y han cobrado fuerza en el debate las posiciones en virtud de las cuales existen ciertos tipos o niveles de procesamiento de información que quizás no puedan llevarse adelante en ausencia de la consciencia fenoménica (Cleeremans & Tallon-Baudry, 2021, p. 2).

En el marco de esta discusión, la distinción entre consciencia-P y consciencia-A ha sido de capital importancia, básicamente por dos razones. En primer lugar, porque la distinción tiene consecuencias directas en la indagación sobre la función de la consciencia: si bien es relativamente simple asignar una función a la consciencia de acceso, asignársela a la consciencia fenoménica ha resultado ser una tarea por demás compleja, sin que se puedan avizorar grandes consensos al respecto. En segundo lugar, y en íntima relación con lo anterior, la distinción está en la base del llamado *problema difícil* de la consciencia, en la medida en que este problema refiere precisamente al carácter subjetivo de la experiencia consciente que asociamos a la consciencia fenoménica, y quedan circunscriptos a los llamados *problemas fáciles* todos aquellos vinculados con la capacidad de que la información presente en un módulo del sistema mental pueda ser accesible para otro módulo, para reportar esta información o para guiar la conducta.

La distinción entre consciencia-A y consciencia-P, por tanto, ha moldeado gran parte de la investigación sobre la consciencia en las últimas décadas, y en este sentido es innegable que tal distinción ha sido el motor para investigaciones particularmente fructíferas tanto en el plano teórico como en lo que respecta a la indagación empírica. Sin embargo, a juicio de Cleeremans y Tallon-Baudry, esta distinción es engañosa, y el hecho de que nos habría conducido a conclusiones incorrectas radicaría, precisamente, en el modo en el que la discusión sobre la consciencia en general ha tendido a disociar su función de sus aspectos fenoménicos.

Esta disociación puede visualizarse, por ejemplo, en la enumeración de funciones propuestas para la consciencia (en general) que elabora Seth (2008) y que hemos referido en este trabajo. Cleeramans y Tallon-Baudry, en relación con esta enumeración, advierten que en aquellos casos en los que se considera la función de la consciencia en relación con la experiencia subjetiva, siempre se entiende esta última como un instrumento o un engranaje al servicio de alguna otra funcionalidad propiamente dicha. Sin embargo, “hay poca discusión sobre la experiencia subjetiva *per se* tanto en el análisis de Baars como en el de Seth: ¿qué diferencia supone el hecho de *sentir* las cosas? ¿Qué función podría tener la *experiencia subjetiva* en sí misma?” (Cleeremans & Tallon-Baudry, 2021, p. 3)

Es interesante el modo en el que estos autores consideran el impacto de la distinción entre consciencia-P y consciencia-A: el hecho de disociar las posibles funciones de la consciencia de sus aspectos fenoménicos (lo cual también se verifica en el concepto de zombi tal como lo utiliza Chalmers en su dualismo naturalista) tiene como consecuencia obturar la consideración de una cierta hipótesis: que la función de la consciencia es, precisamente, la generación de experiencia subjetiva.

A partir de este cambio de enfoque, Cleeremans y Tallon-Baudry desarrollan una hipótesis que podría sintetizarse de la siguiente manera: la

posibilidad de experimentar cosas es lo que mueve a los agentes conscientes a actuar. La experiencia subjetiva tiene, pues, un valor intrínseco. Y las cosas que tienen un valor intrínseco juegan un rol funcional relevante. La generación de experiencia subjetiva es precisamente la función de la consciencia fenoménica. Observaremos con mayor detalle esta línea de argumentación, aunque solamente desde la perspectiva de nuestro objetivo: evaluar su relevancia a la hora de considerar la solidez de la hipótesis BAV.

## 7.1.- Tres supuestos sobre el rol funcional de la consciencia fenoménica

Para comenzar, los autores dan cuenta de tres supuestos de fondo que sirven como un marco general para su análisis. Volveremos a considerar con mayor profundidad estos supuestos más adelante.

a) En primer lugar, la experiencia consciente no es equivalente a la mera sensibilidad (*sensitiviy*). "(…) una lámina de corteza cerebral, un termostato, una computadora, incluso una planta – todos ellos exhiben sensibilidad a estados de cosas particulares, pero pocas personas estarían dispuestas a adjudicarles alguna forma de consciencia". (Cleeremans & Tallon-Baudry, 2021, p. 4) Esta perspectiva, como puede verse, se corresponde con aquellos marcos teóricos que se basan en estados mentales de segundo orden (o superior), como las mencionadas teorías HOT o la teoría del GWS, referidas con anterioridad. Desde esta mirada, la consciencia implica en cierto sentido *saber que uno sabe*; debe existir un nivel de sensibilidad sobre la propia sensibilidad. Este supuesto conduce a los autores a afirmar que, si la consciencia fenoménica tiene una función, ésta debe exceder la mera sensibilidad.

b) Los agentes conscientes tienen la capacidad de aprender en qué estados conscientes *desean* encontrarse. El hecho de *desear* encontrarse en un estado mental más que en otro, desde la perspectiva de los autores, supone la idea de que los agentes fenoménicamente conscientes tienen razones propias por las que actúan, y no responden mecánicamente a razones que les son impuestas por las condiciones de su "diseño". Generalizando el ejemplo que proponen, un programa informático puede tener la habilidad de realizar tareas muy complejas, pero no lo hace *por sus propias razones*, sino como consecuencia de un algoritmo de funcionamiento que le fue asignado.

c) La noción de *razón propia* presupone a su turno la centralidad del agente. Desde la perspectiva de los autores (y se trata de una perspectiva ampliamente extendida), es incongruente hablar de experiencias conscientes sin sujetos experimentadores que las experimenten (cf. Dennett, 2019). La noción de sujeto consciente, por tanto, debería ser central a cualquier teoría sobre la consciencia fenoménica, y es precisamente la ausencia de esta dimensión de sujeto lo que en parte explica que un sistema informático no pueda tener razones propias por las cuales actuar. "A los sistemas meramente sensibles no *les importan* los estados en los que se encuentran porque no tienen un sentido de su propia existencia como agentes autónomos". (Cleeremans & Tallon-Baudry, 2021, p. 4)

## 7.2.- La experiencia subjetiva tiene un valor intrínseco

Es sobre la base de estos tres supuestos que los autores desarrollan su hipótesis sobre la función de la consciencia fenoménica en tanto generadora de experiencia subjetiva. Una breve reconstrucción del argumento que sustenta esta hipótesis comienza estableciendo la afirmación de que la experiencia subjetiva

tiene un valor intrínseco. Con el término *valor intrínseco* nos referimos a que es valiosa de un modo *no derivado*. El valor de una cosa puede ser *derivado* en el caso de que la valoremos por su carácter instrumental de cara a la obtención de otra cosa valiosa; el ejemplo paradigmático de un valor derivado es el del dinero, en la medida en que el valor del dinero no está dado por la posesión misma de billetes, monedas, o números en una cuenta bancaria, sino por su capacidad de franquear el acceso a otras cosas valiosas. Una cosa tiene valor intrínseco cuando es valiosa en sí misma, y no solamente por referencia al valor de otra cosa derivada de la primera.

De los numerosos argumentos que se ofrecen para sustentar la afirmación de que la experiencia subjetiva tiene un valor intrínseco, mencionaremos tres, escogidos particularmente porque se trata de pequeños experimentos mentales que permiten vislumbrar la profunda discusión filosófica que detrás de estas ideas se ha podido tejer, aun si no nos adentramos en esta oportunidad en ella.

En primer lugar, imaginemos que se nos ofrece una enorme suma de dinero a cambio de ser *zombificados*, es decir: de ser sometidos a un proceso de supresión de la consciencia fenoménica que nos permitirá seguir siendo sujetos igualmente funcionales y conductualmente indistinguibles de lo que somos ahora, pero carentes de experiencia subjetiva de ningún tipo. Ninguna suma de dinero parece ser suficiente para aceptar tal cosa, salvo que con ello logremos salvar la vida de otra persona, o alcanzar un cierto estado de cosas cuya realización ponderemos por encima de nuestra propia vida. La experiencia subjetiva parece tener un valor en sí misma. (Siewert, 1998)

En segundo lugar, podemos imaginar que se nos ofrecen las opciones de o bien ser zombificados o bien vivir una vida que se irá deslizando paulatina, pero constantemente, hacia lo displacentero. Al menos en un primer momento, la intuición más fuerte es que cualquier persona elegiría también evitar la zombificación, aun si esto implicara grados crecientes de experiencias subjetivas

desagradables. Sólo ante niveles insoportables de displacer la zombificación se presenta (al igual que la muerte) como una opción deseable. Esto parece indicar que resulta más deseable tener experiencias fenoménicas moderadamente negativas antes que no tener ninguna. (Kriegel, 2019)

Por último, y siguiendo el ejemplo de la "máquina de experiencia" de Robert Nozick (1974) (un dispositivo de realidad virtual que nos permite experimentar cualquier situación irreal simulada por éste) nos podríamos plantear la siguiente disyuntiva: optar entre ser zombificados o ser conectados a esta suerte de *matrix* para vivir una vida fenoménicamente indistinguible de nuestra vida ordinaria pero enteramente ficticia. Esta disyuntiva, propuesta también por Kriegel, es la que a nuestro juicio mejor ilustra este punto. La opción es entre una vida de experiencia subjetiva en un cuerpo inerte conectado a una máquina, o una vida corporal y conductualmente activa y real pero sin ningún correlato fenoménico de ninguna clase. Parece también claro que la enorme mayoría de las personas preferiría la experiencia subjetiva ficticia antes que la zombificación.

Estos tres experimentos mentales contribuyen a mostrar que la experiencia subjetiva parece tener un valor intrínseco, es valiosa en sí misma, de manera independiente del contenido concreto de esa experiencia, al menos en algún sentido[15].

> Dejando de lado circunstancias excepcionales, si nos dan la opción, la mayoría de nosotros probablemente preferiría ser un agente consciente pobre antes que un zombi rico, experimentar la vida antes que carecer de cualquier sensación. En otras palabras, estaríamos bastante dispuestos a renunciar a muchos bienes no fenoménicos sólo para conservar la experiencia fenoménica, aun en condiciones en las que esas experiencias fenoménicas podrían no siempre ser experiencias que queramos tener. (Cleeremans & Tallon-Baudry, 2021, p. 5)

15 La discusión sobre hasta qué punto o en qué medida la experiencia subjetiva debe considerarse valiosa de manera independiente de su contenido, excede ampliamente los alcances de este trabajo, pero puede encontrarse una buena cartografía de esta discusión en Lee (2019).

La afirmación de que la experiencia subjetiva tiene un valor intrínseco, cuya plausibilidad podemos advertir a partir de estos experimentos mentales, encuentra también respaldo en la forma en la que se ha procedido en investigaciones vinculadas con las ciencias experimentales, particularmente en lo que concierne a las áreas de la percepción consciente, la toma de decisiones y el estudio de las emociones. En estos tres dominios, la idea del valor subjetivo ocupa un lugar central, y su consideración permite abrir líneas de reflexión que resultan de interés para la consideración de la hipótesis BAV.

### 7.2.1.- Valor subjetivo y toma de decisiones

En lo que refiere a la toma de decisiones, mucho se ha investigado acerca de cómo los sujetos actúan de acuerdo con el valor que le asignan a determinadas cosas. El valor, a su turno, podría conceptualizarse desde diferentes perspectivas, pero existe un cierto consenso en torno a visualizarlo en términos de una magnitud que ordena diferentes bienes en una escala, con una particular métrica en cuanto a su distancia en la escala, y que permite dar cuenta de diferentes elecciones realizadas por los sujetos, así como de los juicios que elaboran expresando preferencias entre diferentes bienes: el diferente valor que asigno a una taza de café en relación con una taza de té, la forma en la que esta asignación de valor fluctúa respecto de diferentes parámetros tanto objetivos (el tipo de grano de café, la temperatura del té) como subjetivos (la sensación física del momento, el contexto de la elección), el modo en el que esta asignación se ve influida por el lapso temporal que me separa de la obtención del bien, etc., son todos parámetros que en su conjunto parecen dar cuenta de la opción tomada o la preferencia expresada.

Aquí la noción de valor se torna imprescindible para comprender la función de la recompensa, y entender la forma en la que un sistema informático o un organismo en apariencia no consciente puede guiar su conducta hacia la concreción de cierto estado o la obtención de cierto recurso. Un robot puede ser programado para calcular el valor de encontrarse en determinados estados y elaborar una pauta de conducta que maximice el valor posible de los estados en los que puede encontrarse en un determinado momento, al igual que un programa informático que "juega" al ajedrez puede programarse para que guíe su conducta valorando de manera diferencial diversos estados en los que puede encontrarse (conservar un caballo es valioso, pero más valioso es capturar una reina).

Análogamente, un organismo consciente puede ser pensado como un sistema cuya programación (cableada en su cerebro por decenas de miles de años de evolución) implica ciertas funciones de recompensa: se le asignará un determinado valor a encontrar refugio, a procurar alimento (y determinados alimentos sobre otros), a reproducirse, etc.; y estos valores asignados podrían explicar ciertas pautas de conducta. En todos estos casos, se trata de valores *extrínsecos*, fijados por la evolución de un modo análogo al modo en el que un programador fija estos valores para un sistema informático que juega al ajedrez. El problema surge al incorporar en el análisis aquellos valores que no están directamente asociados a un beneficio objetivamente computable.

> En algunos casos, un bien podría tener un valor obvio, en el sentido definido por la evolución, para satisfacer una necesidad básica (…). Pero, ¿cómo computa uno el valor subjetivo que le asigna espontáneamente al sabor de un vino, cuando el sabor, como tal, no satisface una necesidad básica, y cuando el valor asociado al sabor es diferente al valor asociado al contenido calórico del vino? Ninguna regla de la evolución especifica o siquiera constriñe este valor subjetivo del sabor. Los valores subjetivos y las preferencias resultantes son moldeados por el aprendizaje, desde luego, pero la pregunta central es: ¿por qué tenemos valores subjetivos para empezar? En el caso del vino, podemos fácilmente imaginar dos vinos con el mismo contenido calórico, uno que nos gusta y otro que no. ¿Cuál es el propósito de tener diferentes valores subjetivos para dos bienes que

> objetivamente satisfacen la misma necesidad básica? (Cleeremans & Tallon-Baudry, 2021, p. 6)

Desde la perspectiva de los autores, la existencia de una experiencia subjetiva es precisamente lo que habilita a que las cosas tengan un valor subjetivo para los organismos conscientes; lo que los sitúa en un escenario muy diferente al de un sistema informático que responde a recompensas cuyo valor está fijado siempre y únicamente de manera extrínseca. Procurar los nutrientes que uno necesita a través de una vía intravenosa satisface las mismas necesidades objetivas (y por tanto, tiene el mismo valor extrínseco) que comer una comida. Sin embargo, el acto de comer está asociado con el sabor y el aroma de la comida, experiencias que tienen -además- un valor subjetivo que solamente puede existir en la medida en que hay un sujeto para quien este sabor y este aroma son experiencias *en primera persona*.

> La experiencia subjetiva contribuiría entonces al cómputo de la función de recompensa creando un vínculo entre un bien y su valor subjetivo, intrínseco. En otras palabras, la clave para que una recompensa tenga un valor *intrínseco* es precisamente la experiencia subjetiva, dado que en su ausencia, no hay nadie para quien la recompensa signifique nada. (Cleeremans & Tallon-Baudry, 2021, p. 7)

### 7.2.2.- El valor subjetivo de las emociones

En lo que respecta a las emociones, también la noción de valor subjetivo resulta de interés. Si bien existen emociones que típicamente podríamos vincular con comportamientos evolutivamente ventajosos, como el miedo, la agresividad o el deseo sexual, también existen otras emociones cuya ventaja adaptativa resulta más difícil de visualizar, como por ejemplo el remordimiento. Si bien el modo en el que atribuimos valores a determinadas emociones es un asunto aún a dilucidar, los autores destacan el valor subjetivo que atribuimos a ciertas emociones como la

explicación de determinadas conductas evolutivamente desventajosas. El ejemplo propuesto es el del salto BASE, un deporte extremo que consiste en saltar en caída libre desde edificios, torres o acantilados utilizando un traje especial (*wingsuit*) o un paracaídas. Los circuitos neuronales que nos impulsan a salvaguardar nuestra vida sin duda nos inclinan a nunca practicar un deporte de este tipo: el valor extrínseco de la supervivencia nos lo impediría. Sin embargo, si el valor intrínseco de la emoción que esta clase de saltos implica es lo suficientemente grande, puede perfectamente explicarse esta conducta, en la medida en que este valor subjetivo supera o desborda el valor objetivo de evitar conductas innecesariamente riesgosas. De este modo, la existencia de una experiencia subjetiva permite dar cuenta de conductas que, si nos ciñéramos a un enfoque únicamente adaptativo, tendríamos serias dificultades para explicar.

### 7.2.3.- Valor subjetivo y percepción consciente

En lo que respecta a la percepción consciente también es relevante el lugar que puede ocupar la noción de valor subjetivo. Desde la perspectiva de los autores, cuando vemos, olemos o saboreamos algo, asignamos automáticamente un valor subjetivo a tal experiencia, aun cuando no siempre tenemos autoconsciencia de este valor. Existe evidencia experimental que indica que esta valor subjetivo que asignamos a algunas percepciones visuales, por ejemplo, al visualizar rostros humanos, es independiente de la tarea que estamos llevando adelante, por ejemplo, la de identificar la edad de las personas. (Lebreton et al., 2009)

> Tales hallazgos respaldan directamente nuestra hipótesis, que predice que la asignación de valor es un rasgo intrínseco de la consciencia y que ocurre automáticamente. *La asignación automática de valor es precisamente lo*

> *que le otorga a la experiencia perceptiva consciente una función*. (Cleeremans & Tallon-Baudry, 2021, p. 8, cursivas en el original)

Como se ha dicho, el valor subjetivo que automáticamente asignamos a las experiencias perceptivas conscientes tiene implicancias en la construcción de las preferencias de un organismo. Estas preferencias, a su turno, son un componente ineludible a la hora de entender cómo la conducta se ve moldeada por los procesos evolutivos. El hecho de que la experiencia perceptiva consciente esté asociada con un valor subjetivo permitiría explicar, a través de diferentes filogenias, por qué algunos organismos prefieren un determinado color, o un determinado aroma, o una determinada forma, en la medida en que estos parámetros pueden estar asociados con elementos de su entorno ecológico que son relevantes para su supervivencia y reproducción (presas, depredadores, parejas sexuales, etc.).

> El vínculo entre la experiencia sensorial y las preferencias está conectado aún más profundamente con consideraciones evolutivas. Hoffman (2019) argumenta que los sistemas perceptivos no necesitan capturar la realidad 'tal cual es' – sólo necesitan mejorar la aptitud de una creatura, esto es, detectar características del mundo que son relevantes para mejorar la aptitud. Por tanto, *la aptitud vence a la verdad*. En un nivel evolutivo, el punto es simplemente notar que los sistemas sensoriales de diferentes especies están moldeados para traducir aquellos aspectos de la realidad que optimizan la aptitud. La evolución de sistemas sensoriales naturalmente conduce al surgimiento de espacios representacionales internos que están moldeados según el nicho ecológico de cada especie. (Cleeremans & Tallon-Baudry, 2021, p. 9, cursivas en el original)

## 7.3.- Lo que esta hipótesis permite explicar

Ahora bien, habiendo establecido la relevancia y la imbricación del concepto de valor subjetivo en la literatura vinculada con la toma de decisiones, las emociones y la percepción consciente, a juicio de los autores queda suficientemente establecido que la experiencia subjetiva que identificamos con la

consciencia fenoménica tiene un valor intrínseco. Sobre esta base, se intenta responder entonces una de las preguntas centrales a la hora de considerar la hipótesis BAV: ¿cuál es la relevancia funcional de la consciencia fenoménica? Al respecto, Cleeremans y Tallon-Baudry argumentan que el hecho de que la experiencia subjetiva posea un valor intrínseco permite dar cuenta de: a) el comportamiento creativo y flexible; b) el impulso interno a la acción; c) la existencia de una "moneda común" para la mente.

### 7.3.1.- Comportamiento creativo y flexible

En lo que respecta al comportamiento creativo y flexible, ya se ha referido en este trabajo el vínculo entre esta capacidad y la consciencia fenoménica, muy especialmente al abordar casos como el *petit mal*. La consideración del valor intrínseco de la experiencia subjetiva permite visualizar este vínculo desde un nuevo ángulo. El hecho de que las experiencias subjetivas que la consciencia fenoménica habilita puedan tener un valor intrínseco hace que los organismos fenoménicamente conscientes desarrollen comportamientos y conductas que van más allá de las pautas incorporadas al cableado neuronal por procesos evolutivos, y esta capacidad de actuar de manera flexible mediante la creación de nuevas pautas de comportamiento puede redundar, finalmente, en una mayor capacidad para adaptarse a un entorno cambiante y multiforme.

> *La experiencia subjetiva le da valor a lo que no tiene un valor intrínseco tal como lo definirían las reglas de la física o la biología*. La habilidad de un agente para disfrutar -o padecer- la experiencia consciente le otorga un considerable valor adaptativo – el agente puede ahora actuar de maneras que flexiblemente esquivan las constricciones montadas por la evolución en tanto le otorga mucha mayor libertad para actuar de maneras que elige por sus propias razones. (Cleeremans & Tallon-Baudry, 2021, p. 10, cursivas en el original)

### 7.3.2.- Impulso interno para la conducta

A juicio de los autores, el valor intrínseco de la experiencia subjetiva también es relevante para dar cuenta del impulso interno que los organismos fenoménicamente conscientes tienen para actuar. Como se ha dicho, un sistema informático o un organismo no consciente puede -en virtud de la función de recompensa- guiar su conducta hacia la obtención de cierto recurso o la concreción de cierto estado. Ya sea como consecuencia de algoritmos de programación o de una disposición de las redes neuronales, puede asignarse un valor a estos recursos o estados y establecer una pauta de conducta acorde a su maximización. Sin embargo, en estos casos la recompensa no es recompensante *para ellos*. Esta distinción es relevante y se funda en un razonamiento que podría plantearse así:

*1) La experiencia subjetiva (propia de la consciencia fenoménica) tiene valor intrínseco.*

*2) Los sistemas que se basan en funciones de recompensa buscan maximizar esta función, esto es: maximizar las recompensas.*

*3) En el ejercicio de su búsqueda de recompensas, estos sistemas buscan obtener cosas que tienen un valor, valor que puede ser solamente extrínseco, solamente intrínseco, o la suma de ambos tipos de valor.*

*Por lo tanto, los sistemas fenoménicamente conscientes pueden buscar recompensas que son valiosas, parcial o totalmente, de modo intrínseco, a*

*diferencia de otros sistemas cuyas funciones de recompensa están determinadas por valores que les son impuestos extrínsecamente.*

Puede apreciarse en el siguiente pasaje la relevancia que esta conclusión reviste.

> Los algoritmos dirigidos por valores, incluidos los de refuerzo del aprendizaje, funcionan de esa manera. La única razón por la que actúan en absoluto es porque están programados para maximizar una función de recompensa, pero paradójicamente, la recompensa en sí misma no es recompensante *para ellos*. Esto también se sostiene para muchos organismos biológicos, que están programados por la evolución para sobrevivir, reproducirse y en todo caso alcanzar la homeostasis. En otras palabras, obtener recompensas es un poderoso impulso para la conducta. Pero, ¿qué define lo que es una recompensa, más allá de las necesidades biológicas básicas? Si la experiencia consciente tiene un valor, entonces generar experiencia consciente es recompensante. Dado que la función de recompensa es generada internamente, en lugar de ser impuesta por constricciones biológicas, los agentes conscientes actúan por razones propias. *Buscar nuevas experiencias conscientes (y por tanto valiosas), constituye entonces un impulso para llevar adelante conductas que no necesitan estar justificadas por razones adicionales*. (Cleeremans & Tallon-Baudry, 2021, p. 10, cursivas en el original)

### 7.3.3.- Unificación de la experiencia

Tal como se ha referido con anterioridad, la unidad (tanto en un sentido *vertical* como *horizontal*) parece ser uno de los rasgos de la consciencia fenoménica. Si bien -y tal como se mencionó- esta afirmación no está exenta de críticas, la hipótesis de que las experiencias subjetivas que habilita la consciencia fenoménica son intrínsecamente valiosas permitiría también dar cuenta del alegado carácter unificado de este tipo de consciencia.

Esto es así porque el carácter intrínsecamente valioso de las experiencias conscientes *en tanto son conscientes* (esto es, con relativa independencia del *contenido* de estas experiencias) permite considerar al valor intrínseco de las experiencias como una suerte de *moneda común* en el "mercado experiencial" de la mente. Los seres humanos podemos establecer juicios de preferencia entre cosas que tienen pocos aspectos en común, en parte precisamente porque estas experiencias (por más disímiles que puedan ser entre sí) implican un valor *para el agente*, valor que puede ser comparado en términos de escala en una suerte de espacio de coordenadas común. Consideremos por ejemplo la forma en la que elegimos entre dos experiencias muy diferentes, como escuchar un determinado álbum musical o caminar hasta la heladería y tomar un helado de un determinado sabor. Estas experiencias parecen no tener nada en común si consideramos su contenido fenoménico. Sin embargo, ambas implican un valor intrínseco *en tanto experiencias subjetivas*. Precisamente por este hecho, es que podemos comparar ambas experiencias en términos del valor que implican *para uno*. De este modo:

> (…) la codificación de valores, si tiene lugar en el espacio fenoménico (esto es, '¿qué me hace esto *a mí*?') es precisamente el tipo de mecanismo que uno esperaría que habilitara la emergencia de un espacio fenoménico unificado, fundamentalmente a través de la creación de una moneda de recompensa común en el cerebro. (Cleeremans & Tallon-Baudry, 2021, p. 11, cursivas en el original)

Como hemos visto, el hecho de que la consciencia fenoménica habilite experiencias subjetivas que son intrínsecamente valiosas, constituye -desde esta perspectiva- la clave para atribuir a la consciencia fenoménica una relevancia funcional que, entre otros argumentos, contribuye a sostener su condición de adaptación. La ventaja adaptativa que la consciencia fenoménica otorga a los organismos que presentan este rasgo, es la posibilidad de explorar nuevas pautas de comportamiento que trascienden las limitaciones impuestas por mecanismos neuronales que la propia selección natural construyó como algoritmos de recompensa, particularmente porque posibilitan nuevos tipos de recompensa que

exceden la supervivencia, la reproducción o la homeostasis. El hecho de que los organismos fenoménicamente conscientes puedan explorar estas nuevas líneas de comportamiento -si bien en ocasiones puede resultar desventajoso por las consecuencias negativas que pueda tener- en muchos casos permite la adopción de estrategias o el desarrollo de habilidades que no podrían haber surgido si los organismos biológicos basaran su conducta únicamente en función de los valores extrínsecos que estas conductas tienen para su supervivencia.

## 7.4.- Implicancias de esta hipótesis para la disolución del *problema difícil*

Como se ha mencionado, la distinción entre consciencia fenoménica y consciencia de acceso está en la base del llamado *problema difícil* de la consciencia. Cleeremans y Tallon-Baudry, en el marco de esta hipótesis que postula que la consciencia fenoménica tiene una relevancia funcional que la convierte en una adaptación en virtud de su capacidad para generar experiencias subjetivas, implica de cierta manera el abandono de la distinción entre consciencia-P y consciencia-A y, subsidiariamiente, la posible disolución del *problema difícil*.

Puede visualizarse ahora de qué manera a juicio de los autores la distinción entre consciencia-P y consciencia-A resulta engañosa. En primer lugar, es engañosa en tanto conduce a la contraintuitiva suposición de que la consciencia fenoménica es un mero epifenómeno funcionalmente irrelevante. Desde luego, el carácter contraintutivo de esta suposición no es en sí mismo un argumento en su contra; lo que se intenta señalar es que la sospecha epifenomenalista ha podido sostenerse, aun a pesar de su carácter contraintutivo, en parte porque se ha fundado en la distinción P-A, distinción cuya puesta en duda implicaría retirar uno de los principales apoyos a esta sospecha.

En segundo lugar, es engañosa en tanto ha permitido afirmar la posibilidad de experiencias fenoménicas pero sin acceso, como se desprende de algunos de los pasajes precitados de Block (1995) o Searle (1992). Postular la posibilidad de que existan estados mentales P-conscientes que sin embargo no son A-conscientes, implica de cierta forma aceptar la posibilidad de que existan experiencias sin un sujeto que las experimente[16], posibilidad incompatible con esta perspectiva en la medida en que en la hipótesis de Cleeremans y Tallon-Baudry se asume la centralidad del sujeto como núcleo subjetivo de cualquier conjunto de experiencias. Las miradas que disocian el carácter fenoménico del carácter accesible de un estado mental de alguna manera eluden el hecho de que un estado mental cualquiera (sea fenoménico o no) presupone a un sujeto para el cual ese estado mental es *su* estado mental.

Podemos ver, en un pasaje del trabajo de Cleeremans y Tallon-Baudry, de qué manera se articula el cuestionamiento a la distinción P-A con su hipótesis de que la generación de experiencias subjetivas fenoménicas es precisamente la función de la consciencia, y finalmente con la afirmación de que el llamado *problema difícil* se disuelve (podríamos decir: se revela su condición de *pseudoproblema*), una vez que se recupera la centralidad del sujeto que experimenta. En palabras de los autores:

> Nuestro segundo punto central (…) va mucho más allá de los debates sobre las relaciones entre consciencia de acceso y consciencia fenoménica: proponemos que la *función* misma de la consciencia es precisamente proveer a los organismos de *experiencias fenoménicas*. En otras palabras, lo que la consciencia *hace* es posibilitar a los organismos la *experiencia* de

16 Respecto a este punto, es interesante el experimento mental que proponen Michael Cohen y Daniel Dennett (2012) a partir de la hipótesis de un sujeto cuyas áreas cerebrales responsables de la experiencia del color (V4) son desconectadas del resto de su cerebro, pero continúan funcionando *normalmente*. Si le preguntáramos a este sujeto de qué color es la manzana que está observando, posiblemente respondería que no tiene ninguna experiencia de color. Sin embargo, en tanto el área cerebral encargada del color continúa funcionando, hay quienes afirman que los estados mentales perceptivos continúan existiendo aun sin acceso (Fahrenfort & Lamme, 2012). Afirmar que una región cortical aislada está experimentando un estado mental es, a juicio de Cleeremans y Tallon-Baudry, una buena muestra de las consecuencias inverosímiles de disociar la consciencia-P de la consciencia-A.

> sí mismos, del mundo y de otros agentes. Esta perspectiva coloca al *sujeto* experimentador en el corazón de la investigación sobre la consciencia y puede quizás resultar en una disolución del problema difícil: los estados fenoménicos, en lugar de ser carentes de función -un misterioso ingrediente añadido a una sopa cognitiva- constituyen por el contrario la base misma sobre la cual hacemos todo lo que hacemos. (Cleeremans & Tallon-Baudry, 2021, p. 13, cursivas en el original)

## 7.5.- Algunos problemas con esta hipótesis

Al considerar críticamente la perspectiva de Cleeremans y Tallon-Baudry, y tal como lo hemos anunciado, es importante revisar al menos dos de los supuestos en los que se funda esta hipótesis y que fueron referidos en la sección 7.1.

Recordemos que dicho marco general de análisis sobre la consciencia fenoménica implicaba, en primer lugar, distinguir entre lo que serían razones *propias* para actuar y razones impuestas *desde el exterior* del agente, como cuando un programa informático actúa en función del diseño impuesto por su programador. La clave de esta distinción, a juicio de los autores, es que los agentes fenoménicamente conscientes *desean* encontrarse en determinados estados mentales más que en otros. En segundo lugar, y en estrecha relación con el supuesto anterior, esto implica centrar la cuestión de la consciencia fenoménica en el sujeto que la experimenta, en la medida en que “a los sistemas meramente sensitivos no les importan los estados en los que se encuentran en la medida en que carecen de un sentido de su propia existencia en tanto agentes autónomos”. (Cleeremans & Tallon-Baudry, 2021, p. 4)

Ambos supuestos entrañan, a nuestro juicio, distinciones problemáticas. Pensemos por ejemplo en el concepto de *razón*, como cuando se distingue entre

razones impuestas desde el exterior y razones *propias*. Recordemos el ejemplo de las llaves e imaginemos que antes de cambiar las llaves de bolsillo, y a lo largo de un par de metros de mi recorrido, me entretengo arrojando brevemente las llaves hacia arriba y volviéndolas a recoger, encontrando satisfacción en el desafío de volverlas a atrapar y en el sonido que produce el tintineo de las llaves en el aire. Cabría preguntarse cuáles son las razones que me llevan a realizar tal acción.

Desde una cierta perspectiva, se trata de razones *propias*. No está inserto en el diseño "de fábrica" de mis redes neuronales -pautado por decenas de miles de años de evolución- que yo deba emprender tal conducta por la ventaja que ésta reportará para mis chances de supervivencia y de reproducción. Parecería que lo hago *por el mero placer de la experiencia*. Desde esta mirada, las razones que me mueven a ese juego, en tanto no están pautadas por constricciones evolutivas vinculadas con la adaptación son, como decíamos, razones *propias*. Sin embargo, también es posible enfocar el asunto desde otra perspectiva. De hecho, en cierto sentido, el estado neurofisiológico que dio pie a la acción motora de jugar con las llaves es, como cualquier otro estado neurofisiológico, el resultado de una miríada de otros procesos neurofisiológicos, todos ellos -a su turno- el resultado de una complejísima interacción entre un cerebro tal como fue construido con base en la información genética en el útero materno, y todos los estímulos que ha ido recibiendo (y que lo han ido moldeando) desde el inicio de su vida hasta el momento actual en el que se desencadena el juego. La razón que me mueve a jugar con las llaves (el deseo de tener determinada experiencia) podría ser entendida como un eslabón más en la cadena causal que -desde la perspectiva de una clausura causal del mundo físico- se extiende hacia atrás en el tiempo hasta el origen mismo del universo.

Aquí el concepto de *razón* se convierte en una de las formas que puede exhibir el concepto de *causa*, lo que parece disolver la distinción entre razones propias y razones impuestas desde el exterior. Si por razón *propia* entendemos las

causas que se terminan de configurar en el interior del agente, entonces también son propias las razones que mueven al programa informático a desarrollar determinadas jugadas en la partida de ajedrez. La evolución no podía prever que me encontraría en la situación exacta de verme ante la posibilidad de arrojar las llaves al aire para disfrutar de su tintineo, del mismo modo que el programador del algoritmo no podía prever que se llegaría a tal estado particular del tablero con las piezas ordenadas en esta disposición puntual. El impulso a arrojar las llaves termina de configurarse en el interior de mi cerebro, producto de la interacción de mi cerebro con un entorno dinámico, del mismo modo que la acción de mover al caballo termina de configurarse en el interior del programa informático, producto de la interacción entre las pautas de programación y el estado actual del tablero. Por lo tanto, si son propias mis razones son también propias las razones del programa informático.

Por el contrario, si quiere afirmarse que las razones del programa informático no son propias porque el programa responde a pautas generales impuestas por un programador, no queda claro por qué tal enfoque no podría aplicarse también a la acción humana de jugar con las llaves. La decisión de mover al caballo no es una decisión del programador, pero sí lo son las pautas generales que hacen que en determinado escenario el programa escoja mover al caballo. Análogamente, la decisión de jugar con las llaves no es una decisión de la evolución ni de la arquitectura neuronal que exhibía mi cerebro cuando nací, pero sí es un resultado de cómo esta estructura neuronal ha ido moldeándose a lo largo de mi vida producto de la interacción entre su diseño “de origen” y los estímulos a los que fui expuesto.

Parece revelarse entonces un problema que podemos visualizar retomando una cita ya referida del trabajo de Cleeramans y Tallon-Baudry.

> La habilidad de un agente para disfrutar -o padecer- la experiencia consciente le otorga un considerable valor adaptativo – el agente puede

> ahora actuar de maneras que flexiblemente esquivan las constricciones montadas por la evolución en tanto le otorga mucha mayor libertad para actuar de maneras que elige por sus propias razones. (Cleeremans & Tallon-Baudry, 2021, p. 10)

Si la habilidad de tener experiencias conscientes le otorga a los agentes un valor adaptativo, entonces la capacidad de actuar flexiblemente forma parte de esta ventaja adaptativa, en la medida en que les permite descubrir nuevas pautas de conducta que pueden redundar en una mejora en la chances de supervivencia y reproducción. Esta capacidad, si efectivamente implica una ventaja adaptativa, ha sido montada por la evolución en virtud de mecanismos de selección natural. Por lo tanto, parece contradictorio afirmar que un rasgo posee un valor adaptativo en la medida en que permite eludir las constricciones planteadas por la evolución. En otras palabras, si posee un valor adaptativo, fue precisamente montado por la evolución y no constituye, al menos en un sentido relevante, una elusión de sus constricciones.

Volvamos al ejemplo del salto BASE. Desde una perspectiva muy llana, nuestro cerebro debería estar “cableado” para evitar estos riesgos extremos y no deberíamos sentir el deseo de realizar esta práctica. Sin embargo, muchas personas lo hacen. La explicación de los autores es que el valor intrínseco de la experiencia consciente que esta práctica implica sobrepasa esta constricción evolutiva tendiente a la conservación de la propia vida. No obstante, en la misma línea de argumentación se expresa que la capacidad de buscar nuevas experiencias conscientes en virtud de su valor intrínseco implica una ventaja adaptativa para el agente. Por tanto, la práctica del salto BASE no es una forma en la que el agente elude las constricciones evolutivas o la programación de su cerebro. Por el contrario, es una de las conductas a las que la configuración evolutiva de su cerebro lo empuja, en la medida en que su capacidad para experimentar estados mentales fenoménicamente conscientes que son intrínsecamente valiosos lo impulsa a llevar adelante prácticas placenteras aun ante el riesgo de su propia supervivencia. Este error sería análogo al de juzgar que porque un programa de

ajedrez sacrifica a su reina, está de alguna manera eludiendo su programación de origen. Este argumento olvida que en la programación de origen puede perfectamente estar incluida la posibilidad de sacrificar una reina si esto redunda en aumentar las posibilidades de ganar la partida.

Este aspecto problemático de la distinción entre razones *propias* y razones *impuestas desde el exterior* está también vinculado con otro de los supuestos que forman parte del marco general de análisis en el que se desarrolla esta hipótesis: la idea de que a los agentes fenoménicamente conscientes, a diferencia de otra clase de agentes, *les importan* los estados en los que se encuentran. Cabe preguntarse acerca del alcance de expresiones como “les importan” o “son recompensantes *para ellos*”. Si efectivamente puede disolverse la distinción entre razones propias y razones impuestas desde el exterior, entonces -al igual que en el caso anterior- deberíamos aceptar que, o bien a muchos sistemas basados en algoritmos con funciones de recompensa “les importan” los estados en los que se encuentran (aun aquellos que carecen de consciencia fenoménica), o bien a ningún sistema “le importa” el estado en el que se encuentra, en tanto todas las funciones de recompensa responden a constricciones, ya sean de diseño o producto de la evolución.

# 8.- Conclusión

A lo largo del presente trabajo hemos tenido como objetivo principal evaluar críticamente la hipótesis BAV. Para ello, nos hemos detenido en primer lugar en la descripción de un marco general de análisis para la consideración de dicha hipótesis. Este marco nos condujo a una descripción de algunos de los mecanismos involucrados en la evolución biológica de los seres vivos, y particularmente de la relación entre las adaptaciones producto de la selección natural y las enjutas o los accidentes evolutivos producto de las constricciones de desarrollo o de la deriva genética. Asimismo, nos aproximamos a la complejidad que implica la delimitación de un concepto tan amplio como el de consciencia, haciendo luego un énfasis particular en la distinción entre consciencia fenoménica y consciencia de acceso. Para terminar de encuadrar esta discusión, visualizamos la complejidad que involucra el concepto de *función* y las distinciones analíticas que pueden realizarse sobre él.

La hipótesis BAV presupone que la consciencia fenoménica carece de función o que, en todo caso, no es su función lo que explica su surgimiento y continuidad en la filogenia de las especies. Por tanto, fue necesario explorar algunas de las funciones que se le han atribuido a la consciencia fenoménica, lo cual nos condujo a introducir en nuestro marco de análisis algunos de los problemas vinculados con el epifenomenalismo y el inesencialismo de la consciencia fenoménica, y las diversas formas en las que el concepto de zombi ha tomado un lugar central en la discusión filosófica al respecto de estos asuntos.

En este marco, pudimos notar que hay consenso en la falta de consenso. Dicho de otro modo, no existe ninguna función propuesta para la consciencia fenoménica que sea objeto de consenso en la literatura revisada. Si bien las incógnitas respecto de la consciencia en general son numerosas, y se trata sin duda de uno de los aspectos de la realidad más esquivos para la comprensión científica, creemos que en el caso particular de la pregunta por la función de la consciencia fenoménica, esta falta de consenso se apoya fuertemente en dos circunstancias: en primer lugar, las dificultades para superar la perspectiva escéptica que supone el epifenomenalismo; en segundo lugar, la forma en la que la discusión sobre la función de la consciencia fenoménica se ve limitada por la distinción entre consciencia fenoménica y consciencia de acceso, particularmente en la implicancia que esta distinción tiene en la delimitación del llamado "problema difícil" de la consciencia. Como hemos visto, ante muchas de las funciones que se le han atribuido a la consciencia fenoménica pueden establecerse contraargumentos que apunten a mostrar que: a) se trata de una función que podemos atribuir a la consciencia de acceso, antes bien que a la consciencia fenoménica, o b) se trata de una función que la consciencia fenoménica podría llevar adelante, pero que también podría realizarse en su ausencia.

Estas dificultades para establecer con claridad una función para la consciencia fenoménica parecen constituir un apoyo importante para la hipótesis BAV, en la medida en que si pudiéramos establecer una función fuera de toda duda razonable, esto sería un argumento muy potente a la hora de defender el carácter adaptativo de este tipo de consciencia. En otra palabras, la hipótesis BAV parece sostenerse, en parte, gracias a la ausencia de una respuesta clara a la pregunta por la función de la consciencia fenoménica.

Existen, paralelamente, otros argumentos a favor de la hipótesis BAV. Como hemos visto, existen criterios no arbitrarios a los que suele recurrirse para afirmar el carácter de enjuta o de accidente evolutivo de un determinado rasgo, y

la consciencia fenoménica parece cumplir a cabalidad con cada uno de los tres criterios propuestos. Asimismo, puede mostrarse el vínculo lógico entre la hipótesis BAV y muchas perspectivas metafísicas sobre lo mental. Si bien ninguna de estas dos líneas de argumentación ofrecen garantías de que la hipótesis sea correcta, parece sensato afirmar que la convierten en una hipótesis que -al margen de su carácter contraintutivo- no puede descartarse fácilmente.

Sin embargo, a la hora de analizar estos argumentos, surgieron consideraciones críticas que permiten avanzar en la profundización del debate. En primer lugar, si bien la discusión sobre la función de la consciencia fenoménica resulta -por las razones expuestas- central a la hora de considerar la hipótesis BAV, también es posible abordar el carácter adaptativo de la consciencia fenoménica sin comprometerse con ninguna función específica; en otros términos: saldar el debate sobre la función de la consciencia fenoménica no es un requisito indispensable a la hora de establecer si la consciencia fenoménica es o no una adaptación. Los argumentos por la complejidad contribuyen a mostrar que la consciencia fenoménica es compleja en un sentido análogo a la complejidad anatómica, en la medida en que la consciencia fenoménica implica una unificación de la experiencia consciente y presupone la existencia de múltiples canales de entrada de la información, complejidad que no cabría esperar de un rasgo que fuera una enjuta o un accidente evolutivo. Así, si bien de este carácter complejo de la consciencia fenoménica podrían extraerse hipótesis sobre sufunción, se busca mostrar su carácter adaptativo al margen de la discusión sobre cuál sería dicha función.

Este argumento, no obstante, se enfrenta a problemas serios, todos ellos estrechamente vinculados entre sí. En primer lugar, porque puede sostenerse desde una perspectiva epifenomenalista que esta complejidad “anatómica” que exhibe la consciencia fenoménica es compatible con su carácter causalmente inerte, y que los poderes causales deben hallarse, en todo caso, en otro rasgo (como la

consciencia de acceso) del cual la consciencia fenoménica pueda ser un subproducto. En segundo lugar, porque la existencia misma de un "sistema" para la consciencia fenoménica puede ser puesta en duda, máxime si lo que entendemos por "consciencia fenoménica" es un fenómeno sobreviniente a otro u otros rasgos del funcionamiento de la mente. En tercer lugar, porque la pretendida inversión energética que implicaría el mantenimiento de un sistema complejo como el de la consciencia fenoménica no es necesariamente una evidencia de su carácter adaptativo, en la medida en que podría ser un subproducto cuya desventaja en términos energéticos fuera superada por la ventaja que otorga el rasgo adaptativo de base.

Otra línea de argumentación importante, a la hora de poner en duda la hipótesis BAV, es la que surge de retirar uno de los apoyos de la hipótesis, a saber: el argumento que pone en duda que la hipótesis BAV se desprenda lógicamente de algunas intuiciones funcionalistas. Como hemos visto, existe un error en este razonamiento en la medida en que implica la falsa asunción de que un rasgo no pudo haber sido seleccionado por la selección natural si podemos imaginar hipotéticos competidores biológicos de los portadores de ese rasgo. La consciencia fenoménica puede haber evolucionado por selección natural aun si es posible imaginar estos hipotéticos zombis, en virtud de que los rasgos son seleccionados en el marco de un contexto ecológico real y no en el marco de un contexto ecológico hipotético.

Esta línea de argumentación es importante no solamente porque retira uno de los apoyos de la hipótesis BAV, sino porque además contribuye a trazar las líneas generales de un plan de investigación que permitiría avanzar en la dilucidación del debate sobre el carácter adaptativo de la consciencia fenoménica. Aun sin vislumbrar un consenso en el horizonte sobre la consciencia fenoménica en su conjunto, es posible esperar que la construcción de este conocimiento tenga lugar "desde abajo", estableciendo el carácter adaptativo de los elementos que

componen la experiencia fenoménica, como se ha mostrado a partir del caso de la nocicepción.

Las dos primeras líneas de argumentación abordadas eluden la determinación de una función para la consciencia fenoménica, ya sea porque pretenden establecer su carácter adaptativo al margen de esta determinación, o bien porque trazan una perspectiva que debe, en primer lugar, establecer la función y el carácter adaptativo de los componentes o marcadores de la consciencia fenoménica. Existe, adicionalmente, una tercera línea de argumentación que puede contribuir a poner en duda la hipótesis BAV, y que consiste en volver a la pregunta por la función de la consciencia fenoménica para enfocar la interrogante desde una nueva perspectiva.

La hipótesis de que la consciencia fenoménica tiene como función la generación de experiencias subjetivas resulta, a nuestro juicio, la más prometedora de las funciones que se han glosado, en la medida en que permite dar cuenta del carácter creativo y flexible de parte del comportamiento humano (lo que se halla en la base de la discusión que surge a partir del análisis de casos como el *petit mal*), así como del carácter unificado de la experiencia fenoménicamente consciente. No obstante, a la hora de intentar explicar el impulso interno para la acción, la hipótesis revela un flanco débil frente al fisicalismo y al principio de clausura causal del mundo físico: la dificultad de establecer con claridad por qué razón han de considerarse propias las causas que mueven a la acción en un agente fenoménicamente consciente, pero ajenas en el caso de un agente no consciente.

Al margen de esta dificultad, la hipótesis de que la consciencia fenoménica tiene como función la generación de experiencias subjetivas contribuye también a mostrar un vínculo que atraviesa casi todas las reflexiones que han formado parte de este recorrido. Al afirmar que ha sido un equívoco disociar los aspectos fenoménicos de la consciencia respecto de su rol funcional, se visualiza con claridad hasta qué punto la hipótesis BAV está profundamente imbricada con la

distinción entre consciencia fenoménica y consciencia de acceso y, consecuentemente, con la distinción entre los problemas fáciles y el problema difícil de la consciencia. No cabe la posibilidad de poner en duda el carácter adaptativo de la consciencia si no trazamos un límite entre la consciencia fenoménica y la consciencia de acceso, en la medida en que ésta última cumple un rol funcional claro y, por tanto, sería difícil sostener su carácter de enjuta o de accidente evolutivo. Si la distinción entre consciencia fenoménica y consciencia de acceso se demuestra incorrecta, no solamente se disuelve también la distinción entre los problemas fáciles y el problema difícil, sino que la hipótesis BAV se convertiría en una hipótesis ya no falsa, sino carente de sentido.

En síntesis, a partir de la literatura revisada podemos afirmar que la hipótesis BAV ocupa un lugar central en el debate sobre la consciencia fenoménica, a pesar de que en diversas circunstancias se presente bajo diversas denominaciones y enfoques. Se trata de una hipótesis que solo puede sostenerse a partir de aceptar distinciones analíticas muy significativas, como la distinción entre consciencia fenoménica y consciencia de acceso, y guarda relaciones lógicas con posiciones metafísicas muy bien fundadas como el epifenomenalismo o el dualismo de propiedades. A su turno, constituye una forma de acomodar la consciencia fenoménica en el marco general de la evolución biológica, en la medida en que permite vislumbrar cómo puede la teoría de la evolución dar cuenta del surgimiento de la consciencia fenoménica sin que ello implique asignarle una función que la haga objeto de la selección natural. Desde luego, la hipótesis en sí misma no constituye una explicación evolutiva de la consciencia fenoménica, pero permite visualizar una forma que tal explicación podría adoptar a partir de conceptos centrales en la comprensión biológica como el de enjuta o accidente evolutivo.

Los intentos de echar por tierra la hipótesis BAV, al menos en lo concerniente a la literatura revisada, difieren de maneras muy interesantes en su

enfoque y alcance. En algunos casos, se intenta mostrar que la consciencia fenoménica exhibe las mismas características que en otros casos han sido suficientes para establecer el carácter adaptativo de un rasgo, aun sin determinar ninguna función. En otros, se propone dejar entre paréntesis la indagación por el carácter adaptativo de la consciencia fenoménica a la espera de establecer el carácter adaptativo de los elementos que la componen, razonamiento en el marco del cual se logra poner en duda el lugar que los zombis han ocupado en esta discusión y, consecuentemente, retirar uno de los apoyos a la hipótesis BAV en su relación con el funcionalismo. Por último, en otro caso se propone conjuntamente una función para la consciencia fenoménica que contradice la hipótesis BAV y que simultáneamente pone en cuestión la distinción entre consciencia fenoménica y la consciencia de acceso.

La consideración crítica de estos argumentos contrarios a la hipótesis BAV nos conduce a pensar que, por las diversas razones que se han indicado, ninguno de ellos logra refutarla fuera de toda duda razonable. Sin embargo, sí logran poner de manifiesto algunas de sus debilidades. En primer lugar, parece sensato afirmar que el experimento mental de imaginar zombis que compiten en un mismo nicho ecológico con organismos fenoménicamente conscientes no constituye un argumento a favor de la hipótesis BAV, a diferencia de lo que puede intuirse en un primer momento. En segundo lugar, la hipótesis de que la función de la consciencia fenoménica es generar experiencias subjetivas -al margen de los problemas que implica desde una perspectiva fisicalista- permite reafirmar que, si logramos establecer una función para la consciencia fenoménica que permita dar cuenta de un tipo de comportamiento que no hemos podido verificar en agentes no conscientes (como el comportamiento creativo y flexible), se demostraría el carácter adaptativo de la consciencia fenoménica y se mostraría, también, que los razonamientos sobre el *petit mal* no estaban, en última instancia, tan equivocados como afirmaba Block. Sin embargo, sigue tratándose de una hipótesis que debe

superar los problemas que señalamos y encontrar apoyos empíricos que la sustenten.

De este modo, la hipótesis BAV se muestra como una contraintuitiva pero potente especulación, que cuenta con numerosos argumentos a favor, aunque no tan decisivos como sus defensores los consideran. Su relación con sólidas perspectivas metafísicas sustenta su plausibilidad, y no están exentas de problemas las perspectivas que pretenden refutarla. Si fuera correcta, lograría hacer encajar la consciencia fenoménica en el marco de la teoría de la evolución, desafío que no ha sido logrado por otros medios hasta el presente.

La posibilidad de que la hipótesis BAV sea correcta tiene implicancias sumamente interesantes en otras áreas de la reflexión filosófica que no fueron abarcadas en este trabajo. Particularmente reviste interés las consecuencias que dicha hipótesis puede tener en uno de los supuestos más fuertes -y también más controversiales- de la discusión ética, a saber: la existencia o no del *libre albedrío* en los seres humanos. Se trata de un debate que tangencialmente se puede visualizar cuando se discutió acerca de la causación mental aparente, así como el modo en el que la hipótesis de Cleeremans y Tallon-Baudry pretende dar cuenta del impulso interno a la acción. A su vez, la discusión ética sobre la naturaleza, la existencia y el alcance del libre albedrío, tiene consecuencias ineludibles en el campo de la filosofía política. Se trata de temas que, a nuestro juicio, revelan la centralidad que tiene y tendrá en el futuro la investigación sobre la consciencia en los más diversos ámbitos de nuestra comprensión científica y filosófica sobre el ser humano, la naturaleza y la sociedad.

Como en casi todos los ámbitos de la reflexión filosófica, y particularmente en lo que concierne a la filosofía de la mente, una indagación como la que hemos llevado adelante genera más incógnitas y problemas que respuestas certeras y definitivas. Cada puerta que se abre en la dirección de una determinada línea de investigación ameritaría una profundización mucho mayor

que la que el alcance de este trabajo permite. Sin embargo, los tramos recorridos en cada dirección permiten mostrar que la discusión filosófica ulterior sobre la consciencia fenoménica y su explicación evolutiva deberá lidiar, de un modo u otro, con esta singular hipótesis.

# Referencias bibliográficas


Alexander, S. (1920). *Space, Time and Deity*. Dover Publications.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Bayne, T., & Montague, M. (Eds.). (2011). *Cognitive Phenomenology*. Oxford University Press.

Block, N. (1978). Troubles with Functionalism. En C. W. Savage (Ed.), *Perception and Cognition. Issues in the Foundations of Psychology.* (pp. 261-325). University of Minnesota Press. https://doi.org/10.7551/mitpress/2111.003.0006

Block, N. (1991). Evidence against epiphenomenalism. *Behavioral and Brain Sciences1*, *14*(4), 670-672.

Block, N. (1992). Begging the question against phenomenal consciousness. *Behavioral and Brain Sciences*, *15*(2), 205-206.

Block, N. (1993). Review of Consciousness Explained by Daniel C. Dennett. *Journal of Philosophy*, *90*(4), 181-193. https://doi.org/10.5840/jphil199390448

Block, N. (1994). Functionalism; Qualia. En *A companion to philosophy of mind*. Blackwell.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227-247. https://doi.org/10.1017/S0140525X00038188

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200-219. https://doi.org/10.31812/apd.v0i14.1838

Chalmers, D. J. (1999). *La mente consciente*. Gedisa.

Cleeremans, A., & Tallon-Baudry, C. (2021). *The function of consciousness is to generate experience*. https://doi.org/https://doi.org/10.31234/osf.io/jfpw2

Cohen, M. A., & Dennett, D. C. (2012). Response to Fahrenfort and Lamme: defining reportability, accessibility and sufficiency in conscious awareness. *Trends in Cognitive Sciences*, *16*(3), 139-140. https://doi.org/10.1016/j.tics.2012.01.002

Coslett, H. B., & Saffran, E. M. (1994). Mechanisms of implicit reading in alexia. En M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision. Collected tutorial essays.* (pp. 299-330). Erlbaum.

Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, *72*(20), 741-765. https://doi.org/10.2307/2024640

Damasio, A. (1999). *El Error de Descartes: La razon de las emociones*. Andrés Bello.

Dennett, D. C. (1995). *La conciencia explicada*. Paidós.

Dennett, D. C. (2015). *Bombas de intuición y otras herramientas de pensamiento*. FCE.

Dennett, D. C. (2019). Welcome to Strong Illusionism. *Journal of Consciousness Studies*, *26*(9-10), 48-58.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & Van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, *311*(5763), 1005-1007. https://doi.org/10.1126/science.1121629

Dobzhansky, T. (1997). *La genética y el origen de las especies*. Circulo de Lectores.

Edelman, G. M. (1989). The Remembered Present; A Biological Theory of Consciousness. En *Noûs* (Vol. 28, Número 2). Basic Books. https://doi.org/10.2307/2216056

Etcoff, N. L., Freeman, R., & Cave, K. R. (1991). Can we lose memories of faces? Content specificity and awareness in a prosopagnosic. *Journal of cognitive neuroscience*, *3*(1), 25-41. https://doi.org/10.1162/JOCN.1991.3.1.25

Fahrenfort, J. J., & Lamme, V. A. F. (2012). A true science of consciousness explains phenomenology: comment on Cohen and Dennett. *Trends in Cognitive Sciences*, *16*(3), 138-139. https://doi.org/10.1016/j.tics.2012.01.004

Flanagan, O. (1991). *The Science of the Mind* (2nd ed.). Bradford.

Flanagan, O. (1993). *Consciousness Reconsidered*. Bradford. https://doi.org/10.2307/2185744

Flanagan, O., & Polger, T. (1995). Zombies and the Function of Consciousness. *Journal of Consciousness Studies*, *2*(4), 313-321.

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature*. Cambridge University Press.

Gould, S. J. (2004). *La estructura de la teoría de la evolución*. Tusquets.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London - Biological Sciences*, *205*(1161), 581-598. https://doi.org/10.1098/rspb.1979.0086

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology*, *1*(N/A), 4-15. https://doi.org/10.1017/S0094837300004310

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934-946. https://doi.org/10.1038/nrn2497

Hoffman, D. D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. W. W. Norton.

Horgan, J. (2023). A 25-Year-Old Bet about Consciousness Has Finally Been Settled - Scientific American. *Scientific American*.

Huxley, T. H. (1874). On the hypothesis that animals are automata, and its history. *Fortnightly Review*, *16*, 555-580. https://doi.org/10.1038/010362a0

James, W. (1892). The Stream of Consciousness. En W. James (Ed.), *Psychology* (pp. 71-82). Duke University Press.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15*(4), 673-692. https://doi.org/10.1016/j.concog.2006.09.004

Kim, Jaegwon. (1992). «Downward causation» in emergentism and nonreductive physicalism. En A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or reduction?: Prospects for nonreductive physicalism* (pp. 119-138). De Gruyter.

Kim, Jaegwon. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press. https://doi.org/10.2307/2108401

Kim, Jaewgon. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. The MIT Press.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9780511623486

Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J. Y., & Jordan, A. L. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLoS Biology*, *17*(2), 1-17. https://doi.org/10.1371/journal.pbio.3000021

Kriegel, U. (2019). The Value of Consciousness. *Analysis*, *79*(3), 503-520. https://doi.org/10.1093/analys/anz045

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Lassevich, C. L. (2020). *El Problema Difícil de la Conciencia y el Argumento del Conocimiento de Frank Jackson*. Universidad de la República.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431-439. https://doi.org/10.1016/j.neuron.2009.09.040

Lee, A. Y. (2019). Is consciousness intrinsically valuable? *Philosophical Studies*, *176*(3), 655-671. https://doi.org/10.1007/s11098-018-1032-8

Levine, J. (1983). Materialism and qualia. The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354-361.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in the initiation of action. *Behavioral and Brain Sciences*, *8*(4), 529–566.

Lloyd, E. A. (2005). *The Case of Female Orgasm: Bias in the Science of Evolution*. Harvard University Press.

Lucero M., B., & Muñoz-Quezada, M. T. (2015). Estudios en Visión Ciega (Blindsight): aportes fundamentales en la neurociencia y su proyección hasta nuestros días. *Ajayu Órgano de Difusión Científica del Departamento de Psicología UCBSP*, *13*(1), 39-53.

Marcel, A. J. (1986). Consciousness and processing: Choosing and testing a null hypothesis. *Behavioral and Brain Sciences*, *9*(1), 40-41. https://doi.org/10.1017/S0140525X00021439

Mayr, E. (1982). The Growth of Biological Thought. En *American Anthropologist* (Vol. 88, Número 2). Harvard University Press.

Morsella, E. (2005). The function of phenomenal states: Supramodular interaction theory. *Psychological Review*, *112*(4), 1000-1021. https://doi.org/10.1037/0033-295X.112.4.1000

Nagel, T. (1974). What Is It Like To Be a Bat? *The Philosophical Review*, *83*(4), 435-450. https://doi.org/10.1017/s1477175616000336

Nichols, S., & Grantham, T. (2000). Adaptive complexity and phenomenal consciousness. *Philosophy of Science*, *67*(4), 648-670. https://doi.org/10.1086/392859

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259. https://doi.org/10.1037/0033-295X.84.3.231

Nozick, R. (1974). *Anarchy, State and Utopia*. Basic Books.

Ostrom, J. N. (1976). Archaeopteryx and the origin of birds. *Biological Journal of the Linnean Society*, *8*(2), 91-182. https://doi.org/10.1111/j.1095-8312.1976.tb00244.x

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707-727. https://doi.org/10.1017/S0140525X00081061

Polger, T., & Flanagan, O. (2002). Consciousness, adaptation and epiphenomenalism. En J. Fetzer (Ed.), *Consciousness evolving* (pp. 21-41). John Benjamins.

Putnam, H. (1972). Mind, Language and Reality. En *Journal of Pragmatics* (Vol. 2, Números 5-6). Cambridge University Press.

Revonsuo, A. (2005). *Inner presence: consciousness as a biological phenomenon*. The MIT Press.

Robinson, Z., Maley, C. J., & Piccinini, G. (2015). Is Consciousness a Spandrel? *Journal of the American Philosophical Association*, *1*(2), 365-383. https://doi.org/10.1017/apa.2014.10

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*(3), 329-359. https://doi.org/10.1007/BF00355521

Rosenthal, D. M. (2006). *Consciousness and Mind*. Oxford University Press.

Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, *46*(3), 829-840. https://doi.org/10.1016/j.neuropsychologia.2007.11.012

Sacks, O. (2017). *El hombre que confundió a su mujer con un sombrero* (14th ed.). Anagrama.

Schacter, D. L. (1989). On the relation between memory and consciousness: Dissociable interactions and conscious experience. En *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 335-389).

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417-424. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, *13*(4), 585-596. https://doi.org/10.1017/S0140525X00080304

Searle, J. R. (1992). *The Rediscovery of the Mind* (MIT Press (Ed.)). https://doi.org/10.7551/mitpress/5834.001.0001

Sergent, J., & Poncet, M. (1990). From covert to over recognition of faces in a prosopagnosic patient. *Brain*, *113*(4), 989-1004. https://doi.org/10.1093/BRAIN/113.4.989

Seth, A. (2008). Functions of Consciousness. En *Encyclopedia of Consciousness*.

Siewert, C. P. (1998). *The Significance of Consciousness*. Princeton University Press.

Sober, E. (1990). Putting the Function Back into Functionalism. En W. G. Lycan (Ed.), *Mind and Cognition: A Reader*. Blackwell.

Theurer, K. L., & Polger, T. W. (2018). Is consciousness an adaptation? En R. Joyce (Ed.), *The Routledge Handbook of Evolution and Philosophy* (pp. 252-267). Routledge.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140167. https://doi.org/10.1098/rstb.2014.0167

Tye, M. (2021). *Qualia (Stanford Encylopedia of Philosophy)*. https://plato.stanford.edu/entries/qualia/

Van Gulick, R. (1989). What difference does consciousness make? *Philosophical Topics*, *17*(1), 211-230.

Van Gulick, R. (2022). *Consciousness (Stanford Encyclopedia of Philosophy)*. https://plato.stanford.edu/entries/consciousness/

Wedel, M. J. (2012). A monument of inefficiency: The presumed course of the recurrent laryngeal nerve in sauropod dinosaurs. *Acta Palaeontologica Polonica*, *57*(2), 251-256. https://doi.org/10.4202/app.2011.0019

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480-492. https://doi.org/10.1037/0003-066X.54.7.480

Weiskrantz, L. (2012). Some contributions of neuropsychology of vision and memory to the problem of consciousness. En A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science* (pp. 183-199). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198522379.003.0008

Wright, L. (1973). Functions. *The Philosophical Review*, *82*(2), 139. https://doi.org/10.2307/2183766