



**VNiVERSiDAD  
D SALAMANCA**

Facultad de Filosofía

Máster Interuniversitario en Lógica y Filosofía de la ciencia

**Trabajo de Fin de Máster**

**2022-2023**

**La evolución de la autoconsciencia**

Simón Botana Meyerderks

Tutor: Dr. Fernando Martínez Manrique

# Índice

1. Introducción.....	5
2. ¿Qué es la consciencia?.....	7
3. La importancia evolutiva de la consciencia.....	8
4. La separación como origen de la vida y ¿de la consciencia? .....	10
5. El origen de los sistemas nerviosos: generando una unidad desde la pluralidad .....	12
6. Los primeros sustratos de autoconsciencia en organismos pluricelulares .....	15
6.1 Reaferencia y descarga corolaria.....	15
6.2 La psicología ecológica formulada por Gibson .....	18
7. La aparición de los modelos internos y la capacidad conceptual .....	19
8. Pausas para tomar decisiones y comprensión de la propia existencia.....	24
8.1 El surgimiento de simulaciones predictivas .....	24
8.2 Perfeccionamiento del proceso de evaluación mediante la metacognición .....	28
9. Breve apunte sobre el rol de las sensaciones fenoménicas en el desarrollo de la autoconsciencia .....	32
10. Conclusión .....	34

## **Resumen**

Este ensayo intenta resolver la paradoja lingüística de la autoconsciencia, la cual parece indicar que deben existir nociones de autoconsciencia previas al origen del lenguaje. Así, se buscan sustratos prelingüísticos de autoconsciencia en animales no-humanos dentro de un marco naturalista, siendo el principal objetivo del trabajo identificar los mecanismos cognitivos más relevantes que han permitido el origen de la autoconsciencia, estableciendo el orden en el que surgieron, para así exponer cómo distintos grados de autoconsciencia fueron apareciendo a lo largo de la evolución biológica. Se ha tratado de aunar los conocimientos científicos y filosóficos actuales para elaborar un relato coherente, de carácter especulativo, sobre la evolución de la autoconsciencia. Este relato se basa en la visión de que la autoconsciencia se ha construido a través de mecanismos cognitivos con una función evolutiva definida. Se da especial importancia a la complejización de algunos procesos cognitivos presentes en organismos pluricelulares y se identifican como la principal fuente de los avances más significativos que posibilitaron la aparición de grados más elevados de autoconsciencia. En primer lugar, se describen algunos sustratos primitivos de la autoconsciencia, como los elementos propioceptivos de la percepción o los procesos reaférentes. Después, se intenta exponer cómo animales pluricelulares fueron capaces de analizar mejor su entorno a través de modelos internos, los cuales fueron posteriormente empleados en 'pausas para tomar decisiones' para evaluar posibilidades futuras de acción. Se argumenta que a través de la complejización de esas pausas mediante procesos metacognitivos se dio lugar a un grado de autoconsciencia suficiente como para resolver la paradoja lingüística de la autoconsciencia.

**Palabras clave:** modelo interno, simulación predictiva, consciencia, reaférentia, propiocepción, metacognición.

## Summary

This paper attempts to resolve the linguistic paradox of self-consciousness, which seems to indicate that there must be notions of self-consciousness prior to the origin of language. Thus, we search for pre-linguistic substrates of self-consciousness in non-human animals within a naturalistic framework, the main objective of the work being to identify the most relevant cognitive mechanisms that have allowed the origin of self-consciousness, establishing the order in which they arose, in order to expose how different degrees of self-consciousness appeared throughout biological evolution. An attempt has been made to bring together current scientific and philosophical knowledge to develop a coherent speculative account of the evolution of self-consciousness. This account is based on the view that self-consciousness has been constructed through cognitive mechanisms with a defined evolutionary function. Special importance is given to the complexization of some cognitive processes present in multicellular organisms and they are identified as the main source of the most significant advances that made possible the emergence of higher degrees of self-consciousness. First, some primitive substrates of self-awareness are described, such as proprioceptive elements of perception and reafereent processes. Then, an attempt is made to show how multicellular animals were able to better analyze their environment through internal models, which were subsequently employed in 'decision pauses' to evaluate future possibilities for action. It is argued that through the development of these pauses by metacognitive processes, a sufficient degree of self-consciousness was achieved to resolve the linguistic paradox of self-consciousness.

**Keywords:** internal model, predictive simulation, self-awareness, reafereent, proprioception, metacognition.

## 1. Introducción

La intención de este ensayo es trazar un relato del proceso evolutivo que ha llevado a ciertos seres vivos a ser conscientes de que son, es decir, a poseer autoconsciencia. En esta cuestión, partimos desde la certeza de que es posible que un ser vivo sea capaz de tener consciencia de su propia existencia, ya que nosotros, los humanos, somos plenamente conscientes tanto de que existimos como de que dejaremos de hacerlo tras nuestra muerte. Un humano sabe por cuenta propia que es consciente de que es, experimenta ese hecho con una certeza incontestable. Tal es la certeza de esta intuición, que llevó a Descartes a sustentar toda su filosofía de la mente sobre ella. Pero, ¿puede alcanzar un humano una certeza semejante en cuanto a la autoconsciencia de otros seres? La respuesta es no. Sin embargo, existen evidencias que nos permiten trazar especulaciones cuando tratamos de comprobar si otros seres vivos son autoconscientes.

En el caso de otros humanos, el lenguaje nos proporciona una herramienta muy fiable para juzgar su autoconsciencia. Una persona puede comunicarnos sus pensamientos en un lenguaje que compartimos, haciéndonos saber que posee la misma certeza que nosotros respecto a su propia existencia. Esta verbalización de sus pensamientos nos acerca a su punto de vista subjetivo. Pero, ¿qué ocurre cuando queremos averiguar si seres vivos que no poseen lenguaje —al menos, un lenguaje que seamos capaces de comprender— son autoconscientes o no?

Existen, a grandes rasgos, dos opiniones ampliamente difundidas entre filósofos de la actualidad en cuanto a la atribución de autoconsciencia a otros seres vivos: (i) el lenguaje es la base sobre la que se constituye la autoconsciencia, funcionando como una brecha, un antes y un después en la historia evolutiva, y como la capacidad de emplearlo parece exclusiva de nuestra especie, somos el único tipo de seres vivos que tiene consciencia de su propia existencia; (ii) la autoconsciencia no depende de la capacidad lingüística de un animal, sino de la presencia de mecanismos cognitivos clave que parecen estar presentes en diversas especies de animales no-humanos y, a pesar de que sea difícil juzgar con claridad cuáles, hay evidencias empíricas que muestran que por lo menos algunas especies no-humanas dan muestras de tener consciencia de su propia existencia.

En este ensayo me posicionaré junto a aquellos filósofos que sostienen la segunda opción. Uno de los principales argumentos a favor de que la autoconsciencia es previa al surgimiento del lenguaje lo expone Bermúdez (2000) en su obra *La paradoja de la autoconsciencia*. En este libro, el filósofo se percata de la circularidad en que incurrimos cuando consideramos el uso del pronombre ‘yo’ como necesario para albergar una noción de primera persona. Bermúdez se pregunta: si solo aquellos que aprenden a usar

pronombres en primera persona pueden poseer una noción de primera persona, entonces, ¿quién empezó a utilizar estos pronombres y cómo fue eso posible? Parece implausible que los agentes que emplearon por primera vez el pronombre 'yo' fueran incapaces hasta *entonces* de poseer una noción de sí mismos que les permitiera comprender dicho término. Por esta razón, Bermúdez sostiene que deben existir nociones de autoconsciencia primitivas de carácter prelingüístico, lo que tiene como consecuencia que animales sin lenguaje podrían poseer dichas nociones. Estas conclusiones invitan a pensar la aparición de la autoconsciencia como un proceso gradual que se extendió durante largos periodos de la evolución biológica.

La pretensión de este ensayo ha sido identificar de manera más precisa las nociones primitivas que parecen resultar necesarias para resolver la paradoja lingüística de la autoconsciencia dentro de un marco naturalista. La aparición gradual de la autoconsciencia necesita justificarse mediante una selección de mecanismos clave que podrían haberla desarrollado a lo largo de la evolución. El objetivo principal de este trabajo es identificar dichos mecanismos y establecer el orden en el que surgieron, para así construir un modelo teórico que describa la aparición de mecanismos cognitivos que permiten un conocimiento de la propia existencia. Se intenta, más aún, señalar qué eventos de la historia evolutiva dieron lugar a esos mecanismos y en qué orden, indicando con ejemplos de especies concretas aquellas clases de animales que parecen incluirse en cada etapa descrita. Así, se ha tratado de aunar los conocimientos científicos y filosóficos actuales para elaborar un relato coherente, de carácter especulativo, sobre la evolución de la autoconsciencia. Este relato se basa en una visión de que la autoconsciencia se ha construido a través de mecanismos cognitivos con una función evolutiva definida. Se da especial importancia a la complejización de los procesos cognitivos presentes en organismos pluricelulares como la principal fuente de los avances significativos que posibilitaron la aparición de grados más elevados de autoconsciencia. A pesar del carácter especulativo del ensayo, se ha intentado revestir de bibliografía relevante aquellos puntos donde se establecen las decisiones teóricas que marcan el recorrido defendido. A pesar de que se sostiene un relato concreto del origen y desarrollo de la autoconsciencia, en ningún momento pretende ofrecerse una falsa seguridad donde aún no es razonable tenerla. El ensayo se centra, además, en la evolución biológica producida antes de la aparición del ser humano, con la intención de justificar qué mecanismos cognitivos posibilitaron el desarrollo posterior de los elevados grados de autoconsciencia que se alcanzaron durante la evolución cultural de nuestra especie.

## 2. ¿Qué es la consciencia?

La palabra consciencia tiene un significado maleable y difuso, lo cual ha sido aprovechado por filósofos para emplearla de manera diferente en diversos contextos. Vimal (2009) revisó la literatura filosófica del siglo XX y distinguió hasta cuarenta usos distintos del término ‘consciencia’, por lo que conviene detenernos un momento y especificar cómo se empleará esta palabra a lo largo del presente ensayo. Sin embargo, como la generación de la consciencia en animales y seres humanos sigue siendo un misterio en la actualidad y no se conoce con exactitud qué procesos físicos dan lugar a la consciencia ni cómo funcionan, cualquier definición de consciencia necesariamente tiene un carácter tentativo y no llega a una precisión tan elevada como sería deseable.

Una división muy extendida que nos puede ayudar a concretar el término ‘consciencia’, la ofrece Chalmers (2003), quien distingue, a grandes rasgos, dos tipos de consciencia. La primera es una consciencia funcional, materialista, relacionada con la discriminación de estímulos, el manejo de información adquirida a través de órganos sensoriales, la monitorización de estados internos y el control del comportamiento. Esta consciencia está ligada al concepto de *atención consciente* (*awareness*) y sirve para juzgar las capacidades cognitivas de un organismo a la hora interactuar con el entorno en un cierto momento. Una criatura que está dormida no es consciente de su entorno, mientras que la misma criatura sí lo es cuando está despierta. No es difícil ver que es una cuestión de grado atribuir este tipo de consciencia: seres con órganos sensoriales y cerebros más complejos son más conscientes que otros que basados en procesos más simples. Así, construir una historia evolutiva de este tipo de consciencia supone trazar el recorrido por el que han ido surgiendo mecanismos cognitivos y órganos sensoriales más avanzados, los cuales han permitido a ciertos organismos adquirir más información de su entorno y ser capaces de comprenderlo con mayor profundidad, pudiendo dar así respuestas más variadas y mejor adaptadas al medio que les rodea.

La segunda concepción de consciencia propuesta por Chalmers estaría ligada a la experiencia y el ‘cómo se siente’ —desde el punto de vista de primera persona— encontrarse en un cierto estado mental. Esta visión de la consciencia la ofreció ya Nagel (1974) en su influyente artículo *¿Cómo se siente ser un murciélago?*. Según Nagel, un organismo tan solo será consciente si existe un ‘cómo se siente’ ser ese organismo, es decir, si hay una serie de sensaciones y emociones subjetivas que ese organismo siente como experiencias propias. Ya Nagel advirtió que este tipo de consciencia parece imposible de acceder para aquellos que no son el propio organismo. Esta conclusión llevó a Chalmers a afirmar que este tipo de consciencia supone el ‘problema duro de la

consciencia', ya que resulta muy difícil atacar un problema filosófico del cual solo podemos obtener información en primera persona, de nuestra propia experiencia individual. De ahora en adelante llamaremos a esta segunda concepción *consciencia fenoménica* y llamaremos *consciencia funcional* a la primera concepción.

La autoconsciencia es una forma específica de consciencia dirigida hacia uno mismo. Un ser autoconsciente es consciente de que es, es decir, incorpora su propia existencia a la representación que realiza del mundo. Como veremos en las siguientes secciones, existen diversos grados de autoconsciencia que permiten a un organismo tener un conocimiento mayor o menor de sí mismo, lo que invita a establecer tentativamente 'grados' de autoconsciencia a distintas especies. Resulta complicado saber con exactitud el grado de autoconsciencia de un animal que no es capaz de comunicarnos sus pensamientos. Sin embargo, existen diversos experimentos, recogidos a lo largo del ensayo, que indican distintos grados de autoconsciencia en diversas especies.

### **3. La importancia evolutiva de la consciencia**

El problema principal que tiene la división propuesta por Chalmers, expuesta en la sección anterior, es que no se ofrece ningún *motivo evolutivo* por el cual los dos tipos de consciencia que distingue podrían haber surgido. La aparición y permanencia de la consciencia funcional parece fácil de justificar debido a la gran ventaja de supervivencia que tiene un organismo capaz de discriminar estímulos, monitorizar estados internos y controlar su comportamiento de acuerdo a ellos; pero la consciencia fenoménica, tal como es presentada por Chalmers, no parece generar ventaja evolutiva alguna a los seres que la poseen. Esto nos llevaría consecuentemente a adoptar alguna clase de epifenomenalismo: la visión de que la consciencia fenoménica no tiene ninguna consecuencia en los procesos físicos que se dan en el cerebro y, con ello, de que no tiene efecto alguno en el comportamiento<sup>1</sup>.

Aunque este ensayo no pueda dedicar suficiente espacio a discutir pausadamente los argumentos a favor y en contra del epifenomenalismo, creo que resulta un error asumir que la consciencia, tanto en su concepción funcionalista como en la fenoménica —estrechamente relacionadas—, no ha jugado un papel importante en la evolución biológica. La visión de que la consciencia sí ha constituido un factor decisivo en la historia evolutiva ya ha sido defendida con anterioridad (James, 1879; Popper, 1978; Lindahl, 1997).

---

<sup>1</sup> Para una explicación del epifenomenalismo de uno de sus principales propulsores, véase Huxley (1874). Para una versión más actual relacionado con la evolución, véase Robinson (2007).



Barbieri (2011) ofrece un marco teórico que sí aborda de manera directa las posibles causas del origen de la consciencia durante la evolución. Establece que existen tres tipos de 'cerebro', los cuales pueden coexistir en la misma criatura. El primero es el *cerebro cibernético*, que trabaja mediante una continua reacción en cadena entre inputs y outputs, funcionando prácticamente como un conjunto de servomecanismos que emplean ciclos de retroalimentación para generar respuestas. A pesar de generar respuestas de gran capacidad adaptativa, un cerebro cibernético convierte a un organismo en una marioneta del medio y los estímulos que este genera. El segundo es el *cerebro intuitivo*, cuyas órdenes para la acción provienen del propio sistema y no del exterior. Un organismo con un cerebro intuitivo tiene impulsos internos que le dan cierta independencia del entorno donde se encuentra. Un animal con hambre tendrá una motivación interior para explorar ambientes desconocidos, incluso en la ausencia de señales que podrían retroalimentar una respuesta cibernética para avanzar. El tercer tipo es el *cerebro interpretativo*, que surge cuando un organismo es capaz de memorizar estímulos y aprender de ellos para responder de manera más adecuada en futuras decisiones. Esto requiere reorganizar la información que le llega a través de los sentidos, reconocer patrones y, además, establecer un *significado* para las percepciones que recibe. A través de un proceso abductivo, el organismo genera decisiones que, aunque partan de un conocimiento incompleto de la realidad, le permiten ajustarse a ella lo mejor posible.

La división de Barbieri resulta ciertamente artificial. No se han encontrado evidencias de que dicha división tenga un correlato real en cerebros de organismos conocidos. Cuenta, además, con deficiencias teóricas importantes, ya que la división entre cerebro y el resto del organismo resulta demasiado marcada. Sin embargo, la división de Barbieri resulta útil por dos razones: (i) nos permite investigar el origen de ciertas capacidades cognitivas vinculadas al origen de la autoconsciencia a partir de la identificación de aquellos organismos que poseen un cerebro exclusivamente cibernético cuyos descendientes evolutivos dan muestras de poseer un cerebro intuitivo e interpretativo; y (ii) ofrece una justificación evolutiva para los tres tipos de cerebros y, con ello, del origen de la consciencia funcional y fenoménica. El cerebro cibernético parece vincularse a todos los procesos que funcionan de manera inconsciente. El cerebro intuitivo se vincula a la capacidad de experimentar sensaciones, las cuales están estrechamente relacionadas con la consciencia fenoménica. El cerebro interpretativo es aquel que más parece vincularse a la consciencia funcional, con aquellos procesos de evaluación y toma de decisiones en los que el organismo puede participar de manera consciente.

Antes de comenzar, conviene dejar claro que la emergencia de la consciencia no fue un proceso evolutivo lineal. La evolución se da de forma ramificada y multitud de animales han evolucionado capacidades cognitivas admirables a través de procesos independientes entre sí. Es muy importante entender que tanto la inteligencia como la autoconsciencia ha surgido de manera independiente en diversas ramas evolutivas desconectadas (Roth, 2015). Establecido esto, empecemos.

#### **4. La separación como origen de la vida y ¿de la consciencia?**

El origen de los primeros seres vivos terrestres, los organismos unicelulares, trae consigo una distinción fundamental entre el propio cuerpo y el entorno que este habita. En el caso de las primeras células, una membrana lipídica actuó de barrera física, encerró su contenido y generó un subsistema dentro de otro sistema. Se dio lugar así a una diferenciación necesaria para cualquier forma de vida, aquella que produce un ser mediante la separación: la creación de un ‘interior’ y un ‘exterior’, establecidos mediante una frontera. Esta separación básica ha sido formulada mediante el término *yo-corporeizado* (*body-self*) por Jékely et al. (2021, p.3), el cual definen como “una diferenciación no-arbitraria con el entorno [...] que delimita una unidad [...], la cual se convierte en la base para futuras innovaciones evolutivas”<sup>2</sup>.

El hecho de que el *yo-corporeizado* esté presente desde el origen de la vida ha llevado a algunos filósofos y científicos a postular una consciencia para estos seres vivos unicelulares primigenios. Esto llevó a la científica Lynn Margulis (2001, p. 58), reconocida experta en microbiología evolutiva, a afirmar: “El cerebro aparece después, pero la consciencia, la atención consciente (*awareness*) del ambiente, empieza con el origen de la vida misma”. En la misma línea, Reber y Baluska (2021) afirman que la consciencia se originó por primera vez en microorganismos procariotas, al comienzo de la evolución biológica. Esta teoría tiene como nombre *base celular de la consciencia* y tiene como hipótesis que todo organismo con paredes flexibles que posea algún tipo de sensibilidad sensorial de su entorno y capacidad de locomoción tiene las bases biológicas para tener una mente y una consciencia —tanto funcional como fenoménica—. Los mismos autores admiten que es una teoría especulativa sin evidencia experimental.

Los propulsores de esta teoría parecen equiparar el hecho de que las células poseen receptores sensoriales con la posesión de experiencia sensible consciente. La visión de que el mero uso de un receptor sensorial conduce a una experiencia consciente por sí misma conduce a alguna forma de pansiquismo o a un emergentismo muy relacionado

---

<sup>2</sup> La traducción del inglés original, al igual que las demás traducciones del ensayo, es propia.

con esta teoría, donde la consciencia surgiría de procesos muy básicos. A pesar de que haya habido varios defensores acérrimos de esta teoría (James, 1911; Strawson, 2006), el pansiquismo es una teoría que aún no se ha conseguido relacionar convincentemente con el paradigma científico actual que poseemos de la realidad, ya que no se han encontrado los sustratos fundamentales que podrían dar lugar a la consciencia. Asimismo, esta teoría posee la dificultad de explicar el origen de la consciencia sin aludir a mecanismos cognitivos complejos.

Trewawas (2017) y Reber y Balluska (2021) afirman que organismos celulares muestran comportamientos que solo pueden ser explicados satisfactoriamente si les concedemos una inteligencia tan avanzada que requeriría de consciencia. Los principales ejemplos de esta 'inteligencia celular' son experimentos donde organismos unicelulares son capaces de viajar largas distancias a través de laberintos manteniendo una dirección que parece dirigida, anticipando paredes y giros. Sin embargo, Tweedy et al. (2020) mostraron que esos comportamientos aparentemente inteligentes eran explicables mediante mecanismos muy simples, que se encajarían más bien con un proceso cibernético inconsciente más que con la consciencia ligada a procesos intuitivos o interpretativos. Es harto conocido el sistema que permite a organismos unicelulares navegar por un medio a favor o en contra de un gradiente tan solo con unos pocos sensores y órganos motrices a partir de un proceso llamado quimiotaxis<sup>3</sup>. Tweedy et al. descubrieron que, para navegar por los laberintos de manera eficiente, las células empleaban un proceso de autoquimiotaxis, generando un gradiente local por donde pasaban, el cual se renovaba mecánicamente a medida que las células avanzaban. De esta manera, las células eran capaces de sortear obstáculos y esquinas, e incluso encontrar atajos en los laberintos. La causa de esta capacidad no era una inteligencia extraordinaria, sino un sencillo mecanismo cibernético de quimiotaxis donde las células seguían un gradiente generado por otras células.

La teoría de base celular de la consciencia se enfrenta a una gran dificultad para explicar la emergencia de la única consciencia de cuya existencia podemos estar completamente seguros: la humana. Reber y Balluska (2021) afirman que, por algún proceso todavía sin descubrir, los billones de células de un cuerpo humano 'vuelcan' su consciencia al sistema nervioso para dar lugar a nuestra consciencia unificada. Sin embargo, como bien apuntan Mallat et al. (2021), un humano solo pierde su consciencia cuando sufre una lesión cerebral, mientras que la muerte de otro tipo de células no conlleva ninguna pérdida de consciencia, lo que habla en contra de esta idea.

---

<sup>3</sup> Para una descripción ilustrativa de la quimiotaxis véase el capítulo cuatro del libro de Ogas y Gaddam (2022).

En definitiva, la idea de que las células sean conscientes parece altamente improbable, sobre todo porque, sin caer en el pansiquismo —una teoría de dudosa validez en la actualidad—, no parece haber mecanismos cognitivos que justifiquen la atribución de consciencia a organismos unicelulares. Eso conlleva que las células no son autoconscientes en ningún grado. La única forma de autoconsciencia que podría atribuírseles sería esa distinción fundamental de una célula con su entorno, que le lleva a formar un *yo-corporeizado* tal y como lo definen Jékely et al. (2021). Sin embargo, este término no parece dar origen a la autoconsciencia, sino a la vida solamente. En este ensayo se hablará de qué mecanismos parecen ser responsables del surgimiento de la autoconsciencia a lo largo de la evolución, los cuales pertenecerían exclusivamente a organismos pluricelulares.

## **5. El origen de los sistemas nerviosos: generando una unidad desde la pluralidad**

En cierto punto de la evolución, algunas células comenzaron a unirse formando comunidades cooperativas. La unión de varias células trabajando conjuntamente pareció funcionar en términos de supervivencia y fue potenciado por presiones evolutivas, hasta que dio lugar a lo que hoy llamamos *organismos pluricelulares*, cuya característica fundamental es que la comunidad de células por la que están conformados compone un todo inseparable, es decir, una nueva unidad. La unión de muchas células, cuando estas funcionan coordinadamente, da lugar así a una nueva forma de *yo-corporeizado*, una estructura viviente más compleja, pero que consigue también distinguirse del medio mediante una frontera que marca un ‘interior’ y un ‘exterior’.

Los animales pluricelulares más simples y antiguos que se conocen son las esponjas, cuya única estructura es una membrana. Las esponjas más elementales poseen cuatro tipos de células, las cuales realizan un trabajo en equipo admirablemente coordinado. El origen de esta coordinación aún es investigado en la actualidad y no se tiene del todo claro cómo se ha llegado a producir. Lo que sí está claro es que cualquier tipo de coordinación entre células requiere de alguna forma de *comunicación*. En las esponjas, esta comunicación puede estudiarse a través de la observación de algunos movimientos —muy lentos— de contracción que pueden propagarse a lo largo de la esponja, donde la comunicación parece realizarse a través de moléculas señalizadoras o acciones mecánicas transmitidas de célula a célula por contacto (Leys y Meech, 2006).

La presión por obtener medios más rápidos y estables de comunicación celular parece que estuvo presente desde el origen de los organismos pluricelulares, ya que, muy recientemente, un grupo de investigadores situó en un tipo muy concreto de esponjas los

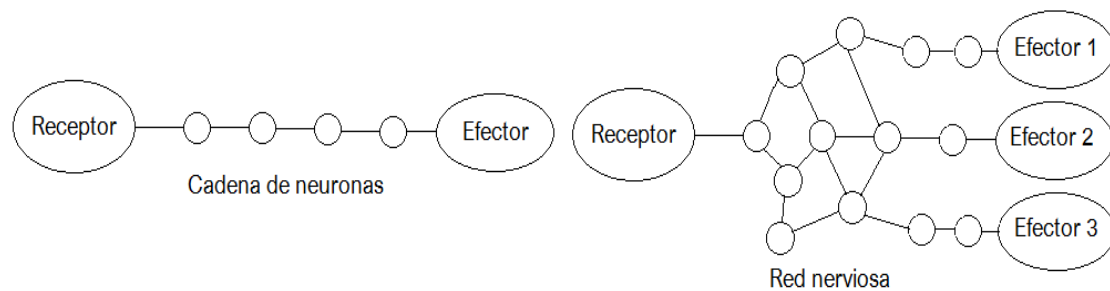
posibles orígenes del sistema nervioso (Musser et al., 2021). Haciendo un estudio genético sin precedentes en la especie *Spongilia lacustris*, identificaron 18 tipos celulares diferentes. Un tipo de ellos eran células neuroides secretoras, localizadas en las cámaras digestivas de las esponjas —huecos dentro de la esponja por donde los coanocitos hacen pasar el agua—. Estas células neuroides expresan genes importantes en la comunicación sináptica y parecen estar especializadas en la comunicación celular encargada de regular la alimentación, es decir, en fomentar un movimiento de los flagelos de los coanocitos cuando se produce una escasez de corriente.

El origen concreto de los primeros sistemas nerviosos, a pesar de estos avances, sigue siendo desconocido. Existen dos teorías que justifican su aparición. La primera se centra en la coordinación interna de un organismo pluricelular y fue postulada por Pantin (1956). Pantin especuló que la coordinación celular necesaria para gestionar un sistema motriz que emplease músculos era demasiado intrincada para ser realizada sin redes nerviosas. Esta estimación parece haberse confirmado, pues todos los animales con sistemas nerviosos poseen músculos y viceversa, siendo la única excepción a esta regla unos parásitos mixozoos cnidarios, que parecen haber sufrido una simplificación evolutiva posterior (Jékely et al., 2015).

La segunda teoría del origen del sistema nervioso, más reconocida, es la que vincula su aparición a la rápida respuesta a estímulos externos mediante la conexión de receptores con efectores mediante un arco reflejo. Un conocido precursor de esta teoría fue Parker (1919), quien especuló que los efectores, órganos capaces de generar acciones, evolucionaron primero, y que las neuronas permitieron conectar esos efectores a sensores concretos. Lo más probable es que el origen esté vinculado a ambos procesos, de los cuales pueden observarse precursores en animales sin sistema nervioso (Jékely et al., 2015).

En definitiva, es un hecho que, en cierto momento, aparecieron células que actuaron como intermediarias entre distintos tipos de células dentro de los animales pluricelulares, posibilitando una comunicación rápida y eficaz: las neuronas. Existen tres tipos de neuronas: (i) las *neuronas sensoriales*, que transforman las señales de un órgano sensorial en una señal nerviosa eléctrica, (ii) las *neuronas motoras*, que reciben una señal nerviosa y activan los órganos motrices y (iii) las *interneuronas*, que conectan los dos primeros tipos de neuronas actuando como un puente entre ellas.

Cuando las interneuronas conectan de manera directa las neuronas sensoriales con las motoras, se produce lo que llamamos acto reflejo. Nosotros los humanos aún tenemos varias de estas conexiones: parpadeamos cuando nuestros ojos perciben un aumento súbito en la intensidad de la luz y sufrimos arcadas cuando se estimula la garganta o la parte posterior de la boca. Sin embargo, en cierto punto las interneuronas dejaron de conectar de manera directa los órganos sensoriales y los órganos efectores. Las conexiones nerviosas de las interneuronas se complejizaron: un cable simple dio lugar a una estructura en red. Esto está ligado a que los receptores, cada vez más complejos, eran capaces de detectar diferencias más sutiles en el tipo y la intensidad de los estímulos. Esto fue aprovechado mediante la aparición de una red de neuronas intermedias, las cuales procesaban el tipo de señal enviada desde el receptor y lo clasificaban según varios parámetros. Tras el procesamiento, la red nerviosa activaba diferentes tipos de efectores según la categorización del estímulo que se hubiese realizado. Esto permitió coordinar y activar una gran cantidad de efectores mediante unos pocos receptores. En la *figura 1* puede verse la complejización del sistema nervioso ejemplificada de manera esquemática.



*Figura 1:* A la izquierda, una cadena simple de interneuronas que puede producir un acto reflejo. A la derecha, una red nerviosa con mayor número de interneuronas, capaz de establecer diferentes respuestas con un mismo receptor.

Las interneuronas aumentaron su densidad en ciertas zonas del cuerpo donde había más actividad, como en las extremidades con capacidad de movimiento o los órganos sensoriales, para así permitir una mayor variabilidad comportamental. Ciertos grupos de especies —como los artrópodos, los cefalópodos y los primeros vertebrados— evolucionaron después una agrupación de interneuronas llamativamente superior al resto en una zona llamada encéfalo, sufriendo a la vez un desplazamiento de la boca y de los principales órganos sensoriales hacia el mismo lugar, dando lugar así a lo que comúnmente se conoce como cabeza. El proceso por el cual se produjeron estos cambios fáciles de percibir en el registro fósil se conoce como encefalización.

Gracias a la mejora de la comunicación que introdujeron los sistemas nerviosos, las células pudieron coordinarse muchísimo más que antes, y los organismos pluricelulares comenzaron, cada vez más, a funcionar como una unidad plural. Esta unicidad no se dio por sí sola, sino que se construyó y perfeccionó a medida que mejoraron la comunicación y la coordinación de las células que conformaban un organismo. Como ya remarcaron Jékely et al. (2021), el *yo-corporeizado* de los organismos pluricelulares es una cuestión de grado precisamente porque se construye. Los sistemas nerviosos significaron un notable avance para la aparición de *yos-corporeizados* más marcados, es decir, unidades plurales que funcionaban más como un todo que otras.

## **6. Los primeros sustratos de autoconsciencia en organismos pluricelulares**

### **6.1 Reaferencia y descarga corolaria**

Durante el proceso de encefalización y complejización de las redes neuronales de los animales se dieron lugar ciertos procesos que suponen el primer avance significativo en la capacidad de autoconsciencia: los mecanismos reaferentes. La reaferencia es cualquier proceso donde una acción de un organismo es detectada por sus propios órganos sensoriales e identificada como tal. En algunas ocasiones, la reaferencia debe ser detectada con el fin de no generar una respuesta poco adaptativa por parte de un organismo frente a un estímulo que no ha producido ninguna posible presa ni ninguna amenaza, sino él mismo. En otras, la reaferencia puede ser empleada como una forma muy particular de percepción, como en el caso de la ecolocalización de los murciélagos.

Jékely et al. (2021) defienden la visión de que la reaferencia ya formó una parte crucial del funcionamiento de los seres vivos primigenios, incluidos los seres unicelulares. Exponen que la quimiotaxis que realizan organismos unicelulares es ya una forma de reaferencia, ya que estos organismos emplean sus propios movimientos para conseguir detectar gradientes. También argumentan que algunas esponjas, como la *Tethya wilhelma*, presentan un tipo de reaferencia por el cual ciertas células son capaces de detectar la corriente autogenerada en el interior de la esponja y mandar señales en caso de que sea insuficiente.

Los mecanismos reaferentes más interesantes se producen en organismos pluricelulares con sistema nervioso. La *descarga corolaria* hace referencia al proceso reaferente por el cual estos organismos son capaces de discriminar estímulos autoproducidos de aquellos que son externos. Podemos distinguir, a grandes rasgos, dos tipos de descarga corolaria: (i) aquella donde unas interneuronas inhibitorias bloquean una respuesta de tipo acto reflejo en momentos donde es probable que ese acto reflejo sea activado por la detección

de un estímulo que proviene del propio organismo; y (ii) aquella donde el organismo es capaz de realizar una *copia eferente* del estímulo emitido, para así sustraerlo por medio de la eliminación de aquellos estímulos que son similares a la copia eferente (Fukutomi y Carlson, 2020). Llamaremos al primer tipo simplemente descarga corolaria y al segundo tipo copia eferente. No resulta complicado intuir que la copia eferente es un proceso más complejo que se da en animales con una mayor capacidad de detección, procesamiento y discriminación de estímulos.

Demos dos ejemplos de descarga corolaria. La babosa marina *Pleurobranchaea californica*, que normalmente realiza un movimiento retráctil cuando detecta algún tipo de contacto en su velo oral, no presenta dicha respuesta mientras come. Davis et al. (1977) detectaron un par de interneuronas que generan una descarga corolaria en el circuito neuronal propio del proceso alimentario, suprimiendo la actividad retráctil de las neuronas motoras que respondían a la estimulación táctil, probablemente para evitar movimientos retráctiles autogenerados. El cangrejo de río americano (*Procambarus clarkii*) realiza unos rapidísimos volteos con la cola que le impulsan con potencia hacia atrás cuando detecta un estímulo táctil en su zona posterior. Los volteos de cola son movimientos bruscos que podrían fácilmente activar de nuevo los mismos mecanismos de huida que los iniciaron. Krasne y Bryan (1973) descubrieron una descarga corolaria que inhibía las neuronas que desencadenaban el mecanismo de huida durante los volteos, que seguramente surgió para evitar la activación de una nueva huida autogenerada.

Las descargas corolarias son una adaptación rígida y tienen la desventaja de que pueden inhibir las neuronas de un animal durante demasiado tiempo, dejándolo insensible a muchos estímulos. En animales más desarrollados, Holst y Mittelstädt (1950) propusieron la idea de que tal vez se generase una ‘imagen negativa’ en el propio organismo a partir de una *predicción* de cómo sería detectado un estímulo autogenerado concreto, para que así pudiese sustraerse ese estímulo del conjunto de estímulos detectados. De este modo, se podrían ignorar los estímulos autogenerados mientras se mantiene la posibilidad de reacción a estímulos exaferentes —generados por una causa exterior al organismo—. Esa ‘imagen negativa’ a la que se refirieron originalmente Holst y Mittelstädt es lo que aquí llamaremos *copia eferente*.

Las primeras evidencias empíricas de copia eferente se obtuvieron de peces mormíridos (*Mormyridae*), una familia de peces muy sociables y con gran inteligencia, los cuales son capaces de producir y detectar pequeños campos eléctricos que emplean para comunicarse entre ellos, orientarse y encontrar alimento. Las frecuencias de los campos eléctricos que producen los propios peces para orientarse y comunicarse son



considerablemente más altas que las bajas frecuencias que deben detectar para poder cazar. Bell (1981) investigó la manera en la que las neuronas electrosensibles de las líneas laterales de los peces mormíridos procesaban las señales eléctricas. Empleó peces paralizados y constató que el modelo teórico de Holst y Mittelstädt (1950) coincidía con sus observaciones.

¿Cómo consiguió esto? Primero, generó un estímulo eléctrico de manera artificial cada vez que el pez emitía uno propio. Esto generó picos de activación en las neuronas observadas, que en un principio detectaron la suma de ambos estímulos como una señal nueva. Tras conseguir que el pez entendiese como ‘propios’ este par de estímulos —esto ocurría cuando las neuronas dejaban de activarse cuando estos dos estímulos eran generados—, Bell dejó de añadir el segundo estímulo eléctrico. Esto tuvo como consecuencia que en las neuronas del pez se diesen unos picos de activación muy similares a la función inversa del estímulo añadido cada vez que detectaba sus propias emisiones, lo cual sugería que el pez había generado una ‘imagen negativa’ de la suma de ambos estímulos. Bell ya indicó que la estructura encargada de generar copias eferentes en los peces mormíridos parecía asemejarse al cerebelo de los mamíferos, lo cual recibió apoyo experimental cuando Brooks et al. (2015) concluyeron que el cerebelo juega un papel clave en la estimación de consecuencias sensoriales causadas por el propio movimiento en primates, que también parece funcionar a partir de mecanismos similares a los de la copia eferente.

Los mecanismos de descarga corolaria y copia eferente son los primeros indicios de autoconsciencia en organismos pluricelulares. Estos mecanismos surgen precisamente porque estos organismos están compuestos por una gran cantidad de células que cumplen funciones muy diversas dentro del organismo. La comunicación, a través de las neuronas —mediante una descarga corolaria o una copia eferente—, es lo único que permite a una parte del organismo saber que otra parte está realizando una acción que afectará su rendimiento. Esto quiere decir que, precisamente por ser pluricelulares, estos organismos han necesitado de este tipo de comunicación destinada a autoconocerse mejor. Las tareas y las funciones que realiza en total el organismo son llevadas a cabo por subgrupos de células separados entre sí, los cuales solo están comunicados gracias a las redes nerviosas. Los organismos pluricelulares construyeron y perfeccionaron estos mecanismos con el fin de coordinar mejor sus componentes. Esta mayor coordinación basada en el conocimiento de las propias acciones les permitió funcionar mejor como un todo, un todo que comienza a construir estructuras destinadas específicamente a autoconocerse mejor.

## 6.2 La psicología ecológica formulada por Gibson

Tanto Bermúdez (2018) como Meijnsing (2022) dan una gran importancia a las investigaciones de Gibson (1979) a la hora de dilucidar los sustratos primitivos de la autoconsciencia. Esto proviene del énfasis que Gibson puso en la propiocepción en sus ya clásicos estudios sobre la percepción. Gibson quería combatir la creencia, antes bastante extendida, de que la percepción sensorial es un proceso pasivo por el cual un organismo simplemente recolecta estímulos del medio, para después procesarlos y poder darles una respuesta. En contra de esta visión, Gibson concebía la percepción como un proceso activo. Esta visión le llevó a considerar la propiocepción como un elemento clave de la percepción: un animal no estaría simplemente tomando estímulos del medio de manera pasiva y arbitraria, sino buscando expresamente aquellos estímulos que posibilitasen sus acciones, teniendo en cuenta su propia fisonomía y estado interno. Esto le llevó a afirmar que “la información acerca de uno mismo acompaña la información sobre el ambiente, ambas son inseparables. Ego-recepción acompaña a la exo-recepción como la otra cara de una moneda. [...] Uno percibe el entorno y se co-percibe a uno mismo” (Gibson, 1979, p. 126).

El concepto de propiocepción de Gibson está estrechamente relacionado con la reaferencia explicada en la sección anterior. Gibson habla sobre propiocepción cutánea cuando un animal se toca a sí mismo en un proceso reafereente táctil, así como sobre procesos reafereentes auditivos, donde un animal escucha sus propias pisadas, su propia respiración, el sonido de sus mandíbulas mientras mastica... Sin embargo, el sistema visual es aquel al que Gibson dedica una mayor atención.

Gibson distingue principalmente tres tipos de propiocepción ligadas a la vista. La primera está relacionada con aquellas partes del propio organismo que forman parte de su campo visual y lo acotan. Un animal tiene unos ciertos límites en su campo visual conformados por su propia fisonomía: las cuencas de los ojos, los pómulos, la nariz y las pestañas componen unos límites u obstáculos que definen el campo visual de una criatura y le aportan información sobre sí misma. Algo parecido puede decirse de aquellas extremidades del propio organismo que pasan a formar parte de su campo visual, a los cuales Gibson llama ‘objetos subjetivos’.

Un segundo tipo de propiocepción resulta de la kinestesia visual, es decir, la capacidad por parte de un organismo de distinguir cuándo su campo visual registra cambios debidos a su propio movimiento o a movimientos ajenos. Gibson investigó cómo funciona la kinestesia visual y descubrió que si el entorno es percibido como una estructura invariante que *fluye* en una cierta dirección, entonces la causa de este cambio

se asocia al propio movimiento. Esto es así incluso sin aporte informacional por parte de los miembros del cuerpo que generan dicho movimiento, como demuestran resultados de un experimento llevado a cabo por Lishman y Lee (1973), donde personas quietas perdían el equilibrio cuando se proyectaba un flujo como el descrito sobre el suelo. Estos resultados sugieren que el mecanismo descrito por Gibson podría ser similar al de una copia eferente que ayuda a interpretar ciertos cambios en la percepción visual, asociándolos a una causa interna precisamente porque dichos cambios se ajustan a un patrón preestablecido con anterioridad.

Un tercer tipo de propiocepción visual proviene de un término que Gibson fue el primero en emplear y que ha cobrado mucha importancia en las últimas décadas en el marco teórico de la ciencia cognitiva: la *affordance*. Una *affordance* se refiere a una cierta *disposición* para actuar que un organismo detecta en el entorno. Un mismo objeto puede suscitar varias *affordances* distintas según el contexto: una taza tiene como *affordances* asociadas tanto el hecho de agarrarla por el asa para beber como el de lanzarla a la cabeza de alguien que nos quiere atacar. Las *affordances* detectadas por un ratón serán muy distintas a las de una serpiente, ya que sus posibilidades de acción dependerán en gran medida de sus fisonomías y capacidades físicas. Por esta razón, las *affordances* traen consigo una cierta propiocepción que informa al organismo de qué acciones puede realizar en un cierto entorno concreto, las cuales serán radicalmente distintas a las de una criatura de otra especie en el mismo medio.

Las propuestas de Gibson (1979) resultan —junto a los procesos reaférentes con las que están vinculadas— muy útiles para establecer una noción primitiva de primera persona en los organismos pluricelulares. A través del componente propioceptivo de la percepción puede construirse un punto de vista subjetivo de carácter primitivo, el cual podría haber sustentado los primeros indicios de autoconsciencia. Sin embargo, a pesar de que los procesos reaférentes y la propiocepción sean eventos donde un organismo emplea información sobre sí mismo como base para adaptarse mejor a un cierto entorno, aún estamos hablando de nociones muy básicas de autoconsciencia. Para poder establecer la existencia de grados de autoconsciencia más elevados, tendremos que cuestionarnos más a fondo cómo los organismos pluricelulares evalúan y deciden sus acciones futuras, y cuán capaces son de comprender el entorno que les rodea y el hecho de que forman parte de él.

## **7. La aparición de los modelos internos y la capacidad conceptual**

En cierto momento de la evolución, algunos animales dejaron de responder tan solo a estímulos concretos y comenzaron a reaccionar ante su entorno tratándolo como un

todo, atendiendo a la combinación resultante del conjunto de sus percepciones. Según la *teoría de la complejidad ambiental* (Godfrey-Smith, 1996), la cada vez más elevada complejidad de los entornos que habitaban los organismos —en parte debido a la coexistencia con otros organismos cada vez más sofisticados— generó una presión evolutiva que convirtió las respuestas demasiado mecánicas y predecibles en una muerte probable para una criatura. Esto impulsó una presión evolutiva hacia la aparición de mecanismos cognitivos que permitiesen una mayor flexibilidad comportamental. La manera en la que esto fue posible fue mediante la creación de modelos internos que simulasen el ambiente que les rodeaba en tiempo real. No se conoce con exactitud el proceso que dio lugar a los primeros modelos internos en organismos pluricelulares, aunque sí que podemos intentar establecer qué animales parecen poseer la capacidad de generar un modelo interno de la realidad.

Un modelo interno es una *representación* del mundo en el interior de un organismo. El uso de la palabra *representación* viene de que el mundo se le presenta al organismo tal y como es, y el organismo lo vuelve a presentar en su interior de manera simplificada. Hay dos razones por las que un modelo interno siempre conlleva una simplificación del entorno que envuelve a un animal. La primera es que un organismo dispone de un número limitado de órganos sensoriales, los cuales recaban información del medio de una *manera* específica, ligada a la morfología de sus órganos sensoriales. Por ejemplo, un murciélago forma un modelo interno de su alrededor mediante un proceso reaférente que consiste en la emisión de sonidos, los cuales rebotan en las superficies y le permiten acotar las distancias con los objetos de su entorno por ecolocalización. Esto último también puede ser conseguido mediante un sistema sensorial que se apoye principalmente en la recepción de fotones —como sucede en nuestra vista—, los cuales también rebotan en las superficies de los objetos y permiten conocer la distancia a la que se encuentran. Ambos sistemas tienen ventajas y desventajas: los murciélagos pueden orientarse en una situación de oscuridad total, pero un sistema visual permite formar un modelo interno más rico en matices, gracias a que en las superficies solo ‘rebotan’ fotones de una frecuencia particular, lo cual puede dar lugar a una categorización de esas frecuencias para diferenciar tipos de objetos, como sucede en nuestro sistema de visión a color. En ambos sistemas, sin embargo, existe una cierta resolución límite, relacionada con la cantidad de información por unidad de tiempo que son capaces de tomar del medio.

La segunda razón por la que un modelo interno siempre es una simplificación de la realidad viene de la capacidad de procesamiento de información de los organismos. Aunque un órgano sensorial fuese capaz de recopilar cantidades enormes de información sensorial del medio, esta debe poder ser procesada, analizada y empleada por el

organismo antes de formar parte de su modelo interno. Las capacidades cognitivas de un organismo no son ilimitadas, lo cual es una fuente de imperfecciones y supone limitaciones en la creación de su modelo interno.

Kant (1781/1981) ya entendió hace casi doscientos cincuenta años que es imposible para un organismo sensible representar en su modelo interno el objeto percibido ‘tal cual es’. La particularidad de sus órganos sensoriales, ligados a la manera concreta en la que un procesa la información recibida para generar un modelo interno, forma una brecha intransitable entre la ‘realidad en sí’ y la ‘realidad percibida’, entre noúmeno y fenómeno. Un modelo interno es una *simulación* de la realidad, que intenta aproximarse lo mejor posible, aunque de manera simplificada, a la realidad misma.

Sin embargo, un modelo interno no tiene por qué representar a la perfección el medio para ser útil. Poseer cualquier modelo interno ya posibilita una cierta independencia del organismo de su percepción momentánea del entorno. Permite, por ejemplo, saber que un depredador que se acaba de ocultar detrás de una roca y del que no se tiene visibilidad *sigue ahí* a pesar de que no se tenga ningún registro sensible de su presencia en ese instante. Permite también recordar el lugar donde se encuentra una fuente de agua que no es visible desde el lugar donde uno se encuentra, o trazar un camino de vuelta hacia la madriguera tras una expedición lejana en busca de alimento.

Los modelos internos pueden ser multinivel, lo que quiere decir que varios modelos internos distintos pueden coordinarse y ser empleados al mismo tiempo por un solo organismo. El caso más básico de este fenómeno se produce en animales que son capaces de generar mapas que les sirven para orientarse en espacios mucho más grandes que aquellos que abarcan sus órganos sensoriales —a la vez que modelan su entorno más inmediato a través de la información actualizada por sus sentidos—. La generación y uso de mapas requiere de memoria, ya que un mapa vacío no sirve para nada. Un mapa debe estar provisto de ítems que pertenecen a diversas categorías, como ‘río’, ‘madriguera’, ‘zona frondosa’, ‘lugar donde se escondió comida’, etcétera, los cuales son atribuidos cuando se detectan y almacenados en el mapa para poder acceder a recursos clave según se necesite.

Generar y operar un mapa interno conlleva una cierta capacidad conceptual, es decir, la capacidad para establecer y operar con conceptos. El término ‘concepto’, al igual que ‘consciencia’, es un término maleable al que diversos autores dotan de distintas definiciones. No es posible establecer aquí las principales líneas de un debate en torno a esta palabra, el cual lleva sosteniéndose desde la antigüedad. Sí dejaremos claro que el término ‘concepto’ será empleado aquí de manera muy radical y básica. Cualquier unidad dentro un proceso cognitivo desligado de la recepción inmediata de estímulos será

considerada un concepto, es decir, cualquier componente diferenciable dentro de un modelo interno que no depende exclusivamente de información perceptual exterior cumple los requisitos para ser considerado un concepto en grado mínimo. Un concepto, al ser diferenciable de otros, debe poder combinarse con otros para dar lugar a conceptos más complejos. También podrá ser agrupado junto a otros conceptos según distintas relaciones, lo que posibilita la formación de categorías y asociaciones. Debe añadirse que aquí se considerará que la capacidad de manejar conceptos está presente en gran cantidad de especies animales no-humanos y que la capacidad conceptual es una cuestión de grado, es decir, que la atribución de conceptos no es tanto un asunto de ‘todo o nada’ sino más bien de ‘más o menos’. Otros autores que sostienen esta visión son Camp (2009), Newen y Bartels (2007) y Carruthers (2009). Para una defensa más extensa de esta postura realizada por el autor de este ensayo véase Botana (2023).

Tras los sorprendentes descubrimientos de Gould (1986) sobre la capacidad de orientación de las abejas, Menzel et al. (2005) realizaron un experimento muy similar utilizando tecnología más avanzada. En concreto, un radar armónico les permitió realizar un seguimiento muy preciso del vuelo de las abejas estudiadas. Al igual que Gould, los investigadores se dedicaron a atrapar abejas que estuviesen yendo a por comida, que los investigadores colocaron en abundancia en unos sitios concretos a cierta distancia de la colmena. Las abejas fueron transportadas en una caja negra y liberadas en lugares aleatorios tras su captura. Menzel et al. observaron que, en un principio, las abejas continuaban volando en la misma dirección en la que se movían en el momento de ser atrapadas. Sin embargo, en cierto momento las abejas parecían darse cuenta de que estaban perdidas, pues iniciaban una estrategia de vuelo muy distinta que consistía en trazar círculos que parecían ayudar a las abejas a recolectar información del medio para orientarse. La orientación solía tener éxito y el vuelo en círculos normalmente era seguido por un vuelo en línea recta hacia el sitio donde estaba la comida o de vuelta hacia la colmena. El hecho de que fueran capaces de volar directamente hacia el punto de comida desde lugares aleatorios sugiere que las abejas poseen un mapa interno que les permite calcular un nuevo camino entre dos puntos cualesquiera siempre que sean capaces de autolocalizarse en el mapa mediante vuelos de orientación.

Respecto a si experimentos como el de Menzel et al. confirman o no la presencia de mapas internos en insectos existe un debate que lleva décadas activo (Dhein, 2023). En este ensayo se argumenta que la evidencia parece ser suficiente como para sostener de forma plausible —aunque no definitivamente— la afirmación de que los insectos poseen mapas internos. Como señala Carruthers (2009), los descubrimientos de Gould (1986) y Menzel et al. (2005) sugieren que las abejas operan mapas simbólicos donde distintos ítems estarían localizados, dentro de los cuales las abejas se autolocalizarían. Esto se ve

apoyado porque las abejas *distinguen* lugares. Vuelan a lugares donde hay diferentes tipos de comida —como polen, néctar o savia de árbol— y también van a por agua a distintos lugares para dar de beber a la colonia. Estudios indican que incluso parecen tener preferencias respecto al lugar donde recogen agua (Ostwald et al., 2016), lo cual sugiere que son capaces de distinguir diferentes tipos de fuente de agua, los cuales podrían estar marcados como ítems diferenciados dentro de su mapa interno. Esto indica que las abejas son capaces de distinguir diferentes categorías de localizaciones —la colmena, varios puntos clave que emplean para orientarse, diferentes fuentes de alimento, fuentes de agua— y de establecer localizaciones que, aunque están en la misma categoría, se consideran lugares individuales y diferenciados. La capacidad de categorización de las abejas se ha evidenciado mediante la realización de otros experimentos, donde las abejas han demostrado ser capaces de entender los conceptos de ‘igual’ o ‘diferente’ con estímulos nuevos bajo condiciones experimentales (Giurfa, 2003).

En todo caso, la autolocalización de un organismo en su propio modelo interno es fruto de una noción muy importante para la evolución de la autoconsciencia. Esta idea ha sido muy trabajada por Peacocke (1992), quien expone que la reidentificación de un lugar por un organismo supone una noción primitiva de autoconsciencia, ya que cualquier reidentificación de un lugar supone darse cuenta de que ‘yo ya he estado aquí’. La reidentificación de un lugar es una capacidad básica sin la cual no sería posible generar mapas internos. Un mapa interno conlleva, más aún, la capacidad de recordar la localización de ese lugar reidentificado en relación a otros lugares reidentificados.

Los experimentos de Gould (1986) y Menzel et al. (2005) revelan que las abejas que participaron en los experimentos eran capaces de relocalizarse dentro del mapa interno que manejaban para orientarse en un cierto espacio. Esto quiere decir que el mapa interno no era egocéntrico, sino aloecéntrico, es decir, que el organismo se incluía en el mapa interno como un elemento conceptual más cuya posición era actualizada constantemente en función de su propio movimiento y la reidentificación de nuevos lugares previamente guardados en el modelo interno. Esta capacidad, como puso de manifiesto Peacocke (1992), y más tarde también Bermúdez (2018), supone un paso adelante en la capacidad de autoconsciencia de los organismos pluricelulares, ya que conlleva la inclusión del propio organismo —como elemento diferenciado— dentro de su modelo interno. Podemos dotar por ello de cierto grado significativo de autoconsciencia, aunque aún primitivo, a cualquier animal que pueda crear mapas internos —y otro tipo de modelos internos—, lo cual, a la vista de los conocimientos actuales, parece incluir diversas especies de animales tanto de invertebrados —sobre todo artrópodos— y vertebrados (Gallistel, 1990). Podemos afirmar con bastante seguridad que la capacidad

para formar modelos internos ha surgido, además, de manera independiente en diversas especies de invertebrados, es decir, que diversas especies consiguieron esta capacidad a través de procesos evolutivos diferentes (Roth, 2015).

## **8. Pausas para tomar decisiones y comprensión de la propia existencia**

### **8.1 El surgimiento de simulaciones predictivas**

Al mismo tiempo que los modelos internos de los organismos pluricelulares se hacían cada vez más detallados y complejos, aparecieron unos mecanismos que emplearon esa capacidad de generar modelos de una manera distinta a la que se ha expuesto en la sección anterior. En vez de crear tan solo un modelo interno —o varios— que generase una representación del mundo en tiempo real, se empezaron a dar casos donde un modelo interno intentaba modelar situaciones futuras.

¿Cómo fue esto posible? Primero, mediante una mejora sustancial de la memoria de los organismos. El perfeccionamiento de la memoria permitió a algunos organismos mantener un registro mayor de situaciones modeladas con anterioridad. La revisión de situaciones parecidas a la actual se convirtió en un mecanismo clave que posibilitó la simulación de posibles eventos futuros a través del conocimiento acumulado durante la vida de la criatura. Si un organismo se encuentra ante una cierta situación problemática y, después de un análisis, consigue descubrir ciertos patrones en el desarrollo de aquellas situaciones pasadas que se parecen a dicha situación conflictiva, entonces puede esperar que se repitan de nuevo esos patrones y actuar en consecuencia. Esto posibilita reaccionar a eventos que aún no han sucedido mediante la predicción de qué sucederá probablemente, a raíz de la información que se ha obtenido del pasado. Estas predicciones, si son acertadas, pueden tener un gran impacto en la supervivencia de una criatura.

El origen de estas predicciones se puede situar en las *respuestas de orientación* (RO). Las RO consisten en la activación de procesos de evaluación y decisión ligados a la detección de un estímulo de tipo concreto, el cual ha sido seleccionado para activar este tipo de respuestas porque ha sido marcado como ‘indicador de peligro’. No se puede entender el origen de las RO sin comprender que, cuando hablamos de animales con una capacidad cognitiva suficiente como para generar un modelo interno de la realidad que les rodea, ya tratamos con animales con cierta inteligencia, los cuales poseen, al menos en cierto grado, plasticidad comportamental —es decir, pueden responder a un mismo estímulo con varias respuestas distintas—. Las RO explotan esta plasticidad



comportamental para activar una evaluación más extensiva de una situación que se considera importante y facilitar una recogida más precisa de datos con el fin de dar la mejor respuesta posible.

Las RO consisten en un proceso donde un animal, al detectar una situación que requiere ‘no equivocarse bajo ningún concepto’, activa un proceso de evaluación para sopesar cómo actuar. Bradley (2009) distingue tres fases en las RO, basándose en un estudio donde unos monos eran expuestos por primera vez a un peluche animal que era introducido en su misma jaula, el cual era inicialmente identificado como una posible amenaza. La primera fase consistía en una pausa prudencial donde el animal congelaba su movimiento y recopilaba toda la información posible del nuevo estímulo mediante sus órganos sensoriales, dirigiéndolos lo máximo posible hacia el nuevo objeto. La segunda consistía en una respuesta positiva y motriz: huida, adopción de una postura defensiva o preparación para el ataque. La tercera, exploratoria, se daba tras un intervalo de tiempo: cuando el peluche, naturalmente, no producía signos de movimiento, los monos realizaban un acercamiento al estímulo novedoso para estudiarlo con más detenimiento, tras haber descartado que resultaba peligroso. Las dos primeras fases pueden considerarse básicas y se darían en casi todos los seres vivos. La última estaría más caracterizada por la ‘avidez de conocimiento’ propia de animales con cerebros muy desarrollados (Dennett, 1991).

Neumann (1990) ilustra las respuestas de orientación con una metáfora. Pongamos que un organismo es un barco donde la tripulación está ocupada realizando tareas rutinarias propias del manejo de una embarcación. De pronto, se detecta algo importante y el vigía grita: “¡todo el mundo a cubierta!” Toda la tripulación abandona su puesto y se enfrenta el estímulo: ¿es un barco enemigo?, ¿se trata de una tormenta peligrosa?, ¿es esa mancha en el horizonte el destino buscado? Se analiza la situación y se debate qué hacer. Tras ello, todo el mundo vuelve a sus tareas iniciales o se prepara para un posible ataque, o una peligrosa tormenta, o para desembarcar... Esto conlleva ligar las respuestas de orientación a un proceso donde se centraliza la atención de la criatura en un proceso concreto de evaluación.

Un ejemplo de respuesta de orientación presente en casi todos los animales, incluidos los humanos, es la detección de una simetría de eje vertical en el campo visual. Gran parte de los seres vivos presentan este tipo de simetría cuando son vistos de frente, por lo que su detección significaría, en la gran mayoría de los casos, la presencia de un animal situado de frente hacia nosotros, es decir, un animal que nos está mirando.

Ante esta situación, casi todos los animales activan una RO para decidir cómo afrontar la situación (Dennett, 1991).

Las simulaciones que tratan de predecir el desarrollo futuro de una situación concreta a partir de la memoria de situaciones anteriores similares ya han sido reconocidas por varios filósofos como una función muy importante de animales con cierta inteligencia. Dennett (1991, p. 177) llega a afirmar que “el propósito fundamental de los cerebros es producir futuro”. Aquí defenderé que la capacidad de producir futuro está estrechamente ligada con la capacidad de un organismo de generar modelos internos. Una vez que un organismo es capaz de generar un modelo interno, puede también generar modelos que no solo acoten el momento presente, sino que jueguen en un plano temporal distinto al de la realidad. Esto posibilita una independencia del momento actual: de pronto se pueden introducir cambios en los modelos que no están inducidos por información sensorial del entorno inmediato, sino por la introducción de patrones de acción extraídos del registro de modelos internos recopilados en la memoria del organismo.

Las respuestas de orientación serían así una ‘pausa para tomar decisiones’, consistente en la activación de un proceso evaluador capaz de analizar la situación y clasificarla en función de experiencias pasadas, que después da pie a otro mecanismo que activa una decisión, ya sea de vuelta a la normalidad —tras una ‘falsa alarma’— o de actuación correspondiente a los resultados del análisis. Aunque los primeros organismos solo experimentasen esa clase de pausas cuando hubiesen detectado un estímulo concreto —como la detección de una simetría de eje vertical— dando lugar a una RO, pronto otras especies empezaron a aplicar las pausas en muchas más ocasiones, en todas aquellas en las que surgiese un problema que no fuese sencillo de resolver.

Pongamos un ejemplo de cómo funcionan estas pausas basadas en simulaciones para producir futuro. Una ardilla está saltando de una rama a otra en un jardín hasta que aterriza en un alféizar. Desde esa pequeña plataforma, la ardilla reconoce pronto que no puede volver a alcanzar la rama por la que ha llegado hasta allí —la red neuronal de su cabeza es experta en medir distancias de salto de acuerdo a sus capacidades—. Un problema ha surgido, lo que crea una ‘pausa para tomar decisiones’, un proceso de evaluación de su complejo cerebro especializado en dar solución a situaciones complicadas. Ese proceso de evaluación tomará en consideración la distancia con el suelo, buscará *affordances* —o combinaciones de ellas (Hunt et al., 2021)— en el entorno que posibiliten una salida. Recopilará toda la información posible para considerar las opciones disponibles y decidir entre ellas la más beneficiosa o, por lo menos, la menos

dañina. En este caso las opciones son muy limitadas, la pared del edificio es lisa y no presenta *affordance* alguna y la rama de origen está fuera del alcance de salto. Finalmente, la ardilla acepta lo inevitable. Para no morir de hambre en el alféizar, el proceso de evaluación elige como única solución viable un salto al suelo desde la ventana, a pesar de que la altura sea suficiente como para ocasionar daño. Antes de tomar esta decisión, la ardilla ha *simulado* en su interior, mediante una atención dirigida por parte del proceso de evaluación, tanto el salto de vuelta, como la caída, como las posibles *affordances* que finalmente fueron descartadas. Esto fue posible mediante el uso de modelos internos focalizados en la generación de situaciones ficticias que se desarrollaron como simulaciones en el interior del organismo durante la ‘pausa para tomar decisiones’.

Este último ejemplo proviene de una observación personal del comportamiento de una ardilla. Pude ver cómo se paralizaba ante la situación problemática, cómo observaba su entorno con suma atención para intentar recabar toda la información útil posible para realizar lo más acertadamente posible sus simulaciones internas, antes de, finalmente, saltar al suelo tras varios minutos de reflexión situacional. Es curioso que la observación directa de una ardilla también indujese a Millikan (2006) a conclusiones muy similares. A partir de la observación de una ardilla que trataba de alcanzar mediante un salto el comedero destinado a pájaros que esta filósofa había colocado en su jardín, Millikan (2006, p. 70) se sorprendió de descubrir a la ardilla “estudiando la situación [...] primero desde un lado, luego desde el otro, después desde debajo [...], analizando la situación desde varios ángulos” antes de probar un salto desde el tejado de la casa para, al segundo intento, conseguir alcanzar el comedero. Esta observación llevó a Millikan a concluir que algunos animales poseen la capacidad de realizar un proceso de prueba y error a partir de representaciones internas que les permitiría experimentar con distintas posibilidades de acción sin involucrarse de manera real. Millikan concluye que esto requiere una cierta racionalidad ligada al nivel perceptual, que funcionaría a partir de situaciones concretas y el análisis de *affordances*.

Los ‘procesos de prueba y error’ que describe Millikan son los mismos a los que me refiero cuando hablo aquí de simulaciones a partir de modelos internos, en un intento de describir más concretamente el proceso cognitivo que posibilita estas capacidades en ardillas y otros animales. ¿En qué animales? Es difícil responder a esta pregunta, porque existen diversas especies que parecen tener la capacidad de generar evaluaciones como las descritas en distinto grado de complejidad. En principio, me aventuro a decir, tentativamente, que todo animal que genera modelos internos es también capaz de realizar ‘pausas para tomar decisiones’, ya sean empleadas con frecuencia o tan solo

como producto de una RO. En consecuencia, también los insectos tendrían, en un grado muy bajo, la capacidad de realizar ‘pausas para tomar decisiones’, aunque estas pausas serían mucho más complejas y frecuentes en animales vertebrados, sobre todo en mamíferos y algunas especies de aves, como los córvidos.

## **8.2 Perfeccionamiento del proceso de evaluación mediante la metacognición**

El proceso de evaluación descrito en la sección anterior se compone de multitud de subprocesos que le permiten realizar un análisis multivariable de una situación. Los modelos internos funcionan combinando distintos tipos de sistemas, los cuales producen información valiosa de carácter diferente, la cual es integrada por el proceso de evaluación en conjunto para tomar una decisión. Existen evidencias de que la mente de los animales y los humanos funciona de manera modular, donde los módulos analizan la información de cada tipo de órgano sensorial a partir de muchos pequeños procesos especializados en una función concreta (Ogas y Gaddam, 2022). Dentro de cada módulo existirían múltiples submódulos encargados de procesar la información en cuanto a aspectos muy concretos. En el módulo visual, por ejemplo, Grossberg (2000) distingue dos procesos cognitivos diferenciados que, combinados, permitirían a un animal distinguir *formas* en su campo visual —y, con ello, reconocer objetos y otros animales, así como estructuras invariantes que posibilitan las *affordances*—. Una de las principales evidencias de la existencia de multitud de procesos cognitivos diferentes localizados en distintas partes del cerebro en humanos son aquellas condiciones mentales producidas por una lesión cerebral localizada, donde solo una parte del cerebro ha resultado dañada, que tienen como consecuencia la pérdida de facultades muy concretas mientras se mantienen intactas otras. Ejemplos de esto son la prosopagnosia —incapacidad para reconocer rostros—, la akinetopsia —imposibilidad de percibir el movimiento con la vista—, la alexia —incapacidad para reconocer palabras escritas en personas que antes podían leer— y la agnosia —incapacidad para diferenciar y reconocer objetos—.

En esta sección queremos hablar sobre la posesión por parte de animales de diversas facultades ligadas a la metacognición. Para ello, es importante tener en cuenta esta visión modular de la mente, ya que, siguiendo a Proust (2013, p. 4), la metacognición se puede entender como un “conjunto de capacidades por las que un subsistema cognitivo es evaluado o representado por otro subsistema de una manera sensible al contexto”. Proust sostiene, desde una corriente que ha denominado *evaluativismo*, que la metacognición ha surgido como una mejora sustancial del proceso de evaluación que describíamos en la sección anterior. El proceso de evaluación de un organismo es siempre falible, por lo

que habrían aparecido subsistemas para reducir la incertidumbre a la hora de establecer decisiones. La teoría de Proust —que tiene muchas similitudes con la defendida en este ensayo— se opone a aquellas teorías que adscriben una gran importancia a la *teoría de la mente* para explicar el origen de la metacognición. Estas teorías sostienen que la metacognición apareció en organismos muy sociales, después de que surgiesen mecanismos cognitivos que ayudaban a ‘leer la mente’ (*mindreading*) de otros organismos en contextos sociales. La metacognición habría surgido cuando estos mecanismos fueron empleados posteriormente por un organismo para ‘leerse la mente’ a sí mismo. Existe evidencia, sin embargo, de que animales sin capacidad de ‘leer la mente’ poseen mecanismos metacognitivos, como se mostrará enseguida.

Proust (2013) distingue tres niveles de metacognición en animales y defiende que los animales no-humanos serían capaces, al menos, de realizar procesos metacognitivos que se encajarían en los dos primeros niveles. El nivel 1 se refiere a aquellos procesos metacognitivos que regulan la percepción y la actividad motora. Nos hemos referido aquí a ellos extensamente en la sección dedicada a la reaferencia, la descarga corolaria y la copia eferente. El nivel 2 se refiere a aquellos procesos que regularían las simulaciones predictivas que hemos descrito en la sección anterior y estaría conformado por procesos metacognitivos que operan en el corto plazo. El nivel 3 se refiere a aquellos procesos metacognitivos donde se simulan y regulan metas del organismo a largo plazo. En este nivel cobra una gran importancia el hecho de que el organismo pueda comprender tanto que cada acción mental que ha ido realizando influye en el largo plazo, como que ha actuado en varias ocasiones con el fin de modificar el curso de sus acciones mentales durante su vida. Proust añade que en los seres humanos estos tres niveles operan de manera semi-jerárquica, porque existiría una jerarquía entre los niveles, pero a la vez cierta independencia entre ellos —el nivel 1 podría operar sin órdenes del nivel 2 o bajo su mando, por ejemplo—.

Con fines ilustrativos, antes de continuar nos detendremos en dos ejemplos de procesos metacognitivos que se encajarían en el nivel 2. El primer ejemplo es la estimación que sería capaz de realizar una persona del grado de certidumbre que tiene de haber reconocido bien un pájaro cuando ha pasado volando rápidamente a su lado —“creo que era un mirlo, estoy casi seguro, al 90%”, frente a “no estoy nada seguro de si ese pájaro era una cacatúa, no he conseguido verlo bien”—. El segundo ejemplo es la sensación de tener un nombre ‘en la punta de la lengua’ mientras se está realizando un proceso de búsqueda inconsciente, lo cual nos informa de que, con un poco más de tiempo y esfuerzo, es posible que se consiga dar con el nombre buscado. Más adelante hablaremos

de cómo la metacognición y los procesos de evaluación se relacionan con la capacidad de sentir ciertas sensaciones.

¿Existe de veras evidencia de que los animales son capaces de llevar a cabo procesos metacognitivos? Sí la hay. A pesar de que hay una cierta dificultad para diseñar experimentos que sean capaces de comprobar la presencia de procesos metacognitivos, décadas de investigación han dado lugar al trazo de diseños metodológicos fiables. Hampton (2009) recogió unas bases metodológicas robustas como respuesta a las críticas constructivas realizadas a los primeros experimentos realizados en el campo de la metacognición. Estas bases asegurarían la presencia de metacognición en un experimento que las cumpliera y tuviera un resultado positivo. Estas bases no se revisarán aquí, pero sí que se citarán algunas investigaciones que las cumplen y cuyo resultado indicó la presencia de metacognición en primates.

Couchman et al. (2010) realizaron un experimento para evaluar y comparar la metacognición en humanos y en macacos Rhesus. Tanto los humanos como los monos respondieron al mismo tipo de pruebas para ello: clasificación de densidades de cuadros rellenos de puntos en 'alta' o 'baja', clasificación de círculos en 'continuo' o 'discontinuo' según el número de puntos que conformaban su perímetro, clasificación en si una elipse era 'plana' o 'redonda' según la proporción entre sus ejes, y otras parecidas. En las pruebas había siempre dos botones de 'respuesta' y un tercer botón '?' destinado a expresar incertidumbre. En humanos, bastó una explicación para que los participantes pudiesen emplear este botón. Los monos recibían una chuche con sabor a fruta si acertaban y recibían una penalización de 20 segundos sin poder realizar más pruebas en caso de fallo. El botón '?' les permitía saltar a la siguiente prueba sin penalización, pero no tenía ninguna recompensa asociada. Los macacos Rhesus demostraron emplear el botón de incertidumbre de manera muy parecida a los humanos, demostrando ser capaces de averiguar mediante procesos metacognitivos las preguntas donde era más difícil responder. Un experimento parecido fue realizado con éxito con orangutanes (Suda-King, 2008). Sin embargo, algunos primates, como los capuchinos gráciles, no han demostrado ser capaces de monitorear su propio desempeño en pruebas similares (Beran et al., 2009).

Kornell et al. (2007) realizaron un diseño experimental donde macacos Rhesus, tras elegir su respuesta en una prueba, podían elegir entre dos opciones: una les daba tres tokens de recompensa en caso de que acertaran y les restaba tres tokens si fallaban; la otra opción les daba un token de recompensa solamente, pero se lo daba acertasen o fallasen. Los monos eligieron consistentemente más la apuesta arriesgada cuando

acertaban que cuando fallaban, lo cual indica una evaluación metacognitiva de su propia certidumbre a la hora de responder. Es notable que los macacos Rhesus han fallado en los experimentos relacionados en los experimentos donde se comprobó su capacidad de 'leer la mente' a otros macacos. Esto parece indicar que carecen de una teoría de la mente y, sin embargo, poseen procesos metacognitivos, lo que apunta a la anterioridad de la aparición de la metacognición en comparación a la teoría de la mente, al menos en algunas especies animales (Proust, 2013).

Los procesos metacognitivos permiten al propio proceso de evaluación realizar un seguimiento de su propia actividad con el fin de estimar la calidad de las respuestas que puede generar. La información metacognitiva informa al proceso acerca de su propio progreso y su probabilidad de éxito, así como sobre la fiabilidad de las estimaciones que se han empleado; todo con el fin de guiar al proceso de evaluación hacia una toma de decisiones lo más acertada posible. Aquí argumentaré a favor de que, en cierto punto, el proceso de evaluación empezó a incluir en el propio proceso de evaluación algunos mecanismos cognitivos destinados a un control del propio proceso. Esto permitiría a un organismo que está evaluando interrumpir una evaluación si esta está consumiendo demasiado tiempo o se concluye que seguramente no produzca los resultados esperados, es decir, permite hacer uso de una especie de *cortafuegos* donde el organismo puede interrumpir, a raíz de información sobre el proceso de evaluación, la pausa para tomar decisiones que lo contiene. Esto supone que el propio proceso de evaluación es capaz de incluir dentro del proceso de evaluación el hecho de que *está procesando*, ya que cualquier decisión tomada por el organismo formará parte de un proceso de evaluación. Una vez que esta metacognición del proceso de evaluación, que en un principio solo se introdujo para poder interrumpirlo y estimar la validez de sus análisis, quedó integrada dentro del proceso, se dieron lugar vías por las que información sobre el propio proceso en su conjunto estuvieron disponibles al propio proceso de evaluación, lo cual permitió una autoconsciencia más explícita que en grados anteriores. Un organismo que durante sus pausas para tomar decisiones es capaz de conocer, de manera intuitiva, que está tomando una decisión, es capaz de saber que es en un grado mínimo.

Creo que esta capacidad dota de suficiente autoconsciencia a un organismo como para reconocerse en un espejo. Esto parece ser corroborado por evidencia experimental. Gallup (1970) diseñó un experimento destinado a averiguar si un animal es capaz de autoreconocerse en una superficie reflectante. Consiste en narcotizar a un animal y colocar una marca colorida en un punto del organismo que solo pueda ver a través de su imagen reflejada. Si un animal se inspecciona la zona coloreada a partir de la observación de su reflejo, se considera que es capaz de autoreconocerse en él. Gallup demostró que

algunos chimpancés son capaces de superar la ‘prueba del espejo’. Más adelante, esta prueba ha sido superada por varias especies de mamíferos — orangutanes (Suárez y Gallup, 1981), orcas (Delfour y Marten, 2001), elefantes asiáticos (Plotnik et al. 2006), delfines (Reiss y Marino, 2001), entre otros— y aves —como las urracas (Prior et al. 2008)—. Existe un debate que aún sigue activo acerca de qué realmente nos dice que un animal sea capaz de pasar la prueba diseñada por Gallup. La prueba de Gallup podría ser interpretada como un fruto directo de que una pausa para tomar decisiones es capaz de incluir el hecho de está llevando a cabo un proceso de evaluación *durante* el proceso de evaluación del reflejo, dando lugar a que se dé la posibilidad de que un animal reconozca que el reflejo, al estar coordinado con sus propias acciones de manera síncrona, sea una proyección de su propio cuerpo sobre la superficie reflejante. Esto parece reforzado por la clara autoinspección a la que proceden algunos de los primates con más inteligencia delante de los espejos, como algunos de los chimpancés y orangutanes que han pasado la prueba.

La autoconsciencia producida por la inclusión de información metacognitiva en el proceso de evaluación permitiría, a algunos mamíferos desarrollados y a especies de aves con gran inteligencia, recibir pequeños ‘destellos’ de autoconsciencia en el pleno sentido de la palabra. Estos destellos, de corta duración, se darían debido a la información de carácter metacognitivo que las pausas para tomar decisiones poseen acerca de sí mismas.

## **9. Breve apunte sobre el rol de las sensaciones fenoménicas en el desarrollo de la autoconsciencia**

Muchos filósofos que han reflexionado sobre las bases primitivas de la autoconsciencia han dado una gran importancia a las experiencias que llevan a sentir un cuerpo como propio en cuanto sustratos de autoconsciencia (Bermúdez, 2018; Meijsing, 2022). Un lobo, aunque sea carnívoro y tenga hambre, no se comerá su propia pata, porque es un ‘objeto subjetivo’ que siente como propio (Gibson, 1979). De manera especulativa, quiero argumentar a favor de que las sensaciones que llevan a sentir un cuerpo como propio están estrechamente relacionadas con el proceso de evaluación expuesto en las secciones anteriores.

Como bien expresa Carruthers (2019), cualquier teoría de la consciencia que se precie debe tener en cuenta que la consciencia no es generalizada y total, sino parcial; es decir, debe mostrar cómo y por qué una gran cantidad de los procesos cognitivos que realizamos son inconscientes y solo algunos se dan de manera consciente. No tengo ninguna consciencia funcional ni fenoménica de cómo transformo mis pensamientos en palabras, simplemente pienso en contenidos y, después, procesos cognitivos de los que



no tengo ninguna información consciente traspasan esos contenidos a veinticuatro símbolos escritos en un orden lineal concreto. No tengo consciencia de qué músculos y tendones activo cuando quiero coger un vaso de agua. Solo unos pocos procesos son conscientes y son aquellos en los que se involucra una ‘pausa para tomar decisiones’, un proceso de evaluación que integra multitud de procesos cognitivos para, después, tomar una decisión que afecta el futuro del organismo al completo.

¿Cómo es posible que el proceso de evaluación sea capaz de integrar la información que proveen multitud de mecanismos cognitivos distintos? Tomemos el proceso inconsciente que transforma un contenido que quiero expresar a un amigo a una frase hablada. Normalmente el proceso funciona perfectamente y no necesito saber nada de él, simplemente deseo expresar un contenido y la palabra aparece, tras lo cual un proceso también inconsciente transforma la palabra en movimientos de la boca y la lengua para comunicarlo con el habla. Pero, a veces, la palabra a la que se vincula mi contenido no es encontrada con suficiente rapidez, lo que activa una sensación de ‘tener la palabra en la punta de la lengua’ que me informa a mí —el proceso de evaluación— de que se está realizando una búsqueda inconsciente. Los múltiples procesos cognitivos se centralizan así en el proceso de evaluación mediante la traducción de su información en sensaciones, los cuales son indicadores naturales de la información que expresan (Proust, 2013). Las sensaciones transmiten información muy concreta, como ‘hambre’, ‘ganas de hacer pis’, ‘cansancio’... De esta manera se puede informar al proceso de evaluación de multitud de contenidos de carácter diferente de manera uniforme. El cómo se crean las sensaciones no queda nada claro. Barbieri (2011) llama a las sensaciones ‘artefactos cerebrales’ para aludir al hecho de que las sensaciones son construidas por el cerebro precisamente para cumplir con una función. Esta función podría ser la de convertir a un ‘idioma’ comprensible por el proceso de evaluación información de carácter muy diverso, es decir, homogeneizar y traducir a una sensación la información proveniente de diversos mecanismos y sensores con el fin de aligerar el proceso de centralización de información necesario para que se produzca un proceso de evaluación. Las sensaciones serían así artefactos temporales —en contraposición a artefactos espaciales como las proteínas— precisamente creados para informar a un proceso que se extiende en el tiempo. Esto conduce a hipotetizar que las sensaciones podrían hacerse conscientes al entrar en contacto con una pausa para tomar decisiones.

Aunque la consciencia funcional y la consciencia fenoménica confluyan en el proceso centralizado de evaluación, parece que sí son fruto de procesos distintos —la consciencia funcional sería fruto de la propia evaluación y la fenoménica de la información mediante sensaciones que le llega al proceso de evaluación por parte de otros mecanismos—. Este

indicio proviene del estudio de la esquizofrenia, donde los pacientes investigados presentan una disociación entre la agencia que impulsa los propios actos —consciencia funcional— y la sensación de experimentar esos mismos actos como propios —consciencia fenoménica— (Proust, 2013). Una persona esquizofrénica puede saber que las voces que escucha en su cabeza se producen en *su* cabeza, pero pensar que están causados por una causa exterior a él. Las disrupciones entre distintas fuentes de autoconsciencia que se producen en pacientes esquizofrénicos —por ejemplo, entre agencia y propiedad— sugiere que la autoconsciencia no es fruto de un mecanismo único, sino el resultado de muchos procesos cognitivos diferentes, los cuales aportan información relevante para la construcción de la autoconsciencia de carácter diverso. Esto no sería solo característico de los seres humanos, sino de todos los animales a los que, en este ensayo, se les atribuye algún grado de autoconsciencia.

## **10. Conclusión**

En este ensayo se ha tratado de mostrar cómo la autoconsciencia parece haber emergido gradualmente a medida que aparecieron ciertos procesos cognitivos durante la evolución de los organismos pluricelulares. La intención principal ha sido no solo argumentar a favor de una visión concreta de la consciencia y su proceso evolutivo, sino incorporar dentro de esa visión una amplia selección de fuentes filosóficas y científicas diversas que tratan la autoconsciencia y su origen. Se ha intentado, así, formar un hilo argumentativo que reuniese la bibliografía en torno a un relato coherente que pudiese enfrentar el reto de describir la emergencia de la consciencia a lo largo de la evolución, indicando qué procesos cognitivos clave parecen haber marcado especialmente el proceso y en qué clases animales habrían aparecido. Se ha excluido de este ensayo en su mayor parte al ser humano, siendo el foco principal de la exposición la evolución que se ha producido antes de su aparición. Se ha evitado profundizar en el caso humano para dar más peso a la evolución de mecanismos cognitivos primitivos que podrían posibilitar una resolución de la paradoja de la autoconsciencia formulada por Bermúdez, y no tanto al proceso cultural que ha dado fruto, en seres humanos, a una autoconsciencia de grado mucho más alto que la observada en animales no-humanos, potenciada por su capacidad lingüística.

## **Agradecimientos**

Me gustaría dar las gracias al Dr. Fernando Martínez Manrique por su guía a la hora de convertir una idea en una investigación académica.

También quiero agradecer a Javier Molina Caro las fructíferas conversaciones del último año. En cierto punto he empleado nociones que no tendría igual de claras sin su ayuda, como el surgimiento de nuevas ‘unidades’ a partir de la unión sistemática de ‘subunidades’.

## **Bibliografía**

- Barbieri, M. (2011). Origin and evolution of the brain. *Biosemiotics*, 4(3), 369-399.
- Bell, C. (1981). An efference copy which is modified by reafferent input. *Science* 214, 450–453.
- Bell, C.; Han, V.; Sugawara, Y.; Grant, K. (1997). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature* 387, 278–281.
- Beran, M., Smith, J., Coutinho, M., Couchman, J.; Boomer, J. (2009). The psychological organization of “uncertainty” responses and “middle” responses: A dissociation in capuchin monkeys. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(3), 371.
- Bermúdez, J. (2000). *The paradox of self-consciousness*. MIT Press.
- Bermúdez, J. L. (2018). *The bodily self: Selected essays on self-consciousness*. MIT Press.
- Botana Meyerderks, S. (2023). *Propedéutica para una evolución de la conceptualidad*. Trabajo Final de la asignatura Filosofía de la Biología, impartida por Ana Cuevas Badallo en el Máster Interuniversitario de Lógica y Filosofía de la Ciencia.
- Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, 46(1), 1-11.
- Brooks, J.; Carriot, J.; Cullen, K. (2015). Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nature neuroscience*, 18(9), 1310-1317.
- Camp, E. (2009). Putting thoughts to work: concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, 78(2), 275-311.

- Carruthers, P. (2009). Invertebrate concepts confront the generality constraint (and win). En R. W. Lurz (Ed.) *The philosophy of animal minds*, (pp. 89-107). Cambridge.
- Carruthers, P. (2019). *Human and animal minds: The consciousness questions laid to rest*. Oxford University Press.
- Chalmers, D. (2003). Consciousness and its Place in Nature. En S. Stich & F. Warfield (Eds.), *Blackwell Guide to Philosophy of Mind* (pp. 102-142). Blackwell.
- Couchman, J.; Coutinho, M.; Beran, M.; Smith, J. (2010). Beyond stimulus cues and reinforcement signals: a new approach to animal metacognition. *Journal of Comparative psychology*, 124(4), 356.
- Davis, W.; Mpitsos, G.; Michael Pinneo, J.; Ram, J. L. (1977). Modification of the behavioral hierarchy of Pleurobranchaea. *Journal of comparative physiology*, 117, 99-125.
- Delfour, F.; Marten, K. (2001). Mirror image processing in three marine mammal species: killer whales (*Orcinus orca*), false killer whales (*Pseudorca crassidens*) and California sea lions (*Zalophus californianus*). *Behavioural processes*, 53(3), 181-190.
- Dennett, D. C. (1991). *Consciousness explained*. Back Bay Books.
- Fukutomi, M.; Carlson, B. (2020). A history of corollary discharge: contributions of mormyrid weakly electric fish. *Frontiers in integrative neuroscience*, 14: 42.
- Gallistel, C. R. (1990). *The organization of learning*. The MIT Press.
- Gallup, G. (1970). Chimpanzees: self-recognition. *Science*, 167(3914), 86-87.
- Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin Company.
- Giurfa, M. (2003). Cognitive neuroethology: dissecting non-elemental learning in a honeybee brain. *Current Opinion in Neurobiology*, 13, 726–735.
- Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press.
- Goff, P.; Seager, W.; Allen-Hermanson, S. (2017). Panpsychism. En E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*.
- Gould, J. (1986). The locale map of bees: do insects have cognitive maps? *Science*, 232, 861–863.

- Gould, S.; Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the royal society of London. Series B. Biological Sciences*, 205(1161), 581-598.
- Grossberg, S. (2000). Filling-In the Forms: Surface and Boundary Interactions. En L. Pessoa and P. DeWeerd (Eds.) *Filling-in: From Perceptual Completion to Skill Learning*. Oxford University Press.
- Hampton, R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms?. *Comparative Cognition and Behaviour Reviews*, 4, 17–28.
- Holst, E.; Mittelstädt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften* 37, 464–476.
- Hunt, N.; Jinn, J., Jacobs, L.; & Full, R. (2021). Acrobatic squirrels learn to leap and land on tree branches without falling. *Science*, 373(6555), 697-700.
- Huxley, T. (1874). On the Hypothesis that Animals are Automata, and its History. *The Fortnightly Review*, 16, 555–580. Obtenido de Sophia Project Philosophy Archives.
- James, W. (1879). Are we automata?. *Mind*, 1-22.
- James, W. (1911). Novelty and Causation: The Perceptual View. En *Some Problems of Philosophy*. Longmans.
- Jékely, G., Godfrey-Smith, P.; Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philosophical Transactions of the Royal Society B*, 376: 20190764.
- Jékely, G.; Keijzer, F.; Godfrey-Smith, P. (2015). An option space for early neural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684), 20150181.
- Kant (1781/1981) *Kritik der reinen Vernunft*. Insel Verlag Wiesbaden.
- Kornell, N. ; Son, L.; Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, 18(1), 64-71.
- Kovac, M.; Davis, W. (1980). Neural mechanism underlying behavioral choice in Pleurobranchaea. *Journal of Neurophysiology*, 43, 469–487.
- Leys, S.; Meech, R. (2006). Physiology of coordination in sponges. *Canadian journal of zoology*, 84(2), 288-306.

- Lindahl, B. (1997). Consciousness and biological evolution. *Journal of theoretical biology*, 187(4), 613-629.
- Lishman, J. R.; Lee, D. N. (1973). The autonomy of visual kinaesthesia. *Perception*, 2(3), 287-294.
- Ludeman, D.; Farrar, N.; Riesgo, A.; Paps, J.; Leys, S. (2014). Evolutionary origins of sensation in metazoans: functional evidence for a new sensory organ in sponges. *BMC evolutionary biology*, 14(1), 1-11.
- Mallatt, J.; Blatt, M.; Draguhn, A.; Robinson, D.; Taiz, L. (2021). Debunking a myth: plant consciousness. *Protoplasma*, 258(3), 459-476.
- Margulis, L. (2001). The conscious cell. *Annals of the New York Academy of Sciences*, 929(1), 55-70.
- Meijnsing, M. (2022). *A Philosophy of Person and Identity: Where was I when I Wasn't There?*. Springer.
- Menzel, R.; Greggers, U.; Smith, A.; Berger, S.; Brandt, R.; Brunke, S.; Bundrock, G.; Hülse, S.; Plümpe, T.; Schaupp, F.; Schüttler, E.; Stach, S.; Stindt, J.; Stollhoff, N.; Watzl, S. (2005). Honeybees navigate according to a maplike spatial memory. *Proceedings of the National Academy of Sciences USA*, 102, 3040–3045.
- Millikan, R. (2006). Styles of rationality. En S. Hurley y M. Nudds (Eds.), *Rational animals?* (pp. 115-132). Oxford University Press.
- Musser, J.; Schippers, K.; Nickel, M.; Mizzon, G.; Kohn, A.; Pape, C.; Arendt, D. (2021). Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science*, 374(6568), 717-723.
- Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, 83(4), 435-450.
- Neumann, O. (1990). *Some aspects of phenomenal consciousness and their possible functional correlates* [Artículo presentado en conferencia]. The phenomenal mind – how is it possible and why is it necessary?. Center for Interdisciplinary Research (ZiF), Bielefeld, Alemania.
- Newen, A.; Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology*, 20(3), 283-308.
- Ogas, O.; Gaddam, S. (2022) *Journey of the Mind: How Consciousness emerged from Chaos*. W. W. Norton & Company.

- Ostwald, M.; Smith, M.; Seeley, T. (2016). The behavioral regulation of thirst, water collection and water storage in honey bee colonies. *Journal of Experimental Biology*, 219(14), 2156-2165.
- Pantin C. (1956). The origin of the nervous system. *History and Philosophy of the Life Sciences*, 28, 171–181.
- Parker, G. H. (1919). *The elementary nervous system*. Lippincott.
- Peacocke, C. (1992). *A Study of Concepts*. MIT Press.
- Plotnik, J.; De Waal, F.; Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, 103(45), 17053-17057.
- Popper, K. (1978). Natural selection and the emergence of mind. *Dialectica*, 339-355.
- Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): evidence of self-recognition. *PLoS biology*, 6(8), e202.
- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. OUP Oxford.
- Reber, A.; Baluška, F. (2021). Cognition in some surprising places. *Biochemical and Biophysical Research Communications*, 564, 150-157.
- Reiss, D.; Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*, 98(10), 5937-5942.
- Robinson, W. (2007). Evolution and epiphenomenalism. *Journal of Consciousness Studies*, 14(11), 27-42.
- Roth, G. (2015). Convergent evolution of complex brains and high intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684), 20150049.
- Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13, pp. 3–31.
- Suárez, S.; Gallup, G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of human evolution*, 10(2), 175-188.
- Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember?. *Animal Cognition*, 11, 21–42.
- Trewavas, A. (2017) The foundations of plant intelligence. *Interf Focus* 7: 20160098.

Tweedy, L.; Thomason, P.; Paschke, P.; Martin, K.; Machesky, L.; Zagnoni, M.; Insall, R. (2020). Seeing around corners: Cells solve mazes and respond at a distance using attractant breakdown. *Science*, 369(6507).

Vimal, R. (2009). Meanings attributed to the term 'consciousness': an overview. *Journal of Consciousness Studies*, 16(5), 9-27.