



Viendo lo que es visto: la atribución de autoconciencia a los animales.

Seeing what is seen: the attribution of self-consciousness to animals.

DOI: [10.32870/sincronia.axxvi.n82.38b22](https://doi.org/10.32870/sincronia.axxvi.n82.38b22)

Alejandro Villamor Iglesias
IES Rosalía de Castro (ESPAÑA)

CE: alejandrovillamoriglesias@yahoo.es / ID ORCID: [0000-0001-7265-7528](https://orcid.org/0000-0001-7265-7528)

Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

Recibido: 27/12/2021

Revisado: 13/04/2022

Aprobado: 11/05/2022

Resumen:

El objeto del artículo consiste en la problematicidad existente en la atribución de autoconciencia a animales no humanos. En particular, se analizan tres vías distintas: la atribución en base al lenguaje, a ciertos experimentos que no requieren habilidades lingüísticas y a la mera posesión de consciencia fenoménica. Si bien todas ellas aportan sólidas razones para la atribución de autoconciencia, también conducen a serios problemas. Este trabajo sostiene la tesis de que no existe ningún criterio infalible para la atribución de autoconciencia y, asimismo, que la consideración de que la autoconciencia es un rasgo humano definitorio es cuestionable.

Palabras clave: Autoconciencia. Autoconciencia animal. José Luis Bermúdez. Test del espejo. Memoria episódica.

Abstract:

The aim of the paper consists on the problematic nature of the attribution of self-consciousness to non-human animals. In particular, three different paths are discussed: attribution on the basis of language, on the basis of certain experiments that do not require linguistic skills, and on the basis of the mere possession of phenomenal consciousness. While all of them provide strong reasons for the



attribution of self-consciousness, they also lead to serious problems. This paper argues that there is no infallible criterion for the attribution of self-consciousness, as well as that the view that self-consciousness is a defining human trait is questionable.

Keywords: Self-consciousness. Animal self-consciousness. José Luis Bermúdez. Mirror test. Episodic memory.

Introducción

La autoconciencia ha sido tradicionalmente considerada un rasgo definitorio de los seres humanos. Como señala Donald Griffin en *El pensamiento de los animales*: “Muchos filósofos, desde Aristóteles a Chomsky, Popper y Davidson, han insistido en que la verdadera experiencia consciente depende del lenguaje humano” (Griffin, 1986, p. 14). De hecho, todavía en la actualidad, buena parte de la comunidad científica y filosófica asume que esta es una peculiaridad exclusiva de los seres humanos (Pérez-Acosta, Benjumea & Navarro Guzmán 2001, p. 312). El lenguaje -y particularmente el dominio de ciertas expresiones como los pronombres personales de primera persona-, es así usualmente aceptado como el criterio determinante a la hora de establecer qué organismos son seres autoconscientes. Puesto que los seres humanos son los únicos organismos con capacidades lingüísticas, a ellos se reduce la autoconciencia.

No obstante, especialmente desde los años setenta, investigaciones pertenecientes a distintos campos de estudio han comenzado a poner esto en cuestión. Desde un terreno experimental, el psicólogo Gordon Gallup (1970) ideó y llevó a cabo un novedoso experimento, conocido como el test del espejo, con el que pretende comprobar mediante un método que no apele al lenguaje si hay animales no humanos autoconscientes. Al ser superado el test por chimpancés, el resultado será que ello es efectivamente así. La superación del test por parte de otros animales ha sido, con el discurrir de los años, una constante. Otro tipo de pruebas (por ejemplo: Eacott, Easton & Zinkivskay, 2005 o Smith, 2005) también fueron realizadas con estos



propósitos, dando como resultado una ampliación del número de especies animales que, de aceptarse el experimento, son autoconscientes en algún grado.

En el ámbito más filosófico o conceptual destacó la aportación de José Luis Bermúdez, quien en su conocida obra *The Paradox of Self-Consciousness* (2000 [1998]) denuncia la paradoja a la que conduce la asunción del lenguaje como criterio para la atribución de la autoconciencia, según la cual se produce una infructuosa circularidad explicativa entre las capacidades lingüísticas requeridas para el análisis de la autoconciencia y las capacidades mentales requeridas para el análisis de esas mismas capacidades lingüísticas. Como veremos, por este motivo el autor concluirá demandando la necesidad de reconocer estados mentales no conceptuales, esto es, no necesariamente lingüísticos, en animales y humanos no lingüísticos como los bebés.¹

El presente trabajo se centrará en las distintas razones aportadas para el reconocimiento de la autoconciencia en animales no humanos. Más en concreto, tendrá como principal objetivo el estudio de la plausibilidad de los distintos criterios aportados para la atribución de autoconciencia con el fin de poner en entredicho la tradicional consideración de que esta actúa como elemento diferencial entre individuos humanos y no humanos. Para ello, este se divide en tres apartados. En el primero de ellos, “La autoconciencia”, se presentará una definición general de autoconciencia que se irá matizando a lo largo del texto. El segundo apartado, denominado “Principales criterios para la atribución de autoconciencia”, se dedica al análisis de tres principales criterios no lingüísticos que se tomarán como fundamentales, así como a la descripción de los distintos experimentos a través de los cuales se evalúan. El tercer punto del trabajo, “La atribución “directa” de autoconciencia”, valora la posibilidad, realizada por algunas teorías de la consciencia, de que la propia posesión de algún grado de consciencia requiera necesariamente una forma de autoconciencia prereflexiva. Para ello, nos centraremos brevemente en los tres enfoques propuestos por Donald Griffin, Rocco Gennaro y Uriah Kriegel. Para terminar, se realiza con una

¹ Cabe clarificar, como se puede observar en su *Thinking without Words* (Bermúdez, 2003), que Bermúdez presentó una serie de argumentos destinados a la defensa de la existencia de contenidos mentales no conceptuales con independencia de la cuestión de la autoconciencia.



recapitulación de los corolarios de los distintos apartados, así como con dos conclusiones finales del artículo.

La autoconciencia

No parece haber ninguna duda de que cualquier persona que se encuentre leyendo estas palabras es autoconsciente en algún grado. Es una asunción general que la posesión de lenguaje nos da una poderosa razón para pensar que los seres humanos somos seres autoconscientes. A pesar de la disparidad de definiciones presentadas (Ben-Artzi, Mikulincer & Glaubman, 1995), así como de los matices que se irán introduciendo, por autoconciencia entenderemos en términos generales la forma de consciencia, y por tanto aquel estado mental, que tiene como referente a uno mismo como algo separado del mundo. Si bien esta es una cuestión polémica que se tratará más adelante, es importante destacar aquí una distinción tradicional presentada entre otros autores por Kant o James (Kriegel, 2007) entre el concepto de “uno mismo” como sujeto de su uso como objeto.² Así, de la oración “soy consciente de mí mismo” cabría diferenciar un “mí mismo” como sujeto de otro “mí mismo” entendido objeto. Esta será una distinción que mantendremos en nuestra definición de autoconciencia hasta el punto 3, en el que se polemizará este aspecto a partir de propuestas como la de Rocco Gennaro, quien establece una diferenciación entre la introspección como “mí mismo” referido a uno mismo como sujeto, de un “mí mismo” presentado como objeto del pensamiento.

La concepción deflacionaria de la autoconciencia

En su paradigmática obra en el estudio de la autoconciencia, *The Paradox of Self-Consciousness* (2000 [1998]), José Luis Bermúdez presenta la que considera que es la posición clásica acerca de la autoconciencia con el nombre de “concepción deflacionaria³ de la consciencia” (p. 10). El pilar

² Ya que, como veremos, algunos autores (véanse: Chisholm, 2002; Gennaro, 1996; Griffin, 1992; Kriegel, 2003 o Sartre, 1957) han sostenido una teoría de la autoconciencia sobre la asunción de que no es posible establecer tal distinción.

³ El uso del concepto “deflacionario” por parte de Bermúdez hace referencia al carácter reduccionista de este tipo de posturas sobre la autoconciencia.



fundamental de esta propuesta consiste en un conocido principio analítico, el “principio del pensamiento-lenguaje”:

The Thought-Language Principle. The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts.

[El principio del pensamiento-lenguaje. La única forma de analizar la capacidad de pensar un rango particular de pensamientos es analizando la capacidad para la expresión lingüística canónica de esos pensamientos] (p. 40).

En particular, esta vinculación entre la posesión de un lenguaje y la atribución de pensamientos se concretaría mediante otros dos principios:

The Conceptual Requirement Principle. The range of contents that one may attribute to a creature is directly determined by the concepts that the creature possesses.

[El principio del requisito conceptual. La gama de contenidos que uno puede atribuir a una criatura está directamente determinada por los conceptos que la criatura posee.] (p. 41).

The Priority Principle. Conceptual abilities are constitutively linked with linguistic abilities in such a way that conceptual abilities cannot be possessed by nonlinguistic creatures.

[El principio de prioridad. Las habilidades conceptuales están constitutivamente vinculadas con las habilidades lingüísticas, de tal manera que las habilidades conceptuales no pueden ser poseídas por criaturas no lingüísticas.] (p. 42).

En la medida en que se encuentra cimentada por estos principios, la concepción deflacionaria toma al lenguaje como criterio para la atribución de autoconciencia.⁴ Concretamente, tomará como

⁴ Donald Davidson fue uno de los filósofos por antonomasia que mantuvo con una mayor fuerza (al sostener la necesidad del lenguaje para que se produzca cualquier tipo de creencia) los dos principios precedentes, rechazando, por ende, la posibilidad de que seres no lingüísticos como los bebés o los animales no humanos puedan tener pensamientos: “una criatura no puede tener un pensamiento a menos que tenga un lenguaje” (Davidson, 2003, p. 148). Aplicado al caso de los animales, el principal argumento de Davidson se desenvolvería, siguiendo a Óscar Horta, como sigue:



referencia el uso semántico del pronombre “yo” en tanto aquel que denota de un modo paradigmático la autoconciencia (Smith, 2017). Desde esta panorámica, por lo tanto, “ser autoconsciente consiste en dominar la semántica del pronombre personal de primera persona (en castellano, “yo”, “me”, etc.) y otros relacionados (como los posesivos “mi”, “mío”)” (García, 2001, p. 167).

Como ya hizo notar Wittgenstein en su momento, existen, con todo, distintos usos posibles del concepto “yo”, debiéndose establecer consecuentemente una diferenciación:

Hay dos supuestos diferentes en el uso de la palabra “yo” (o “mi”), que yo llamaría “el uso como objeto” y “el uso como sujeto”. Son ejemplos del primer tipo de uso los siguientes: “Mi brazo está roto”, “yo he crecido seis pulgadas”, “yo tengo un chichón en la frente”, “el viento sopla sobre mi cabeza”. Ejemplos del segundo tipo son: “Yo veo tal-y-tal”, “yo oigo tal-y-tal”, “yo intento levantar mi brazo”, “yo pienso que lloverá”, “Yo tengo dolor de muelas”. Puede indicarse la diferencia entre estas dos categorías diciendo: los supuestos de la primera categoría implican el reconocimiento de una persona concreta, y en estos casos existe la posibilidad de un error, o como debería expresarlo mejor: se ha tenido en cuenta la posibilidad de un error [...] Por el contrario, no hay problema de reconocimiento de una persona cuando yo digo que tengo dolor de muelas. Carecería de sentido preguntar: “¿Está usted seguro de que es usted quien tiene dolores?” (Wittgenstein, 1968, pp. 100-101).

Wittgenstein establece así la distinción vinculada al uso del “yo” o “uno mismo” como sujeto y como objeto, ya nombrada precedentemente. En “Self-Reference and Self-Awareness” (1968), Shoemaker apuntaló esta distinción de Wittgenstein caracterizando el “yo” empleado como sujeto como “inmune al error por identificación equívoca” (1968, pp. 556-557). Debido al principio del

“(A1) Para tener cualquier clase de actitud proposicional es necesario tener conceptos.

(A2) Para tener conceptos es necesario tener una red de creencias.

(A3') Para poseer una red de creencias es necesario manifestar una conducta compleja.

(A4') Para manifestar una conducta compleja es necesario poseer un lenguaje.

(A5') Los animales no poseen un lenguaje.

(C) Los animales no pueden poseer actitudes proposicionales” (Horta, 2010, p. 58).



requisito conceptual se entiende que, en cualquiera de ambos casos, el uso del pronombre de primera persona requiere de la existencia de unos pensamientos de primera persona conocidos como “pensamientos del yo” (“I-thoughts”). Empero, ya sea empleado de una forma “inmune al error” o no, la posibilidad de llevar a cabo este tipo de proferencias es suficiente acorde a la concepción deflacionaria para poder tener autoconciencia con independencia de si la atribución de una determinada propiedad a uno mismo es o no errónea. Son los “pensamientos del yo” expresados como sujeto de la oración “soy consciente de mí mismo” aquellos susceptibles de ser empleados como criterio para la atribución de autoconciencia. Por ende, para la concepción deflacionaria de la autoconciencia, debido a su capacidad de emplear los pronombres de primera persona, solamente los seres lingüísticos pueden ser considerados seres autoconscientes.

Autoconciencia no conceptual

En el mismo trabajo citado, Bermúdez termina detectando que la aproximación deflacionaria de la autoconciencia incurre, con todo, en una explicación circular o paradójica, a la denomina “paradoja de la autoconciencia” (Bermúdez, 2000 [1998], p. 25). Esta se produce del siguiente modo: si el dominio semántico de pronombres en primera persona como “yo” es una condición necesaria para la atribución de autoconciencia, entonces resulta que la posesión de pensamientos de primera persona (“I-thoughts”) es también una condición necesaria. La razón de que esto sea así responde al principio del requisito conceptual según el cual los contenidos atribuibles a una criatura están vinculados a los conceptos que esta posee (p. 41). Es decir:

[...] dominar la regla de uso del pronombre personal de primera persona implica reconocer que uno mismo es el productor de la emisión de ese pronombre. Ahora bien, este reconocimiento requiere, a su vez, dominio del pronombre personal de primera persona. (García, 2001, p. 167).

Y es que, como apuntan los principios del requisito conceptual y de la prioridad, los contenidos atribuibles a un organismo están determinados por los conceptos que ella posee y estas habilidades



conceptuales están “constitutivamente unidas” (“constitutively linked”) con las habilidades lingüísticas. En consecuencia, acorde a la interpretación de Bermúdez de la concepción deflacionaria solamente los seres que dominen lingüísticamente las expresiones de primera persona son autoconscientes. En particular, las premisas que conducirían a esta posición a la paradoja son las siguientes:

1. The only way to analyze what is distinctive about self-consciousness is by analyzing the capacity to think ‘I’-thoughts.
[1. La única forma de analizar lo que es distintivo de la autoconciencia, es analizando la capacidad de razonar pensamientos del “yo”]
2. The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts (the Thought-Language Principle).
[2. La única forma de analizar la capacidad de pensar un rango particular de pensamientos, es analizando la capacidad para la expresión lingüística canónica de esos pensamientos (el Principio Pensamiento-Lenguaje)]
3. ‘I’-thoughts are canonically expressed by means of the first-person pronoun.
[3. Los pensamientos “yo”, se expresan canónicamente por medio del pronombre en primera persona]
4. Mastery of the first-person pronoun requires the capacity to think ‘I’-thoughts.
[4. El dominio del pronombre en primera persona requiere la capacidad de razonar pensamientos en “yo”] (Bermúdez, 2000 [1998], p. 24).

Ante estas premisas uno podría cuestionarse si la existencia de una circularidad argumentativa por parte de la concepción deflacionaria tiene que conducir, necesariamente, a un círculo vicioso falaz. Y es que, dado que no todo argumento circular tiene que ser vicioso, un defensor de la concepción deflacionaria podría argüir que sus premisas, explicitadas por Bermúdez, no incurren en falacia



alguna. Sin embargo, la existencia de esta falacia argumentativa recaería, asumiendo una posición no innatista, en la asunción, de una parte, de los pensamientos sobre uno mismo ('I'-thoughts) como único camino para explicar el uso de las expresiones de primera persona y, por otra parte, en la asunción de que el lenguaje (acorde al Principio del pensamiento-lenguaje) es la única forma de explicar la capacidad de tener pensamientos sobre uno mismo ('I'-thoughts).⁵ Dado que la aceptación de la posesión del lenguaje como único criterio para la atribución de autoconciencia conduce a una explicación circular, se concluye que la apelación al lenguaje no es un criterio aceptable. Por este motivo, en definitiva, la concepción deflacionaria no consigue explicar qué es ser autoconsciente.

Bermúdez sí acepta otras de las premisas asumidas por las concepciones deflacionarias, como aquella que reza: "5. A noncircular account of self-consciousness is possible" [5. Es posible una explicación no circular de la autoconciencia] (p. 24). Como es de esperar, para ello este tendrá que prescindir del principio del pensamiento-lenguaje pergeñando una distinción entre un modo de representarse el mundo conceptualmente y un modo no conceptual: "Self-awareness, Bermúdez concludes, has its origins not in the analysis of ideas arising from introspective phenomenology, but in an organism's nonconceptual understanding of its world" ["La autoconciencia, concluye Bermúdez, no tiene su origen en el análisis de las ideas que surgen de la fenomenología introspectiva, sino en la comprensión no conceptual de un organismo de su mundo"] (Power, 2001, p. 133). De este modo, defiende la existencia de contenidos no conceptuales de primera persona que consigan alcanzar una explicación no paradójica de la autoconciencia (Bermúdez, 2000 [1998], p. 49) y que, en consecuencia, permitan atribuir autoconciencia a seres no lingüísticos con independencia de si pueden expresarlo a través del pronombre personal. La paradoja a la que se ve conducida la concepción deflacionaria de la autoconciencia por su aceptación del principio del

⁵ En este punto conviene aclarar la identificación presupuesta en este trabajo entre la posesión efectiva de la autoconciencia y la atribución de la misma. Es decir, debido al problema de las otras mentes, asumimos la premisa escéptica de que no debemos aceptar como existente ningún estado mental que no sea propio de uno mismo a no ser que tengamos criterios aceptables para su atribución. Bajo esta perspectiva, el problema de la concepción deflacionaria reside esencialmente en su apelación al lenguaje como criterio para otorgar pensamientos sobre uno mismo ('I'-thoughts).



pensamiento-lenguaje lleva a Bermúdez al reconocimiento de unas habilidades no conceptuales básicas involucradas en la aparición de la autoconciencia. Por lo tanto, en opinión de este autor, es posible atribuir contenidos mentales representacionales a seres no lingüísticos que carecen de los conceptos necesarios para especificar cómo esos contenidos representan el mundo (p. 268). Para lo cual el mencionado principio del pensamiento-lenguaje deberá ser sustituido por otros principios, como el “principio de autonomía”:

The Autonomy Principle. It is possible for a creature to be in states with nonconceptual content, even though that creature possesses no concepts at all.

[El principio de autonomía. Es posible que una criatura esté en estados con contenido no conceptual, aunque esa criatura no posea ningún concepto]. (p. 61)

La teoría de los contenidos no conceptuales gravita sobre una teoría de la percepción no conceptual sostenida, entre otras razones, apelando a experimentos realizados con bebés prelingüísticos que mostrarían que estos son capaces de representarse el mundo no conceptualmente (p. 62-76). Algo que vendría motivado, a su vez, como inferencia a la mejor explicación del comportamiento de criaturas no lingüísticas (p. 271). Sin necesidad de profundizar en los pormenores aportados por este autor, para los objetivos del presente trabajo las ideas fundamentales que podemos extraer de lo dicho son dos: (1) la apelación al lenguaje como criterio para la atribución de autoconciencia fracasa al incurrir en una argumentación circular denominada “paradoja de la autoconciencia”, y (2) es posible proporcionar una teoría de la autoconciencia que apele a unas raíces lógicas y ontogenéticas de la autoconciencia más primitivas que el sistema conceptual que da lugar a formas más elevadas de autoconciencia (p. 294). Es importante aclarar, respecto a este último punto, que Bermúdez no niega completamente el principio de la prioridad presentado más arriba. Este acepta que existe una estrecha relación entre la posesión de habilidades lingüísticas y las conceptuales (García, 2001, p. 168). Por lo que en ningún momento niega una diferenciación gradual de los estados autoconscientes en virtud de su nivel de desarrollo lingüístico.



A falta de un análisis pormenorizado de las razones aportadas por Bermúdez para sostener que efectivamente los contenidos no conceptuales pueden servir de base para la autoconciencia, la relevancia de este punto para la consecución de los objetivos del trabajo recae, primeramente, sobre la dificultad de apelar al lenguaje como criterio para la atribución de autoconciencia y, en segundo lugar, en la posibilidad de presentar criterios no lingüísticos que den cuenta de la autoconciencia.

Principales criterios para la atribución de autoconciencia

Si bien la teoría de las representaciones no conceptuales de Bermúdez nos permite rechazar la apelación al lenguaje como criterio para la atribución de autoconciencia, resta todavía por determinar cuál o cuáles han de ser estos criterios en el caso de las criaturas no lingüísticas. Como acabamos de señalar, aunque el lenguaje no puede ser considerado tal criterio, de aquí no se sigue que su presencia siga siendo un claro indicador de la presencia de autoconciencia, por lo que permanece siendo, a pesar de todo, un criterio aceptable para su aplicación a los seres lingüísticos. El problema se presenta cuando nos referimos a seres que se representan el mundo de un modo no conceptual, como los bebés o los animales no humanos. Por este motivo, debemos encontrar criterios no referidos al lenguaje que nos aporten razones para atribuir autoconciencia a estos. Para ello, tomaremos como referencia en lo que sigue, siguiendo la propuesta de Smith (2017),⁶ tres criterios concretos: (i) el test del espejo, (ii) la posesión de memoria episódica y (iii) la metacognición. Asimismo, en el resto del texto nos centraremos en el caso de los animales no humanos.

⁶ Si bien mantenemos aquí la forma de expresión de estos criterios de Smith (2017), cabe tener presente que estos mismos han sido presentados, bajo otras denominaciones, por otros autores. Como es el caso de Pérez-Acosta, Benjumea y Navarro Guzmán (2001).



El test del espejo

El test del espejo (“mirror test”) fue propuesto en el año 1970 como criterio para la atribución de autoconciencia por parte de Gordon Gallup (1970). En un experimento realizado con chimpancés, y otros primates como los macacos, este investigador concluye que “insofar as self-recognition of one’s mirror image implies a concept of self, these data would seem to qualify as the first experimental demonstration of a self-concept in a subhuman form” [“en la medida en que el auto-reconocimiento de la propia imagen en el espejo implica un concepto de sí mismo, estos datos parecen calificar como la primera demostración experimental de un autoconcepto en una forma subhumana”] (Gallup, 1970, p. 87). Según el norteamericano, la capacidad para reconocerse en el espejo implica por ella misma la posesión de un concepto de “uno mismo” que garantiza la posesión de autoconciencia. Para llevar a cabo el experimento, previa exposición sistemática durante días al espejo, Gallup y su equipo anestesian a los chimpancés con el fin de marcar con tinta inodora roja un punto en su frente (p. 86). Una vez despiertos, los animales son capaces de reconocer la marca a través del espejo, por lo que se concluye que los chimpancés sí son seres autoconscientes. Llevado a cabo el mismo protocolo con otros primates, como los macacos rabones o Rhesus,⁷ se concluye que estos últimos no son capaces, a diferencia de los chimpancés, de reconocerse en el espejo aun a pesar de haber sido expuestos al mismo durante unas 250 horas (p. 87). Por lo que se concluye que este tipo de animales no son autoconscientes.

El éxito de este experimento condujo a su progresiva reproducción en otro tipo de especies con el objetivo de determinar cuáles de ellas son autoconscientes. De esta manera, se han encontrado no sólo otras especies primates que han pasado con éxito la prueba, como los orangutanes (Suárez & Gallup, 1981), sino toda una serie de especies animales más alejadas filogenéticamente de la humana. Este fue el caso, por ejemplo, de animales marinos como los delfines (Reiss & Marino, 2001), quienes son capaces de servirse de espejos para reparar en las marcas que habían sido hechas en sus cuerpos por los investigadores. Los elefantes son otro

⁷ Un estudio publicado en el año 2017 constatará que los macacos Rhesus sí consiguen pasar la prueba del espejo con éxito (véase: Chang et. al., 2017).



paradigmático ejemplo de mamíferos que consiguen pasar la prueba (Plotnik, de Waal & Reiss, 2006) al reaccionar ante el espejo en una serie de fases que van desde las respuestas sociales y la inspección física, hasta su uso para la detección de las marcas dibujadas en su cuerpo (pp. 17053-17054). Asimismo, se considera que el test del espejo ha sido superado con éxito por especies de animales no mamíferos como las urracas (Prior, Schwarz & Güntürkün, 2008). Tal y como sucede en el resto de los casos, el experimento consistente en el reconocimiento de sí mismo ante el espejo a través de una marca realizada sobre el individuo concluye con la evidencia, en vista de su conducta frente al espejo, de que estas tienen conciencia de sí mismas.

Por la contra, el test no ha sido superado con éxito por una amplia gama de especies animales. En algunos casos, este fracaso contrasta sobremanera con algunas de las capacidades propias de esas especies. Por ejemplo, a pesar de su alto nivel de complejidad social⁸ los perros no han superado esta prueba del modo en que ha sido planteada por Gallup (Horowitz, 2017).

El test del espejo ha sido objeto de numerosas críticas de las que destacaremos esencialmente dos. Por una parte, Heyes (1994) destacó que el uso del espejo por parte de un animal no implica que este tenga un concepto de sí mismo:

[...] there is absolutely no reason to suppose that an animal that can use a mirror in this way will also be able to either to contemplate its own mortality, or to appreciate that its reflection resembles the way in which it is viewed by conspecifics.

[...no hay absolutamente ninguna razón para suponer que un animal que puede usar un espejo de esta manera, también podrá contemplar su propia mortalidad o apreciar que su reflejo se parece a la forma en que lo ven sus congéneres] (Heyes, 1994, p. 915).

Para esta autora, el test muestra la capacidad de ciertos animales para poder servirse de un espejo para obtener información acerca de su cuerpo, pero no nos permitiría atribuir a estos un concepto de autoconciencia al requerir esto que los animales sean conscientes de su situación vital en un sentido amplio. Como, por ejemplo, en lo que concierne el reconocimiento de su propia

⁸ Lo cual, como señala Alexandra Horowitz, está altamente asociado a la diferencia entre uno mismo y el otro que se estima necesaria para la superación del test del espejo (2017, p. 17).



mortalidad.⁹ En segundo lugar, en relación con la no atribución de autoconciencia a especies animales como los perros, se ha criticado la prueba del espejo por estar focalizada exclusivamente en el sentido de la vista, por lo que autoras, como la mencionada Horowitz, han decidido recrear este test sirviéndose del olfato (2017). En esta nueva versión del experimento, los perros sí son capaces de reconocerse a sí mismos a través del olfato de su orina, que había sido ligeramente modificada, con respecto a la de la orina de otros perros (Horowitz, 2017, p. 21). Motivo por el cual se derivaría que los perros son seres autoconscientes en virtud del test del espejo.

La conclusión que podemos extraer de lo apuntado hasta aquí es la siguiente: tal y como sucede con el lenguaje, el test del espejo como criterio para la atribución de autoconciencia aporta razones, si bien cuestionables, para considerar que los individuos que lo superan tienen algún grado de autoconciencia,¹⁰ pero no es concluyente al respecto de aquellos que no lo hacen.

Memoria episódica

Otro de los criterios usualmente presentados para la atribución de autoconciencia reside en la posesión de memoria episódica.¹¹ A poco que se cavile acerca de este tipo de memoria, los motivos por los cuales esta ha sido defendida como condición para la posesión de autoconciencia parecen claros. La memoria episódica es un tipo particular de memoria declarativa o explícita. Esto es, el tipo

⁹ En relación con la cuestión del reconocimiento de la propia mortalidad como condición para ser considerado autoconsciente, y a falta de un mayor desarrollo en el artículo citado, cabría plantear a la autora la posibilidad de que aquellos sujetos humanos, como los niños, que no fuesen conscientes de esta mortalidad no podrían ser considerados autoconscientes. Esta podría evadir el problema asumiendo que efectivamente los niños de corta edad ajenos a su limitación temporal carecen de autoconciencia. Pero parece que este criterio sería excesivamente exigente como condición necesaria para la autoconciencia.

¹⁰ Como consecuencia de la problematización que es inherente a la apelación a diversos grados de autoconciencia, cabe aclarar que la hipotética existencia de un grado de autoconciencia “paradigmática” exclusiva a las capacidades cognitivas de los individuos humanos no pone en cuestión la hipótesis que mantiene el trabajo. A saber, que existen criterios no lingüísticos aceptables para considerar que individuos no lingüísticos, como los animales, tienen estados mentales autoconscientes. Aunque, a falta del lenguaje, no se pueda concretar un grado de complejidad específico de los mismos.

¹¹ A diferencia de los otros dos criterios, la memoria episódica es una habilidad conceptual, por lo que su asignación a individuos no humanos no sólo nos daría una razón para considerarlos seres autoconscientes, sino también seres conceptuales.



de memoria encargada de almacenar toda información relativa al mundo externo que resulte conscientemente evocable (Michaelian & Sutton, 2017; Squire, 2009). Dentro de la memoria declarativa es común distinguir la memoria semántica de la memoria episódica. Mientras que la memoria semántica lo es respecto de los contenidos proposicionales o conceptuales independientes de su vivencia biográfica por parte del sujeto, la memoria episódica conlleva una vivencia biográfica propia¹² (Renoult et. al. 2012, p. 550). La clave que permite establecer la relación entre la memoria episódica y la autoconciencia se encuentra en la representación de uno mismo a través del tiempo, que Tulving denomina la capacidad de “autonoesis” (Tulving, 2005, p. 14). Quien consigue proyectarse a sí mismo en el pasado mediante memoria episódica deberá ser considerado autoconsciente. O, dicho en otras palabras, no es posible que un sujeto recuerde ciertas vivencias de su pasado, por ejemplo una experiencia de angustia al quedarse encerrado en un lugar, sin que se recuerde a sí mismo como sujeto de tal vivencia. Lo que presupone la autoconciencia.

Dicho lo dicho, el peso de la argumentación parece recaer en los criterios que disponemos para la atribución de memoria episódica. Endel Tulving sostuvo en “Episodic Memory and Autonoesis: Uniquely Human?” (2005) que solamente los seres humanos son capaces de desenvolver a partir de una determinada edad la autonoesis necesaria para la memoria episódica. Entre las razones aportadas por el psicólogo estonio se encuentran las evidencias conductuales o la posesión de un nivel elevado de consciencia (Tulving, 2005, p. 30). La evidencia referida por Tulving para asignar un “nivel elevado de consciencia” a los individuos humanos comienza vinculándose a la superación del test del espejo (p. 31). Sin embargo, rápidamente rechaza el autor que la superación del test suponga la posesión de autonoesis: “However, passing the mirror test does not mean that the children have autonoetic consciousness, that they are aware of themselves as individuals with a past and a future” [“Sin embargo, pasar la prueba del espejo no significa que los niños tengan consciencia autonoética, que sean conscientes de sí mismos como individuos con un pasado y un futuro”] (p. 31). Finalmente, Tulving sitúa en los cuatro años de edad para los niños humanos el

¹² Mientras que la proposición “El Cairo es la capital de Egipto” constituye un caso de memoria semántica, la proposición “En el verano del 2018 estuve en El Cairo” lo sería respecto a la memoria episódica.



momento en el que se adquiere la memoria episódica (p. 32) basándose en un estudio de Gopnik y Graf (1988) en el que se concluye que es en torno a esa edad cuando los niños son capaces de recordar vivencias pasadas respondiendo a tres cuestiones esenciales: el cuándo, el cómo y el qué.

Tulving dedicó una sección del citado trabajo a la pregunta de si los animales tienen memoria episódica, a lo que responde negativamente. El principal argumento que tomó como referencia para negar su atribución a animales consiste en la imposibilidad de los animales no humanos en realizar los viajes temporales ("time travel") en los que consiste la autoconciencia (Tulving, 2005, pp. 37-40). Dice este: "My more recent thought on the topic was that episodic memory's adaptative values lies in the autoconciencia awareness of subjective time, and especially future time" ["Mi pensamiento más reciente sobre el tema fue que los valores adaptativos de la memoria episódica radican en la conciencia autoconciencia del tiempo subjetivo, y especialmente del tiempo futuro"] (p. 40). Mientras que el valor adaptativo de la autoconciencia es vital en los humanos por su relevancia en la evolución cultura, no sucede lo mismo en el caso de los animales no humanos:

The key notion behind human cultural evolution can be put simply: For anyone to take steps at one point in time that would make the unpredictable, frequently inhospitable natural environment more predictable at a future time, it is necessary to be able to be consciously aware of the existence of a future. An animal that cannot "think future" [...] that is, cannot think about time that has not yet arrived, will not initiate and persist in carrying out activities whose beneficial consequences will become apparent only in the future, at a time that does not yet exist. (p. 40).

No obstante, como señala Joel Smith (2017), sí parecen existir animales que responden al criterio del cuándo, cómo y el qué. El cual, acorde al estudio de Gopnik y Graf (1988) que toma como referencia Tulving, permitiría asignar memoria episódica a un sujeto. Este es el caso, por ejemplo, de las ratas (Eacott, Easton & Zinkivskay, 2005; Kart-Teke, et. al., 2006), la urraca californiana (Correia, Dickinson & Clayton, 2007) o incluso de los perros (Fugazza, Pogány & Miklósi, 2016).



Más allá de este debate acerca de la posibilidad de la memoria episódica en animales, su apelación para la atribución de autoconciencia tiene, al igual que el test del espejo, una serie de inconvenientes. El primero se corresponde con la dificultad para encontrar pruebas concluyentes para la asignación de memoria episódica (Tulving, 2005). Otro problema de relevancia se presenta con casos bien conocidos como el de K.C., en los cuales “[...] due to an accident, someone has lost episodic memory but appears to remain otherwise self-conscious” [...debido a un accidente, alguien ha perdido la memoria episódica pero parece seguir siendo consciente de sí mismo] (Smith, 2017), que mostrarían, precisamente, que la ausencia de memoria episódica no conlleva la ausencia de autoconciencia. De esta forma, a pesar de que un sujeto como K.C. carezca completamente de autoconciencia tanto en lo que respecta al pasado como al futuro (Rosenbaum, et. al. 2005, p. 993), de aquí no se puede colegir que K.C. no sea autoconsciente. Al menos en el sentido en que hemos venido definiendo la autoconciencia en este trabajo. En definitiva, la posesión de memoria episódica no puede ser una condición suficiente para la autoconciencia.

Metacognición

El último de los criterios mencionados se corresponde con la metacognición. Vinculada al pensar sobre el propio conocimiento, la metacognición puede ser definida brevemente como la capacidad para regular los propios estados cognitivos. Dicho de otro modo: “[i]n the literature on human cognition, awareness of what one knows is called ‘metacognition’ and it is associated with a ‘feeling of knowing’” “[e]n la literatura sobre la cognición humana, la conciencia de lo que uno sabe se llama ‘metacognición’ y se asocia con un ‘sentimiento de saber’”] (Allen & Trestman, 2017, p. 71). Así, en tanto control consciente, de orden superior, de la actividad cognitiva de uno mismo, se considera que la posesión de autoconciencia es una consecuencia de la metacognición (Smith, 2009, p. 389). La presencia de esta en humanos raramente se puede poner en duda debido a la posibilidad de estos para dar cuenta mediante el lenguaje de dicha capacidad. Con todo, en los seres no lingüísticos, como los animales, la situación varía al ser necesario recurrir a criterios conductuales. Por ejemplo, Smith destaca dos componentes presentes en los experimentos con animales.



Primeramente, los animales tienen que resolver una serie de pruebas de discriminación perceptiva de un nivel de dificultad tal que pueda propiciar incerteza en ellos (p. 389). En segundo lugar, además de las respuestas basadas en la discriminación perceptiva, se les aporta una respuesta adicional en las cuales pueden negarse a completar el experimento (p. 389). La consecuencia del ensayo consiste en que “[i]f animals monitor cognition accurately, they should prospectively recognize difficult trials as error-risking and decline those trials selectively” [“[s]i los animales monitorean la cognición con precisión, deberían reconocer prospectivamente las pruebas difíciles, como arriesgarse a cometer errores y rechazar esos intentos de forma selectiva”] (p. 389).

La estructura general de los ensayos realizados con animales, así pues, les permiten a estos dar tres tipos de respuestas. Dos de ellas se corresponden a la categorización respectiva fruto del estímulo perceptivo que deba clasificar. La tercera y más relevante recae en la posibilidad de que el animal renuncie a dar ninguna de las dos respuestas:

One common paradigm for testing metacognitive abilities involves presenting subjects with a stimulus that they must categorise in one of two ways. Crucially, they are also given the opportunity to opt out of the test, with correct categorisation resulting in the highest reward, opting out resulting in a lower reward, and incorrect categorisation resulting in no reward. The assumption is that the opt-out response reflects a meta-cognitive judgement of uncertainty.

[Un paradigma común para evaluar las habilidades metacognitivas consiste en presentar a los sujetos un estímulo que deben clasificar en una de dos formas. De manera crucial, también se les da la oportunidad de optar por no participar en la prueba, con la categorización correcta que resulta en la recompensa más alta, la exclusión voluntaria que resulta en una recompensa más baja y la categorización incorrecta que resulta en ninguna recompensa. La suposición es que la respuesta de exclusión refleja un juicio metacognitivo de incertidumbre.] (Smith, 2017).

Este tipo de experimento ha sido llevado a cabo con distintas especies, como es el caso de los delfines. Durante el mismo, los animales deben ir eligiendo, mediante la selección de la palanca



correspondiente, si el tono que están percibiendo supera o no los 2100-Hz (p. 248). Debido al reforzamiento positivo de las elecciones correctas y a la facilidad de las pruebas iniciales, el animal no muestra al principio duda alguna de su elección. Pero conforme la dificultad de las pruebas aumenta, también lo hace progresivamente la incerteza del delfín y, en consecuencia, su elección de la opción neutra (p. 248). La cual, cabe decir, tiene una recompensa menor que la opción correcta y mayor que la incorrecta.

Tornando el estímulo en visual, estudios semejantes han sido llevados a cabo con macacos (pp. 249-251), siendo los resultados parecidos a los de los delfines. Otros estudios también apuntan en la misma línea a la necesidad de reconocer cierto grado de metacognición en animales no mamíferos como ciertas aves (Fujita et. al., 2012). Por el contrario, animales como las ratas son incapaces de plasmar, con el aumento de la complejidad de las pruebas, una incerteza cognitiva que valorara el propio riesgo de la elección (Smith, 2005, p. 260).

Aceptando la relación de necesidad entre autoconciencia y metacognición en base a la capacidad autoreferencial a la que hace alusión esta última, en psicología comparada la problemática de este criterio orbita en torno al medio a través del cual se pretende constatar las capacidades metacognitivas en animales. Como Robert Hampton pone de manifiesto (2010), existen algunas razones para considerar que aquellos animales, como los delfines, que consiguen superar la prueba de la metacognición no lo hacen por poseer tales capacidades, sino más bien como consecuencia de su categorización de la no respuesta como una respuesta más. Respecto, por ejemplo, al caso del delfín destacado por Smith, Hampton presenta algunas posibilidades. Más que ser una consecuencia de la incerteza cognitiva del delfín, su progresiva elección de la opción de no responder podría deberse a la asociación entre la dificultad y la frecuencia de los tonos (Hampton, 2010, p. 21). O, por ejemplo, manteniéndose en una actividad mental de primer orden, el delfín pudo emplear su propia vacilación como estímulo discriminatorio (p. 22). Muy en la línea de las críticas de Hampton, la problematización que pivota sobre la atribución de metacognición a seres no lingüísticos no permite, pues, aceptarla *sensu stricto* como criterio para la atribución de autoconciencia a los animales no humanos.



La atribución “directa” de autoconciencia

Allende la apelación al lenguaje, o a medios más indirectos como los descritos en el apartado precedente, algunos autores han defendido que la posesión de cierto grado de consciencia es condición necesaria y suficiente para la autoconciencia. Entre los precedentes históricos más cercanos de este punto de vista se encuentra reseñablemente el filósofo francés Jean-Paul Sartre, quien asegura en *The Transcendence of the Ego* (1957) que la posesión de consciencia entraña eo ipso autoconciencia en un grado prereflexivo. Es decir, el propio fenómeno de la consciencia solamente puede ser posible si presuponemos un ego previo que la experimente. Algo que debemos distinguir de otro tipo de autoconciencia de orden superior de la que nos podemos percatar cuando hemos tenido un tipo de experiencia perceptiva, como un dolor en nuestro pie (Gallagher & Zahavi, 2014 [2005]). Ciertas posiciones neosartreanas son defendidas en la actualidad por parte de autores como Dan Zahavi, quien en su *Subjectivity and Selfhood* (2005) también amalgama la experiencia perceptiva, o consciencia fenoménica, con la autoconciencia.

Además de estas propuestas surgidas en el ámbito de la fenomenología, otros autores han defendido desde teorías de la consciencia dispares la posibilidad de otorgar autoconciencia a todo organismo que posea estados mentales conscientes. Esta es la postura que denominaremos bajo el epígrafe de atribución “directa” de autoconciencia al no necesitar apelar a ningún criterio, por ejemplo, conductual, ni en el caso de individuos humanos ni en el de no humanos. En ambos casos, las condiciones establecidas para la asignación de consciencia se identifican con aquellas que permiten atribuir autoconciencia. Para ilustrar esta idea, en lo que resta de apartado nos centraremos en las atribuciones “directas” de autoconciencia de tres autores: el zoólogo Donald Griffin y los filósofos Rocco Gennaro y Uriah Kriegel.

Donald Griffin

Tomando como referencia sus estudios etológicos acerca de la consciencia animal, Donald Griffin defendió la necesidad de atribuir a estos estados mentales autoconscientes. Para ello, el zoólogo estadounidense proporcionó tres tipos de argumentos. Dos de ellos, relativos al test del espejo y a



determinadas capacidades mnemotécnicas, ya los hemos tratado precedentemente en este artículo. De las evidencias proporcionadas por la prueba del espejo realizada por Gallup a chimpancés (1970) no mostrará ninguna duda Griffin, a raíz de lo cual llegará a decir que “[t]he question of self-awareness is one of the very few areas of cognitive ethology where we have some concrete experimental evidence” “[I]a cuestión de la autoconciencia es una de las pocas áreas de la etología cognitiva donde tenemos alguna evidencia experimental concreta”] (Griffin, 1992, p. 249). Algo harto semejante dijo respecto de la memoria (p. 250).

Ahora bien, el argumento de Griffin, con claro resabio sartreano, que cabe destacar en este apartado se refiere a su atribución de autoconciencia a los animales que tienen experiencias perceptivas conscientes. De hecho, el problema de aquellos que niegan la autoconciencia animal a la vez que admiten su consciencia perceptiva de primer orden es, para Griffin, evidente:

But those who hold that self-awareness is a unique human attribute often fall back to an insistence that although animals may be perceptually conscious of their own bodies, they nevertheless cannot think such thoughts as “It is I who am running”, or climbing this tree, or chasing that moth.

[Pero aquellos que sostienen que la autoconciencia es un atributo humano único, a menudo vuelven a insistir en que, aunque los animales pueden ser perceptualmente conscientes de sus propios cuerpos, sin embargo, no pueden pensar en pensamientos como “Soy yo quien estoy corriendo”, o subiendo este árbol, o persiguiendo esa polilla]. (p. 248).

Aunque manteniendo la distancia entre ambos, Griffin realiza una distinción semejante a la de Sartre al separar una autoconciencia de orden superior no necesariamente implicada por la experiencia perceptiva, de otra autoconciencia que sí lo está. Por este motivo, el animal que es perceptualmente consciente de su propio cuerpo debe ser también consciente, acorde a Griffin, de las acciones que está llevando a cabo, como correr, trepar a un árbol o perseguir una polilla: “[a]nd if the animal is perceptually conscious of its own body, it is difficult to rule out similar recognition that it, itself, is doing the running, climbing, or chasing” “[y] si el animal es perceptivamente consciente de su propio cuerpo, es difícil descartar un reconocimiento similar de que él mismo está



corriendo, trepando o persiguiendo”] (p. 249). Acorde a Griffin, la propia perspectiva perceptiva en primera persona del animal como sujeto supone que este tiene algún grado de consciencia de estar actuando, él mismo, de un modo concreto, aunque no lleve a cabo una reflexión sobre ello. Dicho de otro modo, descarta la posibilidad de que el animal perciba su cuerpo como una parte más del entorno.

Para ilustrarlo, supongamos el caso de un cerdo que está experimentando dolor en el matadero. Acorde a Griffin, esta percepción está ligada necesariamente a un sujeto que lo experimenta como sucediéndole a él mismo. Por lo que no hay motivos para negar, más allá de una “restricción arbitraria e injustificada” (p. 249), para considerar que ese cerdo es un ser autoconsciente. En tanto que el dolor es ejemplo de estado consciente, y ello, como acabamos de apuntar, implica según Griffin una perspectiva de primera persona donde el propio cuerpo en acción se diferencia del resto del entorno, la propia experiencia del dolor sería una prueba concluyente de la posesión de autoconciencia.

Ahora bien, el del dolor es precisamente un caso que puede dar pie a un problema para la propuesta de Griffin. Adoptando la terminología de los defensores de las HOT (“Higher-Order Theories”) que expondremos a continuación, la de Griffin puede ser considerada una teoría de la consciencia de primer orden (FOT) que se toparía con el problema de aportar algún criterio para diferenciar los estados mentales conscientes de los no conscientes (Gennaro, 2017, p. 93). Este es un problema que teorías de la consciencia de orden superior (HOT) conseguirían evadir introduciendo como elemento diferenciador las representaciones de orden superior (HOR), a la vez que consiguen mantener, al menos desde la posición de Gennaro, la consideración de que los seres conscientes son, en tanto tales, seres autoconscientes.

Rocco Gennaro y la teoría de la consciencia de orden superior (HOT)

Según ciertas teorías de consciencia, conocidas como teorías de la consciencia de orden superior (HOT), los estados mentales que son conscientes consisten en representaciones de orden superior (HOR) en relación con estados mentales de primer orden:



Conscious states are simply mental states we are conscious of being in. And, in general, our being conscious of something is just a matter of our having a thought of some sort about it. [...] We are then in a position to advance a useful, informative explanation of what makes conscious states conscious. Since a mental state is conscious if it is accompanied by a suitable higher order thought, we can explain a mental state's being conscious by hypothesizing that the mental state itself causes that higher-order thought to occur.

[Los estados conscientes son simplemente estados mentales en los que somos conscientes de estar. Y, en general, el hecho de que seamos conscientes de algo es solo una cuestión de que tengamos algún tipo de pensamiento al respecto. [...] Entonces estamos en condiciones de avanzar en una explicación útil e informativa de lo que hace que los estados conscientes sean conscientes. Dado que un estado mental es consciente si va acompañado de un pensamiento de orden superior adecuado, podemos explicar el hecho de que un estado mental sea consciente planteando la hipótesis de que el estado mental mismo hace que se produzca ese pensamiento de orden superior.] (Rosenthal, 1986, pp. 335-336).

En consecuencia, para las HOT los sujetos son conscientes únicamente en el caso de que posean unas habilidades conceptuales lo suficientemente complejas como para elaborar representaciones de orden superior (HOR). Por este motivo, uno de sus principales defensores, Peter Carruthers (2000 o 2019),¹³ mantiene que los animales no humanos no pueden ser considerados ni tan siquiera conscientes. De hecho, este es el caso no solo de los animales, sino de todos los seres no lingüísticos que carecen de la habilidad para elaborar conceptualmente una HOR respecto a alguna experiencia de primer orden (Carruthers, 2000, pp. 201-203). La adscripción de experiencias conscientes a los animales responde, en opinión de Carruthers, a una representación del propio sujeto que realiza esa adscripción en un intento de representársela en primera persona, pero desde el punto de vista del animal (p. 198). Por esta razón, en el intento, los sujetos humanos

¹³ A pesar de que las razones de Carruthers para negar la consciencia de los seres no lingüísticos siguen siendo los mismos, conviene aclarar que en *Human and Animal Minds* (2019) abandonó su tradicional defensa de una teoría de orden superior por la teoría del espacio de trabajo global desarrollada inicialmente por Baars (1988).



están sufriendo una especie de ilusión de consciencia en los animales al asignarles sus propias experiencias fenoménicas. Pero estas experiencias nunca lo son respecto a los propios animales puesto que el sujeto no puede “salirse” de sus propios estados conscientes. No rechaza que los animales no humanos sientan (“feeling”) ciertos sabores o experimenten dolor de primer orden. Lo que niega Carruthers es que estas experiencias o sensaciones sean conscientes. Es decir, estas son sensaciones percibidas sin el factor de la subjetividad (“What-it-is-likeness”) (pp. 200-201). Como se puede ver, Carruthers establece así una distinción entre la experiencia de ver el color verde de la experiencia fenoménica subjetiva de la “verdad”.

Otro paradigmático teórico de las HOT, Rocco Gennaro, admite, como Carruthers, que las experiencias conscientes solamente son posibles cuando el sujeto puede realizar la conceptualización requerida para la elaboración de la HOR. La divergencia entre ambos se encuentra en la complejidad necesaria para llevar a cabo esta representación de orden superior. De este modo, Gennaro introduce la posibilidad de que la HOR se realice mediante una conceptualización básica con conceptos como “estar viendo rojo”: “But why couldn’t those animal HOTs simply contain concepts more like ‘looking-red’ or ‘seeing red’? Is it so implausible to ascribe *these* concepts to most animals? I think not” [“Pero, ¿por qué esos HOT animales no podrían simplemente contener conceptos más como ‘mirar rojo’ o ‘ver rojo’? ¿Es tan inverosímil atribuir estos conceptos a la mayoría de los animales? Yo creo que no”] (Gennaro, 2004, p. 48.). La representación consciente de ver que algo es rojo no tiene por qué residir en un concepto elaborado de la “experiencia del rojo”, sino en otro más sencillo que consista en un estar “viendo rojo”. Respecto al caso del dolor se entiende que, del mismo modo, para que un sujeto sea consciente de una experiencia consciente no tiene que poseer el concepto de “dolor”, sino representaciones como “esto duele” o “esta sensación tan poco placentera”. Finalmente, la atribución de estados conscientes a los animales no resulta para Gennaro (p. 50) una ilusión fruto de las propias representaciones, sino que, por la contra, tiene una mayor capacidad explicativa y predictiva de la conducta de los animales.



Debido a que la consciencia requiere para ser tal una representación de orden superior (HOR), Gennaro defiende que el hecho mismo de la consciencia presupone la autoconsciencia: “So if I am right about what makes mental states conscious, then *having conscious states entails self-consciousness*” [“Entonces, si tengo razón sobre lo que hace que los estados mentales sean conscientes, entonces tener estados conscientes implica autoconsciencia”] (Gennaro, 1996, p. 17). Este filósofo lleva a cabo una diferenciación entre varios grados de autoconsciencia, de tal manera que esta puede llegar a ser un estado mental no consciente al no haber ninguna representación de orden superior respecto a ella (p. 17). Como se puede observar, el quid del asunto reside según la propuesta de Gennaro en la posibilidad de que la autoconsciencia se dé con independencia de si consiste en un estado consciente de algo como si no. Es decir, en si es posible que el sujeto se forme una representación de la propia autoconsciencia. La resistencia a aceptar que la consciencia supone la autoconsciencia respondería a una confusión consistente en identificar la autoconsciencia con la introspección (p. 18). Mientras que los estados mentales conscientes no implican introspección, sí lo hacen respecto a la autoconsciencia en la medida, reitera, en que la autoconsciencia no tiene por qué ser ella misma consciente. Algo que sí sucede con la introspección, entendida como una forma compleja de autoconsciencia (p. 19) referida a un estado mental propio.

Kriegel y la teoría autorepresentacional de la consciencia

Sobre la base de una teoría autorepresentacional de la consciencia, Uriah Kriegel mantiene una posición muy semejante a la adoptada por Gennaro. Según el autorepresentacionalismo de la consciencia, un estado mental es consciente siempre que, sea lo que sea que represente, también se represente a sí mismo (Kriegel, 2006, p. 59). Por ejemplo:

[W]hen you sit in the tub and consciously stare at your toes, it is true both that you are aware of the toes, and that you are aware (though more dimly) of your visual experience of the toes. According to the self-representationalist, this is because your conscious experience at the time represents both your toes and itself.



[[C]uando te sientas en la bañera y miras conscientemente los dedos de tus pies, es cierto que eres consciente de los dedos de los pies y que eres consciente (aunque más tenuemente) de tu experiencia visual de los dedos de los pies. Según el autorrepresentacionista, esto se debe a que su experiencia consciente en ese momento representa tanto los dedos de sus pies como a sí mismo] (p. 59).

Quienes defienden esta posición consideran que los problemas de otras teorías representacionistas de la consciencia, como las teorías representacionistas de primer orden (FOT) y las de orden superior (HOT), son solventados por esta. Así, por ejemplo, a las HOT se les presenta comúnmente un inconveniente conocido como el “argumento a partir de la carencia de objeto de las representaciones de orden superior” (p. 62). Según las HOT existe la posibilidad de que un sujeto pueda tener la impresión de que ser consciente de algo cuando, en realidad, no lo está siendo. Alguien puede tener la impresión de estar saboreando una onza de chocolate cuando de hecho no está comiendo ninguna onza de chocolate. La ausencia de objeto supone que la representación de orden superior está representando incorrectamente y que, en consecuencia, no se está produciendo una experiencia consciente aun cuando el sujeto tenga la impresión de ello. Algo que para Kriegel resulta “absurdo” (p. 62). La posición autorepresentacionista consigue evitar este problema al requerir únicamente que las representaciones, para ser conscientes, tengan que hacer referencia a sí mismas, y no a un objeto que puede estar ausente. De esta manera, esta teoría conseguiría, para sus defensores, preservar la idea de que los estados conscientes son aquellos de los cuales nos percatamos (p. 62).

En una línea semejante a la de Gennaro, Kriegel afirma que la autoconciencia es una “condición necesaria” para la consciencia (p. 106), por lo que esta continuaría siendo una teoría de la consciencia a partir de la cual los animales no humanos deben ser tomados como seres autoconscientes. Kriegel distingue a este respecto dos tipos de autoconciencia. Por una banda, la “autoconciencia transitiva” en donde “*x* is self-conscious of her thought that *p*” y, por la otra, la “autoconciencia intransitiva” donde “*x* is self-conscious thinking that *p*” (p. 103).



Esta es una distinción que guarda gran semejanza con la realizada por Gennaro entre la introspección y la autoconciencia respectivamente. Por lo cual ya podemos atisbar cuál de los dos tipos de autoconciencia es considerada por Kriegel como condición necesaria para la consciencia, la autoconciencia intransitiva. Este vínculo es formulado por el autor como sigue: “A mental state M of a subject x at a time t is conscious only if M represents its own occurrence” [Un estado mental M de un sujeto x , en un tiempo t , es consciente solo si M representa su propia ocurrencia] (p. 108). Digamos que, por ejemplo, mi creencia de que dentro de una caja de zapatos hay una galleta será consciente siempre y cuando me pueda representar esa misma creencia. Tener un estado mental consciente conlleva el tener una representación del mismo o, dicho de otro modo, conlleva ser intransitivamente autoconsciente de ese estado mental. En consecuencia, la atribución de consciencia a los animales conduciría, tal y como sucede en las posiciones de Griffin y Gennaro, a la atribución de autoconciencia a los mismos.

Conclusiones

Este trabajo ha tenido como principal objetivo el cuestionamiento de la idea según la cual la autoconciencia es un rasgo definitorio de los seres humanos que permiten distinguirlo del resto de los animales. Para ello, se ha trazado un recorrido en tres etapas. En la primera de ellas, hemos partido de la común asunción de que el lenguaje es el principal criterio para la atribución de autoconciencia. De esta manera, el dominio de la gramática de la primera persona (por ejemplo, el índice “yo”) sería una condición necesaria y suficiente para que se pueda otorgar con un alto grado de verosimilitud la autoconciencia a seres distintos de uno mismo. Tomando las palabras de José Luis Bermúdez, esta es la tesis mantenida por la concepción deflacionaria de la autoconciencia. Sin embargo, siguiendo al mismo autor, hemos constatado cómo este enfoque incurre en una paradoja insalvable al asumir que el análisis de la autoconciencia se tiene que llevar a cabo mediante el lenguaje y que, simultáneamente, el dominio del lenguaje de primera persona está condicionado por la posesión de pensamientos sobre uno mismo (*‘I’-thoughts*). Este círculo explicativo nos conduce a un primer corolario:



- (1) Si bien el lenguaje continúa siendo una prueba sólida de la posesión de autoconciencia (y, por ende, puede funcionar como condición suficiente), la paradoja denunciada por Bermúdez inutiliza su comprensión como condición necesaria para la atribución de autoconciencia.

Dada la neutralización del lenguaje como principal criterio, la necesidad de encontrar otras pautas para la atribución de autoconciencia, concretamente a animales no humanos, nos ha conducido a la consideración de tres criterios constatables experimentalmente: el test del espejo, la memoria episódica y la metacognición. A pesar de que ciertas especies animales, además de la humana, son capaces de superar uno o incluso todas las pruebas relativas a estos criterios, los inconvenientes que estos entrañan dificultan la posibilidad de que la superación de alguno de ellos, o de todos en su conjunto, sea una garantía de la posesión de estados mentales autoconscientes. Empero, al igual que sucede con el lenguaje, si bien estos criterios no pueden ser tomados como pruebas fehacientes de la autoconciencia, sí nos proporcionan ciertos indicios (especialmente en el caso de especies, como los chimpancés o delfines, que cumplen satisfactoriamente más de una de las pruebas). Esta idea nos conduce a una atribución “bajo mínimos” de autoconciencia:

- (2) Si bien no constituye una garantía acerca de la posesión de autoconciencia, la superación de alguna de las pruebas relativas al test del espejo, la memoria episódica o la metacognición, aporta razones para poder atribuir autoconciencia.

Finalmente, hemos considerado la posibilidad de una atribución “directa” de autoconciencia a través de la atribución de la consciencia perceptiva más básica. Con este fin hemos tenido en cuenta algunas propuestas teóricas acerca de la consciencia cuya asunción puede conllevar que la autoconciencia está implicada por la posesión de consciencia. Esta especie de autoconciencia pre-reflexiva, denominada, por ejemplo, “autoconciencia intransitiva” por Kriegel, sería ella misma la



condición de posibilidad de la consciencia. Puesto que la aceptación de esta atribución “directa” se encuentra supeditada a la aceptación de teorías de la consciencia como las descritas, la conclusión que extraemos a este respecto es la siguiente:

- (3) Si bien se encuentra supeditado a su aceptación, ciertas teorías de la consciencia pueden conllevar la necesidad de atribuir cierto grado de autoconsciencia a los organismos que sean conscientes.

La necesidad de aceptar todo un pertrecho teórico para atribuir autoconsciencia a los seres conscientes no permite derivar un criterio concluyente para quienes no acepten tales teorías. Por mor de ello, la incapacidad para aportar un criterio infalible para la atribución de autoconsciencia permanece. A pesar de que ello es así, la aceptación de las conclusiones del presente trabajo, especialmente de (1) y de (2), contribuyen a poner seriamente en entredicho la asunción de que la autoconsciencia, independientemente de su nivel de complejidad, es una característica diferencial de los seres humanos respecto al resto de animales. Por todo lo mencionado, dos son las conclusiones generales del presente trabajo:

- (4) No se ha hallado ningún criterio que resulte infalible para la atribución de autoconsciencia.

Y, no obstante:

Tomados en su conjunto, los distintos criterios presentados permiten poner seriamente en entredicho la tradicional consideración de que la autoconsciencia funciona como elemento diferencial entre individuos humanos y no humanos.



Referencias

- Allen, C. & Trestman, M. (2017): "Animal Consciousness", en S. Schneider & M. Velmans (Eds.). *The Blackwell Companion to Consciousness*. Chichester, West Sussex: Wiley Blackwell, pp. 63-76.
- Baars, B. (1988): *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Ben-Artzi, E., Mikulincer, M., & Glaubman, H. (1995): "The Multifaceted Nature of Self-Consciousness: Conceptualization, Measurement, and Consequences". *Imagination, Cognition and Personality*, 15, 17-43.
- Bermúdez, J. L. (2000 [1998]): *The Paradox of Self-Consciousness*. Cambridge (EE.UU): MIT Press.
- Bermúdez, J. L. (2003): *Thinking without Words*. New York: Oxford University Press.
- Carruthers, P. (2000): *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Carruthers, P. (2019): *Human and Animal Minds*. Oxford: Oxford University Press.
- Chang, L., Zhang, S., Poo, M. & Gong, N. (2017): "Spontaneous Expression of Mirror Self-Recognition in Monkeys After Learning Precise Visual-Proprioceptive Association for Mirror Images". *Proceedings of the National Academy of Sciences*, 114, 3258-3263.
- Chisholm, R. (2002): *Person and Object: A Metaphysical Study*. London: Routledge.
- Correia, S., Dickinson, A., & Clayton, N. (2007): "Western Scrub-Jays Anticipate Future Needs Independently of their Current Motivational State". *Current Biology*, 17, 856-861.
- Davidson, D. (2003): *Subjetivo, intersubjetivo, objetivo*. Madrid: Cátedra.
- Eacott, M., Easton, A., & Zinkivskay, A. (2005): "Recollection in an Episodic-Like Memory Task in the Rat". *Learning & Memory*, 12, 221-223.
- Fugazza, C., Pogány, A., & Miklósi, A. (2016): "Recall of Others' Actions After Incidental Encoding Reveals Episodic-Like Memory in Dogs". *Current Biology*, 26, 3209-3213.
- Fujita, K., Nakamura, N., Iwasaki, S. & Watanabe, S. (2012): "Are Birds Metacognitive?". En M. J. Beran, J. Brandl, J. Perner & J. Proust (Eds.). *Foundations of Metacognition*. Oxford: Oxford University Press, 50-61.



- Gallagher, S. & Zahavi, D. (2014 [2005]): "Phenomenological Approaches to Self-Consciousness", en E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. URL: <https://plato.stanford.edu/entries/self-consciousness-phenomenological/> [consultado el 29/10/2021].
- Gallup, G. (1970): "Chimpanzees: Self-Recognition". *Science*, 167, 86-87.
- García, A. (2001): "Autoconciencia y contenido no conceptual". *Revista de Filosofía*, 26, 165-185.
- Gennaro, R. (1996): *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. Amsterdam: John Benjamins.
- Gennaro, R. (2004): "Higher-Order Thoughts, Animal Consciousness, and Misrepresentation: A Reply to Carruthers and Levine". En R. Gennaro (Ed.). *Higher Order Theories of Consciousness. An Anthology*. Indiana: John Benjamins, 45-66.
- Gennaro, R. (2017): *Consciousness*. New York: Routledge.
- Gopnik, A. & Graf, P. (1988): "Knowing How You Know: Young Children's Ability to Identify and Remember the Sources of Their Beliefs". *Child Development*, 59, 1366-1371.
- Griffin, D. (1986): *El pensamiento de los animales*. Barcelona: Ariel.
- Griffin, D. (1992): *Animal Minds*. Chicago: University of Chicago Press.
- Hampton, R. (2010): "Multiple Demonstrations of Metacognition in Nonhumans: Converging Evidence or Multiple Mechanisms?". *Comparative Cognition & Behavior Reviews*, 4, 17-28.
- Heyes, C. (1994): "Reflections on Self-Recognition in Primates". *Animal Behaviour*, 47, 909-919.
- Horowitz, A. (2017): "Smelling Themselves: Dogs Investigate their Own Odours Longer when Modified in an "Olfactory Mirror" Test". *Behavioural Processes*, 143, 17-24.
- Horta, O. (2010): "¿Quién puede poseer actitudes proposicionales?". *Teorema*, 29, 55-68.
- Kart-Teke, E., De Souza, M., Huston, J. & Dere, E. (2006): "Wistar Rats Show Episodic-Like Memory for Unique Experiences". *Neurobiology of Learning and Memory*, 85, 173-182.
- Kriegel, U. (2003): "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument". *Canadian Journal of Philosophy*, 33, 103-132.
- Kriegel, U. (2006): "Consciousness, Theories of". *Philosophy Compass*, 1, 58-64.
- Kriegel, U. (2007): "Self-consciousness". *The Internet Encyclopedia of Philosophy*. Documento online: <http://www.iep.utm.edu/self-con/> [consultado el 25/10/2021].



- Michaelian, K. & Sutton, J. (2017): "Memory", en E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy*. Documento online: <https://plato.stanford.edu/entries/self-consciousness/> [consultado el 12/11/2021].
- Pérez-Acosta, A., Benjumea, S. & Navarro Guzmán, J. (2001): "Autoconciencia animal: Estudios sobre la autodiscriminación condicional en varias especies". *Revista Latinoamericana de Psicología*, 33, 311-327.
- Plotnik, J., de Waal, F. & Reiss, D. (2006): "Self-Recognition in an Asian Elephant". *Proceedings of the National Academy of Sciences*, 103, 17053-17057.
- Power, N. S. (2001): "The Origins of Self-Consciousness". *Minds and Machines*, 11, 133-137.
- Prior, H., Schwarz, A., & Güntürkün, O. (2008): "Mirror-Induced Behavior in the Magpie (*Pica pica*): Evidence of Self-Recognition". *PLoS biology*, 6(8), e202.
- 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, 5937-5942.
- Renoult, L., Davidson, P., Palombo, D., Moscovitch, M., & Levine, B. (2012): "Personal Semantics: at the Crossroads of Semantic and Episodic Memory". *Trends in Cognitive Sciences*, 16, 550-558.
- Rosenbaum, R. S., Köhler, S., Schacter, D., Moscovitch, M., Westmacott, R., Black, S., Gao, F., & Tulving, E. (2005): "The Case of KC: Contributions of a Memory-Impaired Person to Memory Theory". *Neuropsychologia*, 43, 989-1021.
- Rosenthal, D. (1986): "Two Concepts of Consciousness". *Philosophical Studies*, 49, 329-359.
- Sartre, J.-P. (1957): [*The Transcendence of the Ego: An Existentialist Theory of Consciousness*](#). New York: Octagon Books.
- Shoemaker, S. (1968): "Self-Reference and Self-Awareness". *Journal of Philosophy*, 65, 555-567.
- Smith, J. D. (2005): "Studies of Uncertainty Monitoring and Metacognition in Animals and Humans", en H. Terrace & J. Metcalfe (Eds.). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York: Oxford University Press, pp. 242-271.
- Smith, J. D. (2009): "The Study of Animal Metacognition". *Trends in Cognitive Sciences*, 13, 389-396.



- Smith, J. (2017): "Self-Consciousness", en E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy*. Documento online: <https://plato.stanford.edu/entries/self-consciousness/> [consultado el 19/11/2021].
- Squire, L. (2009): "Memory and Brain Systems: 1969–2009". *Journal of Neuroscience*, 29, 12711-12716.
- Suárez, S. & Gallup, G. (1981): "Self-Recognition in Chimpanzees and Orangutans, but not Gorillas". *Journal of Human Evolution*, 10, 175-188.
- Tulving, E. (2005): "Episodic Memory and Autonoesis: Uniquely Human?". En H. Terrace & J. Metcalfe (Eds.). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York: Oxford University Press, pp. 3-56.
- Wittgenstein, L. (1976): *Los cuadernos azul y marrón*. Madrid: Tecnos.
- Zahavi, D. (2005): *Subjectivity and Selfhood: Investigating the First Person Perspective*. Massachussets: MIT Press.