Conciencia, primera persona y contenido no conceptual

1. Introducción

Uno de los problemas más fascinantes en la filosofía de la mente es el de entender la naturaleza de nuestras experiencias conscientes. Las experiencias conscientes tienen una dimensión subjetiva: *hay algo que es para el sujeto* tener una experiencia consciente (Nagel, 1974), esto es, se siente de determinada manera tener dicha experiencia. Hay algo que es para mí, por ejemplo, mirar mi computadora mientras escribo, sentir un ligero dolor de cabeza o escuchar de fondo la música de *Tool*. A ese *algo* que es para mí tener una experiencia consciente se le denomina *carácter fenoménico* de la experiencia.

Una forma interesante de aproximarse al problema de la consciencia consiste en diseccionarlo cuidadosamente. Esta estrategia ha sido propuesta por autores como Kriegel (2009b) o Levine (2001), quienes comienzan realizando una distinción conceptual entre dos componentes del carácter fenoménico (con sus problemas asociados): el carácter cualitativo y el carácter subjetivo. Al considerar el problema del carácter cualitativo nos preocupamos de la pregunta acerca de qué caracteriza la diferencia entre las distintas experiencias conscientes, esto es, de la cuestión de por qué, por ejemplo, se siente de distinta manera ver una manzana roja que oler una rosa. El problema del carácter subjetivo deja de lado la forma particular en que se siente tener una experiencia consciente, centrándose en el problema de entender en virtud de qué estar en un estado consciente se siente de alguna manera. Así, podemos pensar en el carácter cualitativo como aquello que hace que una experiencia sea el tipo de estado consciente que es, mientras que el carácter subjetivo es aquello que hace que un estado sea una experiencia consciente en absoluto (Kriegel, 2009b). Este trabajo se centra en este segundo aspecto de la experiencia.

Al reflexionar sobre el carácter subjetivo de la experiencia, debemos preguntarnos por el rasgo que permite individuar las experiencias conscientes: el rasgo que los estados conscientes, y solo los estados conscientes, poseen. A diferencia de otro tipo de estados, los estados conscientes no son estados que ocurran simplemente *en mi interior* —como ocurre,

por ejemplo, con mi proceso digestivo—, sino que de alguna forma *son para mí*: hay algo que es *para mí* tener la experiencia. Al tener una experiencia consciente, digamos, del aroma a epazote que desprende la sopa de tortilla, no me percato (*awareness*) únicamente de dicho aroma, sino que, además, me percato de que estoy en cierto estado. Esta idea, que podemos rastrear hasta autores como Aristóteles, Descartes, Locke o Kant (Rosenthal y Weisberg, 2008), queda capturada por el *principio de transitividad* que defiende Rosenthal (1997, 2005), según el cual que un estado sea consciente consiste en que uno sea consciente de dicho estado. La motivación intuitiva de este principio reside en la idea de que si alguien tiene un estado —ya sea un pensamiento, una percepción o una sensación corporal— del que en absoluto se percata, no parece que tenga sentido catalogarlo como estado consciente. El objetivo será, entonces, entender en qué sentido nos percatamos de cierto estado al tener una experiencia consciente y cuál es la naturaleza de tal percatación.

Una propuesta interesante consiste en no tomar tal relación de percatación como primitiva —lo que simplemente trasladaría el misterio de la conciencia de un nivel a otro —, sino tratar de entenderla en términos representacionales. Si entendemos la percatación en estos términos, la consciencia involucra algún tipo de autorrepresentación (self-representation). Ahora bien, como he discutido en trabajos anteriores (Sebastián, 2012, 2020b, MS; ver también Kriegel, 2003), decir que los estados conscientes son autorrepresentacionales resulta ambiguo: puede querer señalar que un estado consciente se representa a sí mismo, o bien que representa a uno mismo.[1] A fin de evitar confusiones, podemos llamar al primer tipo de autorrepresentacionismo, 'representacionismo centrado en el estado' y al segundo tipo 'representacionismo centrado en el sujeto'. Estas dos posiciones reflejan dos afirmaciones distintas con respecto al principio de transitividad que podemos encontrar en la literatura. Rosenthal y Weisberg, por ejemplo, presentan el principio de transitividad del siguiente modo: “podemos referirnos a la idea de que el hecho de que un estado sea consciente consiste en que uno sea consciente de ese estado como el *principio de transitividad* (Rosenthal, 1997)”[2] (2008, énfasis en el original). Sin embargo, el propio Rosenthal es claro, en múltiples ocasiones, en su preferencia por una lectura en concordancia con un representacionismo centrado en el sujeto, como veremos en detalle más adelante. Al desarrollar su teoría añade: “Que un estado sea consciente consiste en que sea un estado del que uno es consciente de que uno mismo está”[3] (2005, p. 211). Centraré mi trabajo en esta segunda lectura.

Con el objetivo de presentar y defender una teoría que dé cuenta del carácter

---

1 El origen de esta confusión resulta aun más clara en inglés. Decir que un estado es *self-representational* puede indicar que el estado se representa a sí mismo (*itself*) o a uno mismo (*oneself*), dando lugar a lo que he llamado (Sebastián, 2012) respectivamente 'mental-involving' (representacionismo centrado en el estado) o 'self-involving' (representacionismo centrado en el sujeto).

2 “we can refer to the view that a state's being conscious consists in one's being conscious of that state as the *Transitivity Principle.”* (Rosenthal, 1997)

3 “A state’s being conscious consists in its being a state one is conscious of oneself as being in.”

subjetivo de la experiencia, haciendo justicia al principio de transitividad, este trabajo está organizado como sigue. En la siguiente sección discutiré con más detalle la motivación del principio de transitividad para defender una lectura centrada en el sujeto y expondré una semántica que nos permita entender este tipo de representaciones *de se*. Las teorías de orden superior y, más concretamente, las teorías de pensamiento de orden superior, han capitaneado los intentos de dar cuenta del principio de transitividad. A ellas va dedicada la sección 3 donde me centraré en mostrar por qué su propuesta resulta insatisfactoria. Finalmente, en la sección 4 esbozaré una propuesta que no sufre estos problemas.

2. El principio de transitividad: representación centrada en el sujeto

Recordemos que una teoría del carácter subjetivo pretende dar condiciones para la individuación de las experiencias conscientes: qué distingue a los estados conscientes de los que no lo son. El principio de transitividad, unido a un marco representacionista, ofrece una respuesta en este sentido al señalar que las experiencias conscientes son estados autorrepresentacionales. No obstante, para poder discernir la lectura adecuada de este principio, debemos ir más allá de la motivación intuitiva a la que hacía mención en la sección anterior. Cabe considerar tres posibles caminos para intentar mostrar que la percatación de uno mismo sea necesaria para la conciencia.

El primero sería argumentar que se trata de un principio con validez *a priori*. El propio Nagel, al introducir su expresión, hace hincapié en que al tener una experiencia consciente hay algo que es *para el sujeto* estar en ese estado. Sin embargo, esta ruta parece poco prometedora pues como ha mostrado Byrne (2004), resulta dudoso que la expresión 'hay algo que es para X φ-ar' tenga algún tipo de uso especial para describir la subjetividad. Como el mismo Byrne señala, preguntas del tipo '¿cómo fue para el coche ser conducido en el desierto?', parece ser una pregunta inteligible que uno puede hacer. Sin embargo, no derivaríamos de ella ningún tipo de consciencia en el coche.

El segundo camino lo presenta Rosenthal (2005), quien sugiere que la justificación del principio de transitividad la debemos buscar en nuestra psicología del sentido común (*folk-pschology*). Más concretamente, en la idea de que un estado del que no nos percatamos en absoluto, no cuenta en ningún sentido relevante como un estado consciente. Sin embargo, surgen dos problemas en esta propuesta. En primer lugar, no queda claro cómo eso bastaría para justificar un representacionismo centrado en el sujeto —como el que sostiene Rosenthal—. En segundo, y más importante aún, apelar a la psicología del sentido común no parece una opción satisfactoria a menos que podamos justificar a su vez las creencias que la constituyen, en particular las que refieren a nuestros estados

conscientes (Kriegel, 2009a). Para hacer tal justificación, una vez excluido el análisis conceptual, la opción parece ser recurrir al tercer camino: la observación y reflexión sobre el carácter fenoménico de la experiencia, sobre lo que es fenoménicamente dado.

Todas mis experiencias parecen compartir el hecho de que, de alguna forma, están implícitamente marcadas como *mis experiencias.* Las experiencias conscientes ocurren *para mí* (el sujeto que está teniendo la experiencia) de una forma inmediata. Todas presentan un carácter distintivo de primera persona, lo que, a falta de un término más adecuado, Kriegel (2009b) llama *para-mí-idad (for-me-ness*) y Block (2007), *mí-idad* (*me-ishness*). Una descripción más detallada de este rasgo, que en la tradición fenomenológica se reconoce como 'autoconciencia prerreflexiva', la presentan Gallagher y Zahavi (2006):

> Es como algo saborear el chocolate, y esto es diferente de lo que es recordar lo que es saborear chocolate u oler vainilla, correr, permanecer quieto, sentir envidia, depresión o felicidad, o tener una creencia abstracta. Ahora bien, al mismo tiempo que vivo esas diferencias, hay algo experiencial que es, en algún sentido, lo mismo, esto es, su distintivo carácter de primera persona. Todas las experiencias están caracterizadas por una cualidad de *mí-idad* o *para-mí-idad*, por el hecho de soy yo quien está teniendo esas experiencias. Todas las experiencias están dadas (al menos tácitamente) como experiencias que yo estoy teniendo o viviendo. Todo esto, sugiere que la experiencia de primera persona me presenta con un acceso inmediato y no observacional a mí mismo, y que consecuentemente la conciencia (fenoménica) implica una forma mínima de autoconciencia.[4]

Mi experiencia no solamente me dice cómo es el mundo, sino que yo estoy en cierta relación con ese mundo. Al mirar una manzana roja, mi experiencia no solo me dice del mundo que contiene una manzana roja, sino que soy yo el que está en cierta relación con esa manzana. La experiencia consciente no solo concierne a lo que podemos llamar su objeto primario, la manzana y sus propiedades en el ejemplo, o al estado en que me encuentro, sino también al propio sujeto que está teniendo la experiencia como tal, como sujeto de la experiencia.

Entender en qué sentido la experiencia concierne al sujeto es imprescindible para continuar la discusión. Yo soy Miguel Ángel Sebastián (MAS). Cuando tengo una experiencia visual, esta no dicta *que MAS está en cierto estado perceptivo*, sino *que yo*

---

[4] There is something it is like to taste chocolate, and this is different from what it is like to remember what it is like to taste chocolate, or to smell vanilla, to run, to stand still, to feel envious, nervous, depressed or happy, or to entertain an abstract belief. Yet, at the same time, as I live through these differences, there is something experiential that is, in some sense, the same, namely, their distinct first-personal character. All the experiences are characterized by a quality of mineness or for-me-ness, the fact that it is I who am having these experiences. All the experiences are given (at least tacitly) as my experiences, as experiences I am undergoing or living through. All of this suggests that first-person experience presents me with an immediate and non-observational access to myself, and that consequently (phenomenal) consciousness consequently entails a (minimal) form of self-consciousness.

estoy en cierto estado perspectivo. Podemos motivar esta idea de forma sencilla apelando a la creencia que uno está en posición de formarse de forma inmediata en virtud de tener la experiencia. Simplemente, por el hecho de tener la experiencia, estoy en posición de creer *que yo estoy en cierto estado.* Por el contrario, no es cierto que, simplemente por el hecho de tener esa experiencia, esté en posición de creer que MAS está en cierto estado: puedo ignorar el hecho de que yo soy MAS, en cuyo caso no me formaría tal creencia pese a poseer todos los conceptos necesarios. Caracterizar entonces el contenido de la experiencia requiere emplear lo que Perry ha llamado *indéxico esencial* (1979). La experiencia concierne al sujeto como tal e involucra en este sentido una representación de primera persona o representación *de se* (Lewis, 1979). Pasemos pues a analizar cómo entender de forma general este tipo de representaciones que conciernen al sujeto de la misma como tal.

Representaciones *de se*

Una visión muy popular acerca del contenido mental, y que yo adoptaré en este trabajo, es aquella que mantiene que el papel de los estados mentales es distinguir entre diferentes posibilidades (Stalnaker, 1999). Los estados mentales representan el mundo como siendo de cierta manera, y en este sentido resultan correctos o incorrectos dependiendo de cómo sea el mundo: los estados mentales dividen el espacio de posibilidad determinando conjuntos de mundos posibles. Un ejemplo puede ayudar a esclarecer esta idea. Consideremos la creencia de que en México hay policías que aceptan mordidas. Esta creencia distingue dos formas en que puede ser el mundo: puede ser que en el mundo actual en México haya policías que aceptan mordidas o puede ser que no los haya. Esta creencia hace una partición en el espacio de posibilidad distinguiendo unos mundos de otros en función de si hay o no policías que aceptan mordidas en México. No solo eso, sino que, además, toma parte por uno de esos conjuntos, determinando una función que va de mundos posibles a valores de verdad o corrección. Así, decimos que es correcta o verdadera si el mundo actual se encuentra entre los mundos en los cuales hay policías que aceptan mordidas e incorrecta o falsa en caso contrario.

Los estados mentales que conciernen al propio sujeto como tal —a saber, de las representaciones *de se*— suponen un problema para esta forma de entender el contenido de los estados mentales. Comparemos mi creencia de que MAS está tomando café y mi creencia de que el único filósofo con una playera de Los Pumas escribiendo un artículo sobre consciencia en *La Roma* está tomando café. Parece claro, que yo puedo tener una, sin tener la otra, pero no parece el caso que demanden lo mismo del mundo actual para ser verdaderos. Por lo tanto, podemos capturar sin problema las diferencias entre ellos en términos de mundos posibles. Consideremos, por el contrario, mi creencia de que yo estoy tomando café. Como sugería con anterioridad, parece que yo puedo tener esta última creencia sin creer que MAS está tomando café —si por cualquier razón ignorara el hecho

de que yo soy MAS—. Sin embargo, ambas parecen demandar lo mismo del mundo, esto es, que cierto individuo concreto esté tomando café. Es más, no parece que podamos dar las condiciones de corrección de la creencia *de se* (Lewis, 1979; cf. Stalnaker, 2008) en términos de mundos posibles. Por ello, Lewis sugiere que las representaciones *de se* no determinan particiones en el espacio de posibilidad en mundos posibles, sino en mundos centrados, esto es, conjuntos de pares de mundos posibles e individuos (<w,i>). Si podemos pensar los mundos posibles como formas en las que el mundo puede ser, entonces podemos pensar los mundos centrados como la forma en que un mundo es para un individuo (Egan, 2006a). Así, la creencia que expresaríamos mediante la oración 'Yo escribí *Sobre la pluralidad de mundos*' es una creencia que podemos compartir Lewis y yo. Para evaluar las condiciones de verdad de la misma no basta con considerar qué mundo es el actual, necesitamos, además, un individuo. La creencia resulta verdadera con respecto al par <$w_{@}$, Lewis>, pero falsa con respecto al par <$w_{@}$, MAS>. En este sentido, las representaciones *de se* determinan funciones de mundos centrados a valores de verdad.[5]

Entender este tipo de representaciones nos ayuda también a disolver el escepticismo que algunos lectores pudieran tener sobre la idea de que la experiencia concierna al sujeto. Por ejemplo, Hume (1739) rechazaba la idea de que hubiera un sujeto de experiencia, con base en la incapacidad de encontrar algo así como un sujeto en la introspección. Y muchos estarían de acuerdo con esta observación de Hume. Pero, ¿no entra esto en tensión directa con la idea de que la experiencia concierne al sujeto? La respuesta es negativa y es sencillo verlo una vez que hemos entendido el tipo de representación involucrada (Sebastián, 2020b, MS). Para verlo con claridad, consideremos la experiencia visual que tengo al ver la taza de café a la derecha de mi computadora. Esta experiencia representa algo así como —dejando el sujeto a un lado para facilitar la exposición— que hay una taza de café a la derecha de la computadora *desde aquí*, pero no que hay una taza de café a la derecha de la computadora *desde cierto punto en la ciudad de México.* Como prueba de ello, pensemos que el lector tendría una experiencia del mismo tipo si se le presentara una configuración idéntica aun cuando estuviese situado en otra localización distinta. En tales circunstancias, podemos asumir que nuestras experiencias serían idénticas y nos dirían algo así como *que hay una* taza de café a la derecha de la computadora *desde aquí.* En general, nuestra experiencia visual representa el mundo desde un punto concreto en el espacio e involucra de forma no controvertida una representación indéxica *de hinc* (desde aquí). Podemos caracterizar su contenido con la herramienta presentada como colecciones de mundos centrados, tales que

[5] Nótese que cualquier contenido en términos de mundos posibles es expresable en términos de mundos centrados, simplemente resulta correcto o incorrecto con independencia del sujeto que consideremos. Por ello, Lewis (1979) argumenta que toda representación *de dicto* es reducible a una representación *de se*. Egan (2006b) ha distinguido entre contenidos *de se interesantes*, aquellos cuyas condiciones de corrección varían dependiendo del individuo que se considere, y *de se aburridos* —aquellos que no—. En el resto del artículo cuando hable de contenidos *de se* me estaré refiriendo a representaciones *de se interesantes*.

el centro es una localización espacial —situada aproximadamente en mitad de los ojos—. Ahora bien, si le pido al lector que introspecte tal localización, probablemente levante la vista del texto con cara de perplejidad. No encontramos una localización privilegiada al introspectar nuestra experiencia visual. Aun así, no concluimos de ello que la experiencia no concierna a tal localización, es decir, que tal localización no sea parte de las condiciones de adecuación del estado. La forma en la que el sujeto es representado en toda experiencia es análoga a la forma en la que, además, cierta localización es representada en la experiencia visual.[6]

He considerado que los estados mentales hacen particiones en el espacio de posibilidad y en qué consisten las mismas —mundos centrados en el sujeto—. La pregunta que surge a continuación es cómo realizamos este tipo de particiones. En una primera aproximación, podemos decir que hacemos particiones en mundos posibles cuando atribuimos propiedades a los objetos, esto es, cuando representamos los objetos como teniendo propiedades. En el ejemplo de los policías, la creencia en cuestión atribuye la propiedad de recibir mordidas a policías, y es verdadera en caso de que haya policías con esa propiedad y falsa en caso contrario. En el caso de las representaciones *de se*, lo que se precisa es una *autoatribución* de propiedades (Lewis, 1979) —donde, por las razones comentadas, que un sujeto *S* se autoatribuya una propiedad no es reducible a la atribución de propiedades a *S* por parte de *S*—. Así en la creencia de que yo escribí *Sobre la pluralidad de mundos*, me autoatribuyo la propiedad de haber escrito la obra en cuestión, pero la creencia resulta falsa pues carezco de tal propiedad. Cuando Lewis tiene esta creencia, esta resulta, sin embargo, verdadera ya que se estaría autoatribuyendo una propiedad que sí posee.

Continuando con el proceso de desenmarañar la naturaleza de este tipo de representaciones, la siguiente pregunta que debemos afrontar es: ¿cómo hace un sistema para autoatribuirse una propiedad? Lewis no nos da muchas pistas y simplemente señala que “La autoatribución de propiedades es la atribución de propiedades a uno mismo bajo la relación de identidad”[7] (1979, 543). En la sección 4 desarrollaré mi propuesta a partir de esta idea. Sin embargo, una respuesta que puede venir de forma inmediata a nuestra mente es apelar a la utilización del pronombre 'yo' o a su equivalente en el lenguaje del pensamiento. Es aquí donde las teorías de orden superior, y más concretamente las teorías de pensamiento de orden superior, entran en juego. A ellas y a los problemas que enfrentan para dar cuenta del carácter subjetivo de la experiencia va dedicada la próxima sección.

---

6 Para una discusión de lo que ocurre con la autoconsciencia en casos de experiencias alteradas, véase Sebastián (2020a).

7 “Self-ascription of properties is ascription of properties *to oneself under the relation of identity.”*

3. Teorías de orden superior

Con base en el principio de transitividad y en la idea de que un estado consciente es uno en el que me percato de estar, las teorías de orden superior tratan de dar cuenta de la diferencia entre estados conscientes y no conscientes, esto es, de ofrecer una teoría del carácter subjetivo apelando a una representación de orden superior. Los estados conscientes son el objeto de algún tipo de proceso de orden superior o representación. En el caso de los estados conscientes hay algo de orden superior, un metaestado, que no está presente en el caso de otro tipo de estados. El principal punto de desencuentro entre este tipo de teorías radica en la naturaleza del estado de orden superior. Los defensores de las teorías del *sentido interno* (Armstrong, 1968; Carruthers, 2000; Lycan ,1996) defienden que tiene forma de percepción, mientras que las teorías de *pensamiento de orden superior* —teorías HOT (*Higher-Order Thought*). Véase Brown, 2015; Gennaro, 1996, 2012; Rosenthal, 1997, 2005; Weisberg, 2011— mantienen que la representación de orden superior tiene la forma de un pensamiento.

Las teorías de sentido interno defienden que para tener una experiencia consciente hace falta un estado (cuasi-) perceptivo dirigido al estado de primer orden, que hace que me percate de él. Sirva de ilustración de este tipo de teorías un breve esbozo de la teoría presentada por Carruthers (2000). Carruthers parte de una *semántica de consumidor* (*consumer semantics*), según la cual, a grandes rasgos, el contenido de un estado mental depende de los poderes del organismo que consume dicho estado (Millikan, 1984, 1989; Papineau, 1993; Peacocke, 1995). Por ejemplo, lo que un estado representa depende del tipo de inferencias que el sistema cognitivo esté dispuesto a hacer en presencia del estado. Según Carruthers, algunos de los estados perceptivos de primer orden adquieren al mismo tiempo un contenido de orden superior en virtud de su disponibilidad a la facultad de la teoría de la mente —responsable de que podamos apreciar la diferencia entre cómo las cosas son y cómo parecen—, combinado con la validez de alguna versión de una teoría semántica de consumidor. De esta forma, un percepto de rojo puede ser al mismo tiempo una representación de que algo *es rojo* y, estando disponible para la teoría de la mente, una representación de que algo *parece rojo*, pasando, gracias a esto último, a ser una experiencia consciente. Este tipo de teorías puede dar cuenta del principio de transitividad bajo una lectura centrada en el estado, al explicar en qué sentido me percato de estar en cierto estado, pero no parece poder dar cuenta de la lectura deseada —centrada en el sujeto — según la cual la experiencia consciente concierne al propio sujeto que está en ese estado.[8] Pasemos pues a ocuparnos de las teorías HOT.

[8] En una línea similar, pero más general, Rosenthal (2011, p. 25) mantiene que si fuéramos conscientes en virtud de un sentido interno, tan solo podríamos percatarnos del estado mental —ya que las sensaciones representan solo las correspondientes propiedades perceptibles—, pero en ningún sentido también del

De acuerdo con las teorías HOT, cuando miro una manzana roja y tengo una experiencia visual, me encuentro en un estado con cierto contenido que podemos llamar ROJO.[9] Para que el estado sea consciente debo tener, además, un pensamiento adecuado (asertivo y no inferencial) de orden superior que tenga al anterior como objeto, un pensamiento del tipo '*yo* veo rojo' o '*yo* estoy teniendo una representación de rojo' (Lau y Rosenthal, 2011). Esta representación de orden superior representa que *uno mismo* está teniendo o instanciando cierta representación de primer orden. A diferencia de lo que ocurría con las teorías de sentido interno, las teorías HOT acomodan *prima facie* a la perfección la idea de que la experiencia consciente concierne al sujeto de la experiencia como tal. Estas teorías sufren, sin embargo, problemas graves como veremos a continuación.

Problemas de la teorías HOT

En esta subsección pretendo presentar tres problemas que la teorías HOT enfrentan para mostrar por qué no resultan satisfactorias: el problema de la motivación, el problema de la fundamentación de la referencia y el problema de la inmunidad.

El problema de la motivación

El primer problema para las teorías HOT se deriva de su incapacidad para justificar el principio que las motiva. La razón estriba en lo siguiente: el estado, en virtud del cual un estado es consciente, es un estado inconsciente, por lo tanto, parece que estas teorías tienen que sostener que el carácter subjetivo no es fenoménicamente dado. Pero de ser así, y como habíamos visto, no queda claro cómo se justifica la creencia en el principio de transitividad. Procedamos con más detalle.

Según el principio de transitividad, que un estado sea consciente requiere que el sujeto sea consciente de sí mismo como estando en dicho estado. Ahora bien, ¿qué requiere ser consciente de uno mismo como estando en ese estado de primer orden? Según la teoría, otro metaestado que me haga consciente de que soy consciente de mí mismo como estando en el estado de primer orden, y así *ad infinitum* (Caston, 2002; Kriegel, 2009b; Williford, 2006). Rosenthal bloquea este regreso vicioso al mantener que el estado en virtud del cual un estado es consciente no necesita ser a su vez ser consciente. El estado de orden superior solo es consciente cuando hago introspección, lo que a su vez requiere un pensamiento de tercer orden que tenga al segundo como objeto, pero ese pensamiento de tercer orden sería

sujeto al que pertenece ese estado mental.

[9] Merece la pena ser cauteloso en este punto, pues Rosenthal (2005) niega que el estado de primer orden en casos como este sea intencional y defiende que las cualidades mentales correspondientes a ese estado de primer orden representan de forma no-intencional. Este tipo de detalles no son relevantes para los propósitos de este trabajo, ya que se centra en la representación de orden superior.

a su vez inconsciente. El precio de esta solución resulta ahora claro: nada acerca del propio sujeto —¡ni del estado en cuestión!— es fenoménicamente dado, pues el pensamiento de orden superior es típicamente inconsciente y, en consecuancia, el principio de transitividad se queda sin el soporte necesario.[10]

Hay, no obstante, una posición en el espíritu de las teorías HOT que no estoy convencido de que esté afectado por este problema. En un artículo reciente, Richard Brown (2015) ha defendido una teoría HOT según la cual tener una experiencia consciente consiste en tener un pensamiento que dicte que uno mismo está en cierto estado representacional.[11] La sutil diferencia es que el estado consciente se identifica con el de orden superior en lugar del estado de primer orden. Con ello se viola cierta lectura del principio de transitividad según la cual me percato de mí mismo como teniendo una experiencia consciente, sustituyéndolo por un principio suficientemente cercano —y a mi parecer más plausible— según el cual, al tener una experiencia consciente, me percato de mí mismo como estando en cierto estado.[12] En este caso, alguien podría pensar que su versión de la teoría niega que la experiencia dicte algo del sujeto —y con ello que el carácter subjetivo sea fenoménicamente dado—, si uno acepta que lo que es fenoménicamente dado queda determinado por lo que el pensamiento de orden superior representa. El problema de esta posición radica en explicar la diferencia entre el carácter fenoménico de, digamos, la creencia de que yo estoy en un estado que representa rojo (asumiendo que hay algo que es para mí tener una creencia) y la experiencia de ver algo rojo. Con la teoría de Rosenthal claramente puedo explicar esa diferencia. En la experiencia visual yo tengo un pensamiento inconsciente de estar en un estado que representa rojo, mientras que este pensamiento no está presente en la creencia inconsciente. Y si la creencia es consciente, tampoco hay problema: el pensamiento de orden superior representaría que yo tengo la creencia de que estoy en un estado que representa rojo. Por el contrario, si, como sugería con anterioridad, alguien quisiera hacer uso de la propuesta de Brown para mantener que el carácter subjetivo es fenoménicamente dado, esta opción no estaría disponible. Tanto en el caso de la creencia como en el de la experiencia, tengo el pensamiento asertivo de que yo estoy en un estado que representa rojo. A dicho estado puedo haber llegado de forma que, subjetivamente, parezca no inferencial y cuyo contenido es *que yo estoy en un estado que representa rojo*. Por lo tanto, no queda claro cómo dar cuenta de la diferencia fenoménica que existe entre creer que uno está en un estado que representa rojo y, por otro lado, mirar un objeto rojo.

---

10 Ver Kriegel (2009a) para una elaboración más detallada de esta objeción.

11 Brown llama a su teoría HOROR (Higher-Order Representation Of a Representation).

12 Sin embargo, parece que supone abandonar la idea de que un estado consciente es un estado del que me percato. En Sebastián (MS) argumento que no hay razón para abandonar el principio de transitividad una vez entendemos que hay dos usos distintos del término ‘estado consciente’: uno para referirse al estado en virtud del cual tengo la experiencia y otro para referirse al estado del que me percato.

El problema de la fundamentación de la referencia

De acuerdo con las teorías HOT, tener una experiencia consciente depende de tener un pensamiento que incluye la utilización del concepto YO. Muchos han objetado que hay razones para poner en duda que bebés y animales no humanos, a los que tenemos pocas razones para negar que puedan tener experiencias conscientes, sean capaces de poseer este tipo de conceptos.

En respuesta, Gennaro (2012) afirma que hay diferentes conceptos YO que podrían jugar el papel requerido por la teoría. En orden de sofisticación menciona: i) YO *qua* esta cosa (o cuerpo), por oposición a otras cosas físicas, ii) YO *qua* experimentador de estados mentales, iii) YO *qua* cosa pensante que persiste, iv) YO *qua* pensador entre otros pensadores. Gennaro discute una variedad de evidencia empírica que sugiere que tanto los bebés como la mayoría de los animales tienen, como mínimo, un concepto YO como el dado por (i). Sin embargo, no está claro cuál es la relación entre dicho concepto y el contenido de la experiencia como contenido *de se*, que parece requerir algo en línea con (ii). La posesión de tal concepto parece depender de la existencia de experiencias, algo que el defensor de las teorías HOT no puede permitirse. Puede ser que Gennaro esté en lo correcto al señalar que bebés y animales pueden poseer una variedad de conceptos YO, y ello, sin duda, le permite no tener que afirmar que carecen de experiencias conscientes. Sin embargo, el verdadero problema no radica ahí, sino en explicar los mecanismos que fijan la referencia del concepto YO relevante sin apelar a las experiencias conscientes, ya que, de acuerdo con la teoría, la posesión del concepto es anterior a la experiencia misma: es la posesión del concepto YO el que fundamenta la posibilidad de tener experiencias conscientes y no al revés.

Un autor que recoge ese guante es Rosenthal (2011b). Él está de acuerdo en que la referencia del concepto YO en el pensamiento de orden superior refiere a uno mismo como tal, como el individuo que representa: “la percatación del estado representa tácitamente el estado como perteneciendo al mismo individuo que se percata” (2011b, p. 271).[13] Rosenthal afirma que, aunque el pensamiento de orden superior no describe al individuo como el pensador de ese pensamiento, el individuo tiene la disposición a realizar esa identificación si se le pregunta. Como la pregunta rara vez se da, el individuo no tiene por qué hacer tal identificación. Esta disposición supone una identificación tácita y explica, de acuerdo con Rosenthal, cómo el pensamiento de orden superior refiere a uno mismo de forma esencialmente indéxica. Esta explicación no resulta satisfactoria (Sebastián, 2018b).

En primer lugar cabe notar que el mecanismo de referencia está fundamentado en la disposición a identificar al individuo que tiene el pensamiento de orden superior con el referente de ese pensamiento. Como consecuencia, la disposición a hacer tal identificación

[13] “One's awareness of the state tacitly represents that state as belonging to the very individual that has that awareness.”

es una condición necesaria para poder emplear el correspondiente concepto YO y, por tanto, según la teoría HOT, para tener el pensamiento de orden superior del cual depende a su vez la posibilidad de tener experiencias conscientes. Ahora bien, como hemos visto, el pensamiento de orden superior es típicamente inconsciente, pues el pensamiento de orden superior se vuelve consciente solo cuando uno tienen un pensamiento de tercer orden que tiene al anterior por objeto. Pero si el pensamiento es inconsciente, resulta complicado entender cómo el sujeto podría tener la disposición a identificar al individuo al que el pensamiento refiere. Esto se debe a que mientras el pensamiento sea inconsciente, el individuo no sabe el sujeto de qué pensamiento debe identificar: si el pensamiento es inconsciente, no está accesible realizar la identificación en cuestión.

En respuesta a este argumento, uno puede notar acertadamente que, pese a que el pensamiento de orden superior sea inconsciente, uno sigue teniendo la disposición a formarse el pensamiento de tercer orden que lo haga consciente, y que esto es suficiente para garantizar que uno tenga la disposición a identificar al individuo que tiene el pensamiento de orden superior con el referente de ese pensamiento. Sin embargo, no basta para salvar la propuesta de Rosenthal, ya que ese pensamiento de tercer orden incluiría el concepto YO —su contenido sería algo así como *que yo estoy teniendo un pensamiento de segundo orden*—. Para explicar cómo el concepto YO empleado en ese pensamiento de tercer orden refiere, tendríamos que apelar a su vez a la disposición a formarse un pensamiento de cuarto orden que permitiera hacerlo consciente, y así poder tener la disposición a identificar el individuo que tiene el pensamiento de tercer orden con el referente del mismo. Pero el estado de cuarto orden también hace uso del concepto YO y tenemos que explicar cómo es que ese concepto refiere…, y así sucesivamente. Rosenthal fundamenta la capacidad de tener experiencias conscientes en la disposición a tener una jerarquía infinita de pensamientos de orden superior, y parece razonable dudar que tengamos esa capacidad y con ello tal disposición. Más aun, si aceptamos que para tener una disposición, ha de existir la posibilidad de que dicha disposición se manifieste,[14] ha de ser posible que, en determinadas circunstancias, esa cadena de disposiciones, de hecho, se manifieste. Pero dado que nuestras capacidades cognitivas son limitadas, no hay forma de tener el pensamiento de que yo estoy en un estado que representa que yo estoy en un estado que representa que .... (y así una infinidad de veces) ... que yo estoy en un estado que representa rojo. Por tanto, un sistema de capacidad limitada, como somos nosotros, carece de la disposición que fundamenta la referencia del concepto YO que requiere la teoría de Rosenthal. Pero, de hecho, los sistemas con capacidad cognitiva limitada como nosotros tienen experiencias conscientes, por lo tanto debemos rechazar la teoría de Rosenthal.[15]

---

[14] Pensemos en una propiedad disposicional como la ser frágil. Un objeto es frágil si y solo si tiene la disposición a romperse con facilidad. Para ello, ha de ser posible que, dadas ciertas circunstancias, el objeto de hecho se rompa.

[15] Véase Sebastián (2018b) para una presentación pormenorizada de este argumento.

En segundo lugar, y de forma relacionada, el mero hecho de que el pensamiento de segundo orden haya de ser consciente, o al menos haya de poder serlo —de forma que tengamos la posibilidad de identificar al individuo que tiene el pensamiento con el referente del mismo—, hace que la capacidad de tener pensamientos de tercer orden sea una condición necesaria para la conciencia. Este hecho eleva excesivamente los requisitos para tener una experiencia consciente. Además, hay razones para dudar que los pensamientos de tercer orden puedan existir en seres extralingüísticos como son la mayoría de los animales y los bebés humanos. Pues parece razonable pensar que el lenguaje es al menos necesario para tener pensamientos acerca de estados puramente intencionales como son los pensamientos de orden superior (Rosenthal, 2005, ch.10).

El propio Rosenthal considera un problema semejante derivado de la disposición requerida, pues, aun cuando concedamos la capacidad de tener pensamiento a animales no humanos y a bebés humanos, estos carecerían de la disposición para hacer la identificación en cuestión. Como respuesta a esta cuestión, Rosenthal argumenta que los requisitos en estos seres son más débiles, dado que el requisito de que el pensamiento de orden superior refiera a uno mismo de forma esencialmente indéxica se debe a que queremos excluir formas irrelevantes de referirse a alguien que, de hecho, es uno mismo. Así, por el hecho de pensar que MAS está en cierto estado, no tengo una experiencia consciente a pesar de que, de hecho, MAS soy yo. Rosenthal afirma que los bebés y los animales no humanos no tienen formas no irrelevantes de referirse a sí mismos. Ellos tienen la capacidad de distinguirse de cualquier otra cosa y, por tanto, pueden referirse a ellos mismos en el pensamiento, “pero sus HOT no requieren indéxico esencial, pues distinguirse de los demás da la única forma que tienen de referirse a ellos mismos”[16] (Rosenthal, 2011b, p. 34). Esta afirmación de Rosenthal me parece totalmente gratuita y creo que hay buenas razones para creer que es falsa. Al ver mi cara en el espejo, yo puedo creer que la persona que estoy viendo está enfada sin por ello sentir enfado. Las teorías de orden superior explican este hecho, pues la atribución de un cierto estado mental (enfado) la estoy haciendo a alguien que simplemente ocurre que soy yo mismo. Sin embargo, la teoría requiere, para que haya consciencia, que lo haga de forma esencialmente indéxica. El problema es que, al igual que yo puedo hacer esa atribución sin por ello tener la experiencia correspondiente, lo mismo, *pace* Rosenthal, ocurre en animales. La capacidad de atribuir estados mentales a otros ha sido recientemente demostrada, por ejemplo, en córvidos y cánidos (Bugnyar y Heinrich, 2006; Hare y Tomasello, 2005; Stulp et al., 2009; Udell et al., 2008). Es de esperar que al verse presentados con su propia imagen, puedan atribuirle cierto estado mental al animal que están viendo. Por lo tanto, no parece cierto que no tengan otras formas de referirse a ellos mismos y, consecuentemente, el problema de fijar la referencia del concepto YO que los pensamientos de orden superior requieren persiste en los animales.

[16] “But their HOTs do not require the essential indexical, since distinguishing themselves from everything else provides the only way they have to refer to themselves.”

El problema de la inmunidad

El último problema que quiero presentar es el de la inmunidad por fallo en la identificación de la primera persona (IEM por sus siglas en inglés).

Una vez aceptamos que el contenido de la experiencia es *de se*, cabe hacer una distinción entre dos formas en las que uno se representa a sí mismo, ambas corresponden a lo que Shoemaker (1968), siguiendo a Wittgenstein (1958), ha llamado representación como sujeto y como objeto. La mejor forma de ilustrar esta diferencia es utilizando un ejemplo. Consideremos dos estados mentales como puedan ser la creencia de que yo tengo tengo dolor de muelas —formada con base en mi dolor de muelas— y la creencia de que yo tengo el brazo enyesado —formada tras ver una imagen en el espejo—. De acuerdo con Shoemaker, esta última creencia ilustra una representación como objeto, pues requiere la identificación de un determinado objeto, una persona en este caso, y por tanto la posibilidad de error en tal identificación. En este caso, cabría preguntarme si estoy seguro de ser yo el que tiene el brazo enyesado. En contraste, como señala Wittgenstein (1958, pp. 66-67) “no tiene caso preguntar por el reconocimiento de una persona cuando digo que tengo dolor de muelas. Preguntar '¿estás seguro de que eres tú el que tiene dolor?' carece de de sentido”.[17] Shoemaker afirma en esa dirección que la representación como sujeto ilustrada por este caso es inmune al error por fallo en la identificación relativo al uso del pronombre de primera persona (IEM). Esto es, “no puede ocurrir que esté equivocado al decir 'tengo dolor' porque, aunque sé que alguien tiene dolor, me equivoco al creer que esa persona soy yo” (ibíd., p. 567). La atribución de estados mentales al tener la experiencia resultan los casos paradigmáticos de IEM.

Shoemaker va un paso más allá y argumenta que no toda autoatribución puede estar fundamentada en la identificación de un objeto como uno mismo. Para él, identificar algo S como uno mismo requiere que, o bien encontremos algo que es verdadero de S que uno independientemente sepa que es verdadero de uno mismo, o bien encontrar que S está en una relación con uno mismo en la que solo uno mismo puede estar con uno. Sin embargo, ambas opciones exigirían a su vez que haya alguna propiedad o relación que uno ya se haya atribuido a sí mismo y que no esté fundamentada en la identificación en cuestión. De este modo, haría falta otra identificación. Así que, si queremos evitar un regreso vicioso, no toda atribución a uno mismo puede estar fundamentada en una identificación. Es más, dado que la identificación de un objeto como uno mismo está acompañada de la posibilidad de error en la identificación, la representación como sujeto no puede depender de la identificación si ha de ser IEM.

---

[17] “there is no question of recognizing a person when I say I have toothache. To ask 'are you sure it is you who have pains?' would be nonsensical.”

Una vez entendido el fenómeno de IEM, podemos regresar a la propuesta de Rosenthal. Como vimos, era la disposición a identificarse como el pensador del pensamiento superior lo que garantizaba la referencia del mismo en la forma requerida. Sin embargo, al estar fundamentada en una identificación, no hay IEM. Rosenthal acepta esa conclusión y a favor de su postura presenta un caso en el que la inmunidad al error parece fallar.

Rosenthal comienza por reconocer que, desde el punto de vista subjetivo, no parece haber nada que medie entre aquel que se percata del estado y el estado mismo. No obstante, señala que no hay razón para creer que la ausencia de mediación sea real, dado que el hecho de que no haya mediación no es necesario para explicar la correspondiente apariencia, pues para ello bastaría que no nos percatáramos de los procesos que median. Aun así, Rosenthal reconoce que si el fallo en la identificación fuera posible, deberíamos ser capaces de describir un escenario en que esa situación resultara creíble. Y él considera que lo hay: los casos de Trastorno de Identidad Disociativo (TID).

El TID está descrito en el DSM IV como la existencia de dos o más personalidades en un individuo. Cada una de las identidades presenta su propio patrón de percepción y acción. Al igual que ocurre con individuos ordinarios, las personalidades que comparten cuerpo en el caso del TID presentan una integración y conexión entre sus propias experiencias, memorias y pensamientos que permite establecer un patrón de comportamiento. Para diagnosticar TID, al menos dos de estas personalidades deben tomar el control del comportamiento del individuo de forma rutinaria y estar asociadas con un grado de pérdida de memoria más allá de lo que resulta normal. Rosenthal señala que, en ocasiones, una personalidad puede percatarse de forma subjetivamente inmediata de los estados mentales de otra personalidad y puede, en dicho caso, repudiar el hecho de estar en cierto estado y atribuírselo a otra personalidad. En este caso, una personalidad tendría un HOT suficientemente bien integrado en su vida mental que describiría a otra personalidad como estando en cierto estado. Rosenthal considera que tenemos un criterio lo suficientemente robusto como para determinar a quién pertenece un estado mental en función del nivel de integración con otros estados de la personalidad. Por ende, la siguiente posibilidad parece viable. Llamemos Taylor y Jack a las dos personalidades que comparten cuerpo. Consideremos un estado mental M y aceptemos, por mor de la argumentación, que tenemos buenas razones para creer que M pertenece a Taylor. La sugerencia de Rosenthal es que puede ocurrir que Jack se autoatribuya la propiedad de estar en M cuando en realidad es Taylor el que está en ese estado. En este caso Jack se percataría de sí mismo como estando en M y lo que está haciendo es identificarse erróneamente con Taylor, que es quien en realidad está en M. Si esta descripción es congruente, efectivamente mostraría una situación en la que falla la IEM, al menos como la presenta Shoemaker, pues Jack cree que él está en M, porque alguien está en M y Jack cree que la persona que está en M es él

mismo.

Dejando a un lado la controversia que el TID suscita, el ejemplo de Rosenthal sugiere que puede haber casos de creencias *de se* en las que hay un fallo en la identificación de la primera persona. Obviamente, nadie ha negado eso, como ilustran los casos en los que el sujeto se representa como objeto —como ocurría en el caso de la creencia que uno se forma tras verse en el espejo—. Pero los ejemplos usados por Wittgenstein y Shoemaker muestran que los casos paradigmáticos de IEM son casos de creencias acerca de las experiencias conscientes formadas de forma inmediata. El ejemplo de Rosenthal no muestra que IEM falle en esos casos y que, si yo me percato de mí mismo como viendo una manzana roja o teniendo dolor, tenga sentido el equivocarme acerca de quién es el que está teniendo la experiencia de sentir dolor —en el ejemplo anterior, Jack se autoatribuye el estado y el que, de hecho, tiene la experiencia consciente, pese a no estar en el estado de primer orden.

Si mis argumentos en esta sección son correctos, las teorías que dependan del uso de conceptos para explicar el carácter subjetivo de la conciencia parecen no tener forma de garantizar la referencia del concepto YO. Por ello, en la siguiente sección presentaré una propuesta alternativa según la cual la experiencia consciente representa al propio sujeto de manera no conceptual y satisface IEM.[18]

4. Contenido no conceptual. Representacionismo centrado en uno mismo

La aproximación representacionista pretende dar cuanta de la experiencia consciente en dos

[18] De esta forma podemos fundamentar la posesión del concepto YO en la experiencia consciente, una idea que ya propone Recanati (2007, 2012). Recanati considera que nuestra experiencia representa *implícitamente* al sujeto que tiene la experiencia (cf. Perry, 1986) . En línea con Searle (1983), mantiene que es el modo en que la experiencia representa lo que determina la relación que se da (si no hay fallo en la representación) entre el sujeto de la experiencia y el objeto primario de la experiencia: al tener una experiencia visual uno “tiene derecho a autoatribuirse la propiedad de estar enfrente de y estar causalmente afectado por lo que su experiencia visual representa [el objeto primario de la experiencia]”. Cuando el contenido es *de se* explícito, el sujeto se hace explícito. El mecanismo que permite hacer explícito lo que ya estaba presente en la semántica, aunque de forma implícita, lo llama Recanati (2000) 'reflexión'. Dado que este proceso de reflexión no requiere inferencia alguna ni nueva evidencia, cuando nos formamos un pensamiento de esta forma, este hereda la IEM de la experiencia que está garantizada por el modo experiencial.
La propuesta de Recanati acomoda mediante la distinción implícito/explícito la diferencia que hay entre el sujeto y el objeto primario de la experiencia. Además, ofrece una explicación de la inmunidad al error de ciertos juicios. Su propuesta sufre, no obstante, de un problema fundamental, pues parte de la naturaleza de la experiencia para explicar el fenómeno en cuestión, y con ello fundamenta algo que resulta misterioso en algo más misterioso aún. Este problema es más agudo en el marco de este trabajo donde, recordemos, el objetivo era dar cuenta del carácter subjetivo de la experiencia en términos representacionales.

pasos. En primer lugar, busca entender qué tipo de información ofrecen nuestros estados conscientes. En este sentido, según hemos visto, nuestra experiencia nos dice algo no solo acerca de su objeto primario —digamos la manzana que estoy viendo en el caso de mi experiencia visual—, o acerca de cierto estado, sino también acerca del propio sujeto que tiene la experiencia como tal. La experiencia consciente representa el mundo desde la perspectiva de la primera persona y, en consonancia, en el marco considerado, involucra una representación *de se*. El segundo paso para determinar qué estados son conscientes consiste en determinar qué es lo que se requiere de un estado para que nos dé ese tipo de información de primera persona: ¿qué hace falta para tener representaciones no conceptuales *de se*?

Regresando al marco presentado en la sección 2, vimos que era posible caracterizar los contenidos no indéxicos como colecciones de mundos posibles. Tales particiones quedaban determinadas mediante la atribución de propiedades. Si la consciencia involucrara representaciones no indéxicas, tendríamos que determinar qué es lo que se demanda de un sistema cognitivo para hacer tales atribuciones. En la literatura encontramos diversas teorías en este sentido. Estas teorías apelan comúnmente a la noción de *indicación* entendida como alguna forma de covarianza entre el vehículo de la representación y su objeto. La idea que subyace es que un estado mental representa aquello que lo causa en *condiciones normales* (para acomodar la idea de fallo en la representación, no queremos que un estado mental represente todo aquello que lo causa). Ahora bien, "condiciones normales" es una noción normativa que no resulta aceptable en un marco naturalista. Las distintas teorías tratan de descargar esta noción en términos compatibles con el naturalismo. La ruta más popular para solventar este problema apela a la *función propia* del estado (Millikan, 1984), resultando que un sistema representacional es aquel que tiene la función propia de indicar que tal-y-cual es el caso, siendo la función propia la que determina cuál, entre todas las cosas con las que el estado covaría, es aquella que representa.[19] Por lo tanto, atribuir a un objeto la propiedad de tomar café es cuestión de estar en un estado que tenga la función de indicar que ese objeto toma café.

Pero nuestra experiencia consciente involucra una representación *de se* que determina particiones en mundos centrados. Tales mundos centrados quedan determinados, siguiendo a Lewis, por la autoatribución de propiedades. Entender la atribución de propiedades no basta para entender la autoatribución: MAS le puede atribuir la propiedad de tomar café a MAS sin que MAS se autoatribuya la propiedad de tomar café (yo puedo creer que MAS está tomando café sin creer que yo mismo estoy tomando café). No basta con un estado que tenga la función de indicar que tal-y-cual es el caso, sino que hace falta

[19] Este ejemplo tan simplista tiene únicamente el objetivo de mostrar el espíritu de las teorías teleológicas. Para un mayor detalle y elaboración véase, por ejemplo, Artiga, 2016; Dretske, 1988; Martínez, 2013; Millikan 1984, 1989; Mossio et al., 2009; Neander, 1991; Schroeder, 2004.

que un estado que tenga la función de indicar que tal-y-cual es el caso *para el mismo individuo que está representando*. Pero, ¿qué tipo de entidad es un individuo? Para responder a esta pregunta, debemos pensar en el tipo de entidades que son susceptibles de tener experiencias. La respuesta más natural en un marco naturalista apunta a los organismos. Pero, ¿qué es un organismo?

Para responder a esta pregunta debemos trasladarnos al campo de la biología donde una respuesta ampliamente aceptada sugiere que son sistemas automantenidos. La noción de sistema automantenido tiene una larga tradición en filosofía desde Aristóteles (Godfrey-Smith, 1994; McLaughlin, 2001). Aunque dentro de la ciencia contemporánea esta noción creció en el seno de la cibernética, recientemente muchos científicos han favorecido aproximaciones basadas en la termodinámica en detrimento de una aproximación cibernética, especialmente siguiendo el trabajo de Ilya Prigogine sobre *estructuras disipativas* —que serían la mínima expresión de sistema automantenido—. De forma sencilla, un sistema automantenido es aquel que se mantiene dentro de unos rangos de funcionamiento que están lejos del equilibrio termódinamico (baja entropía) al que todo sistema tiende. Resulta relativamente sencillo ilustrar esta idea mediante un ejemplo. Pensemos en un montón de arena. Podemos tomar un puñado e ir dejando los granos de arena caer. Como resultado final continuaremos teniendo un montón de arena: hay una multitud de formas en que los granos de arena pueden reorganizarse de tal forma que continuemos teniendo un montón de arena. Ahora comparemos esa estructura con un castillo de arena: hay muchas menos formas en las que los granos pueden reorganizarse. La entropía mide este hecho, esto es, en cuántas formas los constituyentes pueden estar organizados al nivel micro (los granos de arena) de forma consciente con la estructura a nivel macro (el castillo de arena). Formalmente, la entropía es una medida logarítmica del número de estados que tienen una probabilidad significativa de ser ocupados. El montón de arena es una organización con una entropía mucho mayor que la del castillo. Este ejemplo ilustra la segunda ley de la termodinámica que señala que la entropía de un sistema aumenta o permanece constante —tan solo consideremos la tendencia que tiene el castillo a acabar como el montón de arena.

Los organismos son sistemas con baja entropía, sistemas que trabajan lejos del equilibrio termodinámico: aunque admiten cierta variación interna, esa variación tiene que ser mantenida dentro de unos valores restringidos. Los organismos tienen la capacidad de mantener esa condición interna estable compensando los cambios externos e internos mediante el intercambio de materia y energía con el exterior, lo que se conoce como 'homeostasis'. Resulta, por tanto, razonable fundamentar una teoría de la autoatribución en los mecanismos responsables de tal estabilidad y que dan cuenta de lo que el organismo es. En esta dirección, Antonio Damasio (2000, 2010) hace una interesante propuesta al presentar su noción de 'proto-self'. El proto-self es una colección integrada de patrones

neuronales distribuidos por todo el cerebro que representan los aspectos más estables de las estructuras físicas del organismo y que, además, están involucrados en el proceso de regular el estado del organismo. El proto-self es una estructura neuronal que no solo se encarga de regular los procesos homeostáticos del organismo, monitorizando el entorno extracelular (*internal milieu*), sino también, por ejemplo, la estructura músculo-esquelética y la musculatura visceral. Damasio ha argumentado que este conjunto de patrones neuronales son constitutivos del mecanismo neuronal que subyace a nuestra experiencia consciente, y en trabajos anteriores he apelado a estas estructuras para explicar las condiciones que permiten que nuestra experiencia tenga un contenido *de se* (Sebastián, 2012, 2018a). La idea consiste en explicar cómo un organismo se autoatribuye la propiedad de estar en cierto estado mediante la interacción que se da entre este proto-self y lo que he llamado 'estado proto-cualitativo'. Si el proto-self regula el medio interno y tiende a mantener la estabilidad que el sistema automantenido requiere, el estado proto-cualitativo en una estructura neuronal que tiene la función de indicar que tal-y-cual es el caso, por ejemplo, la función de indicar que hay un objeto rojo. El estado proto-cualitativo es el responsable de las diferencias entre experiencias conscientes, por ejemplo, de la diferencia que hay entre oler una rosa y escuchar una melodía. Ahora bien, estar en un estado proto-cualitativo no basta para tener una experiencia consciente. Estos estados no tienen el tipo de contenido requerido. Tampoco el proto-self lo tiene. El estado consciente es el estado resultante de la interacción entre el estado proto-cualitativo y el proto-self. El complejo resultante de esa interacción tendrá la función de indicar que el mismísimo organismo que el proto-self regula está siendo afectado por el objeto que el estado proto-cualitativo representa. De esta forma, los estados conscientes están constituidos por estas dos estructuras y por los mecanismos que se precisan para la interacción requerida. Veamos un ejemplo para entenderlo con mayor claridad.

Al mirar la manzana roja frente a mí, tengo una experiencia consciente. Mi sistema visual genera una representación de las propiedades de la manzana que constituyen lo que he llamado el estado proto-cualitativo.[20] Este es aún un estado inconsciente, no hay algo que es para mí estar simplemente en el estado proto-cualitativo. Además, mi organismo tiene un subsistema, el proto-self, que monitoriza y controla la homeodinámica del organismo y otros invariantes como hemos visto.[21] Esta representación se ve alterada por el procesamiento de la manzana: hay cambios, por ejemplo, en la retina o en los músculos que controlan el globo ocular, así como cambios en los músculos lisos de las vísceras que

[20] El estado proto-cualitativo estará implementado en distintas áreas del neocórtex dependiendo de la modalidad. Por ejemplo, en el caso de la visión, en el córtex visual.

[21] El proto-self, a nivel neuronal, corresponde con varios núcleos del tallo cerebral, que incluyen el tegmento, el hipotálamo, la corteza insular y S2 (Damasio 2000; 2010). Damasio incluye, además, el giro cingulado motivado por los casos de mutismo akinético, una condición en la que parece fallar la conciencia. Sin embargo, casos de pacientes que se han recuperado han reportado que recuerdan las preguntas que les hacía el doctor (Laureys y Tononi, 2008, p. 385). Por esta razón resulta razonable pensar que giro cingulado forma parte de las estructuras de interacción.

corresponden a respuestas emocionales (muchas de ellas innatas) y que son registrados por el proto-self. Esta interacción explica que el contenido del estado complejo que confirman el proto-self y el estado proto-cualitativo sea *de se*.[22] De acuerdo con la familia de teorías que estamos considerando, el contenido del estado depende de aquello que el estado tiene la función propia de indicar. La función propia de este complejo no es ni indicar que hay un objeto con tal-y-cual reflectancia de superficie, ni indicar que tal-y-cual estado corporal se da, sino indicar que el mismo sistema que está realizando la representación (dado que el proto-self es responsable de mantener el organismo como lo que esencialmente es, es parte de ese complejo) está en cierto estado perceptivo. Cuando un organismo, ORG, se encuentra en este estado complejo, no atribuye simplemente una propiedad a ORG, sino que se la autoatribuye. ORG atribuye la propiedad de estar en cierto estado a sí mismo, al mismísimo organismo que el proto-self regula, "bajo la relación de identidad" (Lewis, 1979, p. 483). Por tanto, el complejo representa que el organismo mismo está presentado con el objeto que el estado proto-cualitativo representa. El sistema atribuye la propiedad al mismo sistema que está haciendo la representación sin necesidad de que tal atribución sea mediada por una identificación y, por tanto, tal autoatribución es inmune al error por fallo en la identificación.

5. Conclusión

Una teoría del carácter subjetivo de la experiencia pretende explicar la diferencia que existe entre los estados que son fenoménicamente conscientes y los que no. El principio de transitividad ha sido propuesto como una vía en esa dirección. En el marco representacionista considerado, se sigue que las experiencias conscientes son representaciones *de se*.

Las teorías de pensamiento de orden superior han tratado de dar cuenta de este hecho por medio de un pensamiento que incluye el uso del concepto YO. Sin embargo, he argumentado que estas teorías resultan insatisfactorias al no poder justificar el principio de transitividad —el componente de primera persona no puede ser fenoménicamente dado— y no ser capaz explicar el mecanismo de referencia del concepto YO, que es necesario para la formación del pensamiento de orden superior, del cual depende la experiencia.

Finalmente, he esbozado una teoría que no está afectada por estos problemas. Para ello, y sobre la idea lewisiana de autoatribución, he explicado cómo un sistema se puede

[22] Finalmente, podemos especular sobre las estructuras cerebrales responsables de la interacción entre el proto-self y el estado proto-cualitativo. Estas áreas deben tener ramificaciones a las áreas correspondientes a ambos subestados. Además, debería haber evidencia empírica que vinculase la actividad de estas estructuras con la experiencia consciente. Ejemplos de estas áreas son el colículo superior, el giro cingulado, el tálamo, el córtex parietal medio (véase Damasio 2000, 2010; Laurey y Tononi, 2008 para una discusión de la evidencia empírica que vincula la actividad de estas áreas y nuestra experiencia consciente).

autoatribuir la propiedad de estar en cierto estado sin mediar el uso de conceptos. Este tipo de autoatribución, al no requerir una identificación, resulta inmune al error por fallo en la identificación y puede, además, fundamentar otro tipo de identificaciones de uno mismo —representación como objeto— y dar soporte a una teoría que fundamente la posesión del concepto YO en nuestra experiencia consciente.[23]

Bibliografía

Artiga, M.: 2016, Teleosemantic modeling of cognitive representations. *Biology and Philosophy* 31 (4), 483-505.

Amstrong, D.: 1968, *A Materialist Theory of the Mind*, London: Routledge.

Block, N.: 2007, Consciousness, accessibility, and the mesh between psychology and neuroscience, *Behavioral and Brain Sciences* 30, 481-548.

Brown, R.: 2015, The HOROR theory of phenomenal consciousness, *Philosophical Studies* 172, 1783-1794.

Bugnyar, T. y Heinrich, B.: 2006, Pilfering ravens, corvus corax, adjust their behavior to social context and identity of competitors, *Animal Cognition* 9, 369-376.

Burge, T.: 2007, *Foundations of Mind (Philosophical Essays)*, Oxford University Press, USA.

Byrne, A.: 2004, What phenomenal consciousness is like, en R. Gennaro (ed.), *Higher-Order Theories of Consciousness: An Anthology*, John Benjamins Publishing Company.

Carruthers, P.: 2000, *Phenomenal Consciousness: a naturalistic theory*, Cambridge: Cambridge University Press.

Castañeda, H.-N.: 1966, 'he': A study in the logic of self-consciousness, *Ratio* 8, 130-157.

Caston, V.: 2002, Aristotle on consciousness, *Mind* 111(444), 751-815.

Damasio, A.: 2000, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, 1 edn, Mariner Books.

Damasio, A.: 2010, *Self Comes to Mind: Constructing the Conscious Brain*, 1 edn, Pantheon.

Dretske, F.: 1988, *Explaining Behavior: Reasons in a World of Causes*, MIT Press.

[23] Me gustaría agradecer a Angélica Pena Martínez por sus comentarios, así como a los asistentes al ciclo de conferencias dentro del proyecto Experiencia Visual en la UAM donde presenté partes de este trabajo, especialmente a Eduardo Berúmen, Ignacio Cervieri, Maximiliano Martínez, Álvaro Peláez y Laura Pérez. Agradezco también a Víctor Sánchez Alonso por su apoyo en la corrección de estilo. Investigación realizada gracias al programa UNAM-DGAPA-PAPIIT IG400219 e IA400520.

Egan, A.: 2006a, Appearance properties?, *Noûs* 40(3), 495-521.

Egan, A.: 2006b, Secondary qualities and self-location, *Philosophy and Phenomenological Research* 72(1), 97-119.

Gallagher, S. y Zahavi, D.: 2006, Phenomenological approaches to self-consciousness, http://plato.stanford.edu/. URL: http://plato.stanford.edu/

Gennaro, R.: 1996, *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, John Benjamins.

Gennaro, R.: 2012, *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*, MIT Press.

Godfrey-Smith, P.: 1994, A modern history theory of functions, *Noûs* 28(3), 344-362.

Hare, B. y Tomasello, M.: 2005, Human-like social skills in dogs?, T*rends in Cognitive Sciences* 9(9):439-444.

Hume, D.:1739. *A Treatise of Human Nature*. Oxford University Press, USA.

Kriegel, U.: 2003, Consciousness, higher-order content, and the individuation of vehicles, *Synthese* 134(3), 477-504.

Kriegel, U.: 2009a, Self-representationalism and phenomenology, *Philosophical Studies* 143, 357-381.

Kriegel, U.: 2009b, *Subjective Consciousness: A Self-Representational Theory*, Oxford University Press, USA.

Lau, H. y Rosenthal, D.: 2011, Empirical support for higher-order theories of conscious awareness, *Trends in Cognitive Sciences* 15(8), 365-373.

Laureys, S. y Tononi, G.: 2008, *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, 1 edn, Academic Press.

Levine, J.: 2001, *Purple Haze: The Puzzle of Consciousness*, Oxford University Press.

Lewis, D.: 1979, Attitudes de dicto and de se, *Philosophical Review* 88(4), 513-543.

Lycan, W. G.: 1996, *Consciousness and Experience*, The MIT Press.

Martínez, M.: 2013, Teleosemantics and indeterminacy. *Dialectica* 67 (4), 427-453.

McLaughlin, P.: 2001, *What Functions Explain. Functional Explanation and Selfreproducing Systems*, Cambridge: Cambridge University Press.

Millikan, R. G.: 1984, *Language, Thought and Other Biological Categories*, MIT Press.

Millikan, R. G.: 1989, Biosemantics, *Journal of Philosophy* 86, 281-297.

Mossio, M., Saborido, C. y Moreno, A.: 2009, An organizational account of biological functions, *British Journal for the Philosophy of Science* 60(4), 813-841.

Nagel, T.: 1974/2002, What is it like to be a bat?, en D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.

Neander, K.: 1991, Functions as selected effects: The conceptual analyst's defense, *Philosophy of Science* 58(2), 168-184.

Papineau, D.: 1993, *Philosophical Naturalism*, Blackwell.

Peacocke, C.: 1995, *A Study of Concepts*, The MIT Press.

Perry, J.: 1979, The problem of the essential indexical, *Noûs* 13, 3-21.

Perry, J.: 1986, Thought without representation, *Proceedings of the Aristotelian Society* 60, 137-151.

Recanati, F.: 2000, *Oratio Obliqua, Oratio Recta. An Essay on Metarepresentation.*, Cambridge, MA: MIT Press/Bradford Book.

Recanati, F.: 2007, *Perspectival Thought: A Plea for (Moderate) Relativism*, Oxford:Oxford University Press.

Recanati, F.: 2012, Immunity to error through misidentication: What it is and where it comes from, en S. Prosser and F. Recanati (eds), *Immunity to error thorugh misidentication: New Essays*, Cambridge: Cambridge University Press.

Rosenthal, D.: 1997, A theory of consciousness, en N. Block, O. J. Flanagan y G. Guzeldere (eds), The Nature of Consciousness, Mit Press.

Rosenthal, D.: 2005, Consciousness and mind, Oxford University Press.

Rosenthal, D.: 2011, Awareness and identication of self, en J. Liu y J. Perry (eds), *Consciousness and the Self: New Essays*, Cambridge: Cambridge University Press.

Rosenthal, D. y Weisberg, J.: 2008, Higher-order theories of consciousness, Scholarpedia 3(5), 4407.

Schroeder, T.: 2004, New norms for teleosemantics, en H. Clapin (ed.), *Representation in Mind*, Elsevier.

Searle, J.: 1983, *Intentionality*, cambridge: Cambridge University Press.

Sebastian, M. Á.: 2012, Experiential awareness: Do you prefer "it" to "me"?, *Philosophical Topics* 40(2), 155-177.

Sebastián M. Á.: 2018a, Embodied Appearance Properties and Subjectivity. *Adaptative Behaviour*.

Sebastián M. Á.: 2018b, Drop it like it's HOT: a vicious regress for higher-order theories. *Philosophical Studies* 176 (6), 1563-1572

Sebastián, M. Á.: 2020a. Perspectival Self-Consciousness and Ego-Dissolution: an analysis of (some) altered states of consciousness. *Philosophy and the Mind Sciences* 1(I):9

Sebastián, M. Á.: 2020b. Subjective Character, the Ego and De Se Representation: Phenomenological, Metaphysical and Representational Considerations on Pre-reflective Self-awareness *ProtoSociology* 36.

Sebastián, M. Á.: MS. First-Person Perspective in Experience: Core de se representations as an explanation of subjective character.

Shoemaker, S.: 1968, Self-reference and self-awareness, *The Journal of Philosophy* 65, 555-567.

Stalnaker, R.: 1999, *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford University Press, USA.

Stalnaker, R.: 2008, *Our Knowledge of the Internal World*, Oxford University Press.

Stulp, G., Emery, N., Verhulst, S. y Clayton, N.: 2009, Western scrub-jays conceal auditory information when competitors can hear but cannot see., *Biology Letters* 5, 583-585.

Udell, M., Dorey, N. y Wynne, C.: 2008, Wolves outperform dogs in following human social cues, *Animal Behavior* 76, 1767-1773.

Weisberg, J.: 2011, Misrepresenting consciousness. *Philosophical studies*, 154(3):409-433.

Williford, K.: 2006. The self-representational structure of consciousness. En U. Kriegel y K. Williford (eds), *Self-Representational Approaches to Consciousness*. The MIT Press.

Wittgenstein, L.: 1958, *The Blue and Brown Books*, New York: Oxford