

(Versión para imprenta)

EL SUJETO ENUNCIADO*
Carlos J. Moya. Universidad de Valencia

Tras los términos 'sujeto' y 'subjetividad' se encierra un amplio conjunto de cuestiones, de extrema dificultad, relacionadas entre sí de modo sumamente intrincado. En el presente trabajo comenzaré abordando la cuestión de la subjetividad del modo más directo posible, tratando de obtener una caracterización preliminar de esta noción que sea capaz de introducirnos de modo natural en algunos de los vericuetos que parten de ella. El cuerpo central del trabajo consistirá en una exposición crítica del tratamiento de la subjetividad en la filosofía analítica de la segunda mitad de nuestro siglo, en particular en el marco del conductismo lógico, la teoría de la identidad y el funcionalismo. Este cuerpo central irá precedido de una breve presentación de la muy influyente concepción cartesiana y seguido de una exposición de la reacción neocartesiana y de las respuestas a la misma. Finalmente, ofreceremos, de modo tentativo, una hipótesis alternativa propia para el tratamiento de algunos de los problemas vinculados a la subjetividad.

1. La subjetividad.

¿Qué es la subjetividad? ¿Qué es ser un sujeto? En primer lugar, no creemos que ser un sujeto sea una cuestión de todo o nada. Hay grados de subjetividad, lo que no excluye la existencia de un límite que separa a los seres humanos de otros seres. Ser un sujeto requiere, como mínimo, cierto grado de profundidad o complejidad psicológica. En primer lugar, un sujeto ha de poseer un punto de vista sobre la realidad en términos del cual interactúa con ésta. Este punto de vista incluye, de modo especial, la posesión de creencias y deseos, o de estados análogos a éstos, caracterizados por la tensión y discrepancia potencial entre ellos y la realidad de las cosas. Podemos denominar este conjunto de estados el componente o aspecto *intencional* de la subjetividad. En segundo lugar, un sujeto ha de poseer cierto nivel de sensibilidad a los cambios en el entorno y en sí mismo. Sensaciones y sentimientos son las manifestaciones más claras de este segundo aspecto de la subjetividad, que podemos denominar *sensitivo*. En tercer lugar, un sujeto ha de ser capaz de desarrollar una actividad autónoma, a partir de su propia elección y decisión. Decisión y acción intencional integran en especial este tercer aspecto de la subjetividad, que podemos denominar *práctico*.

Sin embargo, la mera presencia de estos componentes en un determinado ser no nos ofrece todavía la plenitud del concepto de sujeto que aplicamos paradigmáticamente a los seres humanos y que nos lleva a considerarlos responsables de sus acciones. La responsabilidad por las propias acciones es el carácter que distingue a un sujeto personal, a una persona.¹ La mera posesión de los aspectos indicados no nos permite todavía la atribución de esta cualidad a un determinado ser. Así,

* La investigación conducente a este trabajo ha sido financiada por la DGICYT, en el marco del proyecto PB93-1049-C03-02. Agradezco a esta institución su generosa ayuda y estímulo. Agradezco asimismo al profesor Manuel Cruz su amable invitación a colaborar en el presente volumen, y a mis compañeros T. Grimaltos, J. Marrades y V. Sanfélix sus valiosas observaciones y sugerencias.

¹ J. Marrades me sugiere la conveniencia de introducir un aspecto social en la caracterización de la subjetividad: ser sujeto es ser capaz de reconocimiento recíproco. Esta observación, de inspiración hegeliana, se halla no obstante recogida, en parte, en la última sección del presente trabajo.

consideramos natural atribuir a determinados animales, sobre la base de la complejidad de su comportamiento y de sus reacciones, creencias, expectativas y deseos, sensaciones y sentimientos, así como decisiones y acciones intencionales, o al menos estados análogos a éstos, si decidimos reservar estos últimos a los seres humanos. Sin embargo, no atribuimos a los animales responsabilidad por sus acciones. En un importante artículo,² Harry Frankfurt sostuvo que, para ser un sujeto responsable de las propias acciones, una persona, es necesario adoptar actitudes volitivas de segundo orden, actitudes estimativas hacia los propios deseos, intenciones y decisiones. Es claro que, para adoptar estas actitudes de segundo orden, un sujeto ha de tener cierto grado de conocimiento de sus propias actitudes de primer orden, de sus deseos, intenciones y decisiones. Las actitudes de segundo orden presuponen, pues, cierto grado de autoconocimiento en el plano mental. Akeel Bilgrami, por su parte, ha establecido también una relación de dependencia entre la responsabilidad por las propias acciones y el autoconocimiento.³ Un supuesto de la responsabilidad por las propias acciones es que sepamos lo que estamos haciendo intencionalmente y, con ello, que conozcamos nuestras propias creencias, deseos e intenciones. Si esto es correcto, la subjetividad, en el grado en que la atribuimos a los seres humanos adultos, requiere, pues, no sólo la posesión de estados mentales, sino cierto conocimiento de los mismos por parte del sujeto que los posee. Tal vez, pues, es la carencia de esta capacidad reflexiva, de autoconocimiento o autoconciencia, lo que nos impide atribuir responsabilidad a los animales, por más que les atribuyamos propiedades mentales.

Es notorio que el autoconocimiento ha generado una amplia discusión a lo largo de la historia de la filosofía, y no es extraño que haya sido así. El autoconocimiento, cuya manifestación más característica es el uso competente de verbos y predicados psicológicos en primera persona (un ejemplo emblemático es el cartesiano *Cogito*), constituye un enigma, de cuya correcta resolución depende en gran medida la comprensión rigurosa de la naturaleza de la subjetividad. Descartes fue el primero en advertir la singularidad epistemológica de estos enunciados,⁴ haciendo de ellos un bastión de resistencia frente al escepticismo absoluto y el pilar central de toda su metafísica. Aunque sin duda Descartes sobrevaloró la inmunidad de estos enunciados al error, es, no obstante, cierto que, en circunstancias normales, exhiben una presunción de verdad que está ausente de los enunciados basados en observaciones o inferencias. Este rasgo del autoconocimiento resulta enigmático: sin esfuerzo aparente, un sujeto es capaz de emitir enunciados sobre sus propias creencias, deseos, sensaciones, sentimientos, intenciones y decisiones para los cuales el error resulta sólo excepcional.

Esta presunción generalizada de verdad para la atribución de estados mentales a uno mismo se pierde cuando se atribuyen a otros. Los enunciados psicológicos en tercera (o segunda) persona no gozan de una especial autoridad o inmunidad al error. La atribución de estados mentales a otras personas descansa normalmente en la observación de su comportamiento, lingüístico y no

² 'Freedom of the Will and the Concept of a Person', en G. Watson (ed.), *Free Will*, Oxford University Press, Oxford, 1982, pp. 81-95.

³ A. Bilgrami, *Belief and Meaning*, Blackwell, Oxford/Cambridge Mass., 1992, p. 250.

⁴ Al menos en la medida en que pone de manifiesto la importancia de su formulación en primera persona. Estrictamente, la primacía histórica correspondería a Agustín de Hipona, como me ha recordado V. Sanfélix. La profesora Anscombe subraya esta primacía agustiniana en su 'The First Person', en S. Guttenplan (ed.), *Mind and Language*, Oxford University Press, Oxford, 1975, pp. 45-65.

lingüístico, y en la inferencia a partir de él, y está sometido a las dificultades características de otros juicios empíricos y, tal vez, a otras adicionales.

Así, como condición de la subjetividad plena, el autoconocimiento no es cualquier clase de conocimiento de uno mismo, sino que involucra una especial intimidad o inmediatez. No conocemos nuestras creencias, sensaciones e intenciones sobre la base de los criterios objetivos, especialmente conductuales, que otros usan para atribuirnos tales estados. La perspectiva de uno mismo sobre su propia mente es, pues, distinta de la que tiene sobre la mente de los demás. Llamemos a aquélla la perspectiva de la primera persona y a ésta la perspectiva de la tercera persona, por seguir una terminología usual.

Así, pues, los enunciados psicológicos se distinguen de cualesquiera otros por la sistemática asimetría que presenta su atribución en primera y en tercera persona. Como Davidson señala, esta asimetría no tiene parangón en otros tipos de conceptos: 'Muchos conceptos pueden ser aplicados sobre la base de criterios múltiples, pero no hay conceptos en cuya adscripción los sujetos *tengan* que usar criterios diferentes en ocasiones particulares.'⁵ Strawson, por su parte, considera que esta asimetría en la atribución resulta esencial a los predicados mentales: '... Es esencial al carácter de estos predicados el hecho de que tienen usos atributivos en primera y en tercera persona, que pueden ser a la vez adscritos a uno mismo sobre una base distinta de la observación de la conducta propia y adscritos a otros sobre la base de criterios de conducta.'⁶ Sin embargo, los predicados mentales no tienen un significado ambiguo, no son equívocos, a pesar de este cambio sistemático de criterios para su adscripción. Que yo diga 'me duelen las muelas' es un criterio, entre otros, que los demás usan para creer y decir de mí que me duelen las muelas, pero no es un criterio que yo mismo utilizo. Y, sin embargo, en los enunciados 'me duelen las muelas' y 'le duelen las muelas', el verbo 'doler' no se emplea con significados distintos. Strawson subraya correctamente esta unidad de significado de los predicados mentales: 'No se trata de que estos predicados tengan dos tipos de significado. Se trata más bien de que es esencial al tipo único de significado que poseen que ambas formas de atribuirlos estén perfectamente en orden.'⁷ A pesar de esta asimetría, las atribuciones en primera y en tercera persona tienden a engranar armónicamente: no puede ser accidental que yo diga 'me duelen las muelas' o 'voy a tomarme un café' precisamente en las ocasiones en que los demás considerarían también correcto atribuirme ese dolor o esa intención.

La hipótesis que pretendemos sugerir es que una comprensión correcta de la subjetividad depende, al menos en gran medida, de una comprensión adecuada de la interacción y armonía entre las perspectivas de la primera y la tercera persona, así como entre las atribuciones psicológicas correspondientes a ambas. Ninguna de las dos perspectivas y formas de atribución ofrece, considerada aisladamente, una imagen correcta del significado de los términos mentales y de la naturaleza del sujeto al que son atribuidos.

2. La tradición cartesiana.

En la tradición inspirada en el cartesianismo, que incluye no sólo la ortodoxia cartesiana, sino también las tendencias fenomenistas del empirismo británico y del empirismo lógico, así como las corrientes fenomenológicas, la subjetividad tiende a caracterizarse en términos de la conciencia

⁵ D. Davidson, 'First Person Authority', *Dialectica* 38 (1984), pp. 101-111, esp. p. 107.

⁶ P. F. Strawson, *Individuals. An Essay in Descriptive Metaphysics*, Routledge, Londres/Nueva York, 1959 (reimpr. 1990), p. 108.

⁷ *Ibid.*, p. 110.

inmediata que un sujeto tiene, en la perspectiva de la primera persona, de entidades y eventos no físicos que pueblan su mente: 'contenidos psíquicos', 'datos de conciencia', 'datos sensoriales', 'vivencias', son algunas de las expresiones usadas para referirse a esas entidades y eventos, cuya presencia en la mente de un sujeto sólo puede ser sospechada o conjeturada por otros a partir de signos externos. Esta orientación tiene, al menos a primera vista, recursos para dar cuenta de la asimetría en la atribución de predicados mentales y del privilegio epistemológico de las autoatribuciones. La inmunidad, o la notable resistencia al error de las autoatribuciones se debería a que en ellas el sujeto está informando de su conciencia inmediata de un evento o estado mental, cuyo ser consiste precisamente en ser objeto de conciencia: no hay aquí distinción entre apariencia y realidad y, por tanto, no hay resquicios por donde el error pudiera introducirse. Esta explicación tiene cierto atractivo inicial, pero en cuanto reflexionamos más profundamente sobre ella, tropezamos con un conjunto de dificultades clásicas que han jalonado la historia de esta concepción filocartesiana. En primer lugar, la explicación postula la existencia de una forma de percepción distinta de la sensorial,⁸ llamada a veces sentido interno, conciencia o introspección, cuyo objeto no son entidades, eventos o hechos ordinarios y cuya naturaleza resulta, así, extraordinariamente elusiva o misteriosa. Los objetos de esta facultad perceptiva no sensorial no existen sino mientras son percibidos, con lo que su naturaleza no resulta menos elusiva o misteriosa que la de la propia facultad. En segundo lugar, esta concepción genera todo un conjunto de dificultades relacionadas con el llamado problema de las otras mentes. ¿Cómo puedo saber que los demás poseen también estados mentales, que son sujetos como yo y no meros autómatas? Si, en un arranque de generosidad, concedo que tienen estados mentales, ¿cómo puedo saber que cuando dicen, por ejemplo, 'tengo dolor' tienen lo mismo que yo tengo cuando digo 'tengo dolor'? ¿Cómo puedo saber que cuando digo 'Juan tiene dolor', pretendiendo atribuir a Juan lo que yo tengo cuando tengo dolor, estoy diciendo la verdad si sólo sé qué es el dolor en mi propio caso? La respuesta a todas estas preguntas es: 'De ningún modo'.⁹ El privilegio que esta orientación concede a la perspectiva de la primera persona priva de toda justificación razonable nuestra convicción de que los términos mentales tienen un significado común cuando se aplican tanto en primera como en tercera persona. Por otra parte, el recurso a un modelo perceptivo para explicar el autoconocimiento conduce a esta concepción a asimilar indebidamente los estados intencionales, como la creencia, a las sensaciones.¹⁰ La plausibilidad inicial que esta explicación pueda tener para la autoadscripción de sensaciones se pierde cuando se trata de la autoadscripción de creencias, pues éstas no son 'sentidas' ni poseen cualidad fenomenológica, y, por otra parte, tienen una relación constitutiva con enunciados condicionales sobre la conducta futura del sujeto, en circunstancias no necesariamente previstas por éste.¹¹ Ambos aspectos de las creencias (que pueden extenderse a otros estados

⁸ O distinta de la percepción sensorial en sentido ordinario, realista, según el cual lo que percibimos son objetos públicos. Debo a T. Grimaltos esta observación: para el fenomenismo, el objeto inmediato de la percepción sensorial son ideas, datos sensoriales, etc.

⁹ O tal vez por analogía con mi propio caso. Ha sido una vez más V. Sanfélix quien me ha recordado el venerable 'argumento por analogía' como respuesta al problema de otras mentes. No creo necesario insistir en las críticas a este argumento.

¹⁰ Sobre esta observación, cf. C. Wright, 'Review of McGinn's *Wittgenstein on Meaning*', *Mind* 98 (1989), pp. 289-305. Cf. tb., del mismo autor, 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention', *Journal of Philosophy* 86 (1989), pp. 622-634. J. Marrades ha llamado mi atención sobre el *Teeteto* platónico como el primer ejemplo histórico de esta asimilación.

¹¹ Cf. *ibid.*

intencionales) las distinguen de las sensaciones y hacen que el modelo de la 'percepción mental' se muestre inadecuado para dar cuenta de la autoridad del sujeto sobre las mismas.

Resulta difícil exagerar la influencia que la concepción filocartesiana de la mente ha tenido hasta tiempos relativamente recientes, no sólo en el ámbito filosófico, sino también en los supuestos y fundamentos implícitos en distintas escuelas psicológicas y en muchas ciencias humanas. Según esta concepción, hay hechos irreductiblemente subjetivos, que sólo se revelan a la conciencia individual. Estos hechos son inaccesibles, no sólo en la práctica, sino en principio, a los métodos objetivos de investigación, de modo que cualquier conocimiento objetivo del mundo será necesariamente incompleto.¹² Veremos cómo esta conclusión será defendida de nuevo en tiempos muy recientes, lo que muestra la fuerza, persistencia y poder de atracción de la concepción cartesiana de la subjetividad como un ámbito privado, accesible sólo a la perspectiva restringida de la primera persona.

3. El conductismo lógico.

En conexión con el programa de la unidad de la ciencia en el positivismo lógico, y favorecidos por el auge del conductismo en psicología, surgieron, hacia los años treinta, los primeros intentos filosóficos de oposición global al cartesianismo y de integración de la mente en el mundo natural bajo la forma del conductismo lógico.¹³ Con matices diferenciadores importantes, otras posiciones dominantes en nuestro siglo, como la teoría de la identidad y el funcionalismo, han heredado del conductismo lógico la orientación reduccionista y la oposición al cartesianismo. Por otra parte, el original pensamiento de Wittgenstein contiene asimismo un formidable ataque al cartesianismo fundado en un cuidadoso análisis del problema del significado y del uso de los términos mentales, en el que muchos han visto afinidades con el conductismo lógico, aunque no hallamos en Wittgenstein simpatía alguna por el cientifismo y la unidad de la ciencia.

El conductismo lógico se concibe a sí mismo como una hipótesis semántica sobre el contenido de las proposiciones psicológicas,¹⁴ según la cual estas proposiciones son traducibles, sin pérdida de significado, a proposiciones que contienen únicamente descripciones de la conducta observable de los seres humanos y otros animales y/o de la conducta que éstos desarrollarían en circunstancias externas determinadas. Así, hallarse en un estado mental de un tipo determinado no es sino desarrollar cierta conducta o tener una disposición a desarrollarla. Idealmente, la conducta en cuestión debería ser especificada en un lenguaje puramente físico.¹⁵

Una de las dificultades centrales del conductismo es su incapacidad para dar cuenta de determinados aspectos de lo mental relacionados con la subjetividad, en especial el autoconocimiento y la autoridad de la primera persona. La razón es clara: si lo mental se reduce a

¹² A pesar de los intentos paradójicos de convertir estos hechos, bajo la forma de ideas, datos sensoriales, y otras, en el fundamento mismo del conocimiento objetivo. El derrumbe del programa de reducción fenomenista testimonia la incoherencia básica involucrada en este tipo de proyecto.

¹³ Representan esta posición, entre otros autores, R. Carnap, 'Psicología en lenguaje fisicalista', en A. J. Ayer, *El positivismo lógico*, tr. de L. Aldama y otros, F.C.E., México, 1965 (el artículo de Carnap fue publicado originalmente en 1932); C. G. Hempel, 'The logical analysis of psychology', recogido en N. Block (ed.), *Readings in Philosophy of Psychology*, Harvard University Press, Cambridge Mass., 1980; G. Ryle, *El concepto de lo mental*, tr. de Eduardo Rabossi, Paidós, Buenos Aires, 1967.

¹⁴ Cf., p. ej., Carnap, op. cit.

¹⁵ Esta restricción no afecta a Ryle, pero sí a Carnap: cf. Carnap, op. cit., pp. 195, 203.

la conducta, que es un proceso públicamente observable, no hay motivos para suponer que un sujeto posea una autoridad sobre sus propios estados mentales que no puedan tener otros observadores; no hay motivos para pensar que un sujeto tenga un conocimiento de su propia mente más íntimo e inmediato que el que pueda tener cualquier otra persona que conozca su conducta y las circunstancias en que se desarrolla. Así, para Gilbert Ryle, tanto el conocimiento de la propia mente como el de la mente de otros descansan en la conducta. La diferencia entre ambos es sólo de grado: conozco mejor mi propia conducta porque siempre estoy presente cuando tiene lugar. En palabras de Ryle: '... Me doy cuenta que un alumno es haragán, ambicioso o ingenioso siguiéndolo en su tarea, prestando atención a sus excusas, escuchando su conversación y comparando su comportamiento con el de los demás. Que yo sea el propio alumno no establece ninguna diferencia importante. En este caso, puedo escuchar un mayor número de "sus" conversaciones debido a que soy el sujeto a quien se dirigen sus soliloquios mudos. Además, presto atención a todas "sus" excusas porque, cuando las formula, estoy siempre presente.'¹⁶ La eventual plausibilidad del texto de Ryle deriva de los ejemplos de predicados mentales que emplea: ser haragán, ambicioso o ingenioso son rasgos de carácter de naturaleza puramente disposicional, para los que, como se reconoce hoy en general, no rige la asimetría entre la primera y la tercera persona. Mis propios rasgos de carácter se me revelan, básicamente, a través de mi comportamiento y muchas veces otras personas son mejores jueces que yo mismo en esta cuestión. El error de Ryle se pone de manifiesto si se pretende generalizar su posición a los estados intencionales y a las sensaciones. Parece bastante claro, en este caso, que un sujeto no sabe, por ejemplo, que cree algo o que siente dolor observando su propio comportamiento. El conductismo lógico no tiene recursos para explicar las peculiaridades del autoconocimiento en ninguno de los tres aspectos de la subjetividad, intencional, sensitivo y práctico, que indicamos en el primer apartado.

El conductismo lógico se enfrenta con problemas importantes para dar cuenta de las especiales características que presentan los enunciados psicológicos en primera persona, por ejemplo su peculiar resistencia al error. Los conductistas lógicos advierten que la concepción cartesiana de estos enunciados tiene consecuencias difícilmente admisibles. Así, para Carnap, la tesis según la cual estos enunciados 'se refieren a algo no físico (a algo "psíquico", un "contenido de vivencia", un "dato de conciencia", etc.), lleva directamente a la consecuencia de que [todo enunciado de este tipo]... es significante sólo para su autor.'¹⁷ Wittgenstein, al que, en *este* aspecto, podemos asimilar al conductismo lógico, fue más allá que Carnap para señalar, con su argumento contra la posibilidad de un lenguaje privado, que ni siquiera su autor podría entender ese enunciado.¹⁸ Pero las alternativas que estos autores ofrecen no carecen de problemas. Así, un modo de abordar esta espinosa cuestión, característico del conductismo lógico, consiste en negar que los enunciados en cuestión sean enunciados cognitivos. 'Tengo dolor' no es un informe o una descripción de un estado del sujeto, sino una expresión de dolor. Es parte de la conducta de dolor. Decir 'tengo dolor' es como gritar '¡ay!'. Y, del mismo modo que '¡ay!' no puede ser verdadero o falso, tampoco 'tengo dolor' puede serlo. La emisión de 'tengo dolor' reemplaza otras expresiones de dolor no lingüísticas por una expresión lingüística más refinada. Citemos un texto de Wittgenstein (quien sin embargo presenta su tesis cautamente como 'una posibilidad'): '¿Cómo

¹⁶ Ryle, op. cit., p. 151.

¹⁷ Carnap, op. cit., pp. 199-200.

¹⁸ Cf. L Wittgenstein, *Investigaciones Filosóficas*, tr. de A. García Suárez y C. U. Moulines, U.N.A.M./Ed. Crítica, Barcelona, 1988, par. 258.

aprende una persona el significado de los nombres de sensaciones? Por ejemplo, de la palabra "dolor". Una posibilidad es la siguiente: las palabras se conectan con la expresión primitiva, natural, de la sensación y se ponen en su lugar. Un niño se ha lastimado y grita; luego los adultos le hablan y le enseñan exclamaciones y más tarde oraciones. Ellos le enseñan al niño una nueva conducta de dolor.¹⁹ Este tipo de análisis es insatisfactorio por varias razones. Ciertamente explica por qué una oración como 'tengo dolor', emitida sinceramente por un sujeto con adecuada competencia lingüística, no puede ser falsa; pero no puede explicar por qué se la supone verdadera. No puede ser falsa porque no es propiamente una oración descriptiva. Pero esto le priva también de cualquier posibilidad de verdad. En realidad, este tipo de análisis niega la existencia misma del *autoconocimiento*. Pero esta consecuencia es francamente implausible. Por otra parte, 'tengo dolor' parece ser genuinamente descriptivo. Pensemos en descripciones mucho más complejas y detalladas que podemos ofrecer de muchos de nuestros dolores. En estos casos, que no son esencialmente distintos del primero, no tenemos ninguna tentación de considerar los enunciados en cuestión como meras sustituciones lingüísticas de conducta primitiva.²⁰ Finalmente, parece correcto decir que cada uno de nosotros sabe si tiene dolor y que no infiere este hecho a partir de su propio comportamiento.²¹ Sin embargo, si el dolor es una disposición a desarrollar cierta conducta, como sostiene el conductismo lógico, aun suponiendo que proferir 'tengo dolor' sea parte de la conducta en cuestión, cada uno de nosotros sabría si tiene dolor por inferencia: nos oíríamos proferir 'tengo dolor' y concluiríamos que tenemos dolor. Esto nos conduce de nuevo a la posición de Ryle.

4. El materialismo de la identidad.

El conductismo lógico concibe la mente humana desde la perspectiva de la tercera persona. Su caracterización de la mente parte del análisis de las atribuciones de estados mentales a otras personas y del conocimiento de la mente ajena. Pero fracasa lamentablemente al tratar de entender la autoatribuciones y el autoconocimiento, es decir, los aspectos de lo mental directamente relacionados con la subjetividad. Por otra parte, su concepción de los estados mentales como disposiciones para la conducta resulta más plausible en relación con los estados intencionales que con los estados cualitativos o fenomenológicos (como el dolor o las imágenes mentales). Aunque no puede dar cuenta de las autoatribuciones de estados intencionales, en este aspecto no está peor situado que el cartesianismo, ya que el modelo cartesiano de la percepción mental no resulta plausible en relación con la autoadscripción de este tipo de estados, según indicamos más arriba.

En realidad, los primeros representantes de la propuesta conocida hace algún tiempo como 'materialismo del estado central' y más conocida hoy como 'teoría de la identidad de tipos o

¹⁹ Ibid., par. 244.

²⁰ Cf. P. Carruthers, *Introducing Persons. Theories and Arguments in the Philosophy of Mind*, Routledge, Londres/Nueva York, 1986 (reimpr. 1992), p. 115.

²¹ Wittgenstein consideraba incorrecto decir 'sé que tengo dolor' (cf. op. cit., par. 246). Tal vez mis intuiciones lingüísticas estén irremediamente viciadas, pero no veo nada incorrecto en ello. En determinadas circunstancias es plenamente admisible. Supongamos que un médico me administra un fuerte analgésico y yo me sigo quejando. Si el médico me dice: 'No puedes tener dolor ahora', la réplica 'sé que lo tengo' parece perfectamente adecuada. J. Marrades y V. Sanfélix coinciden en señalar que esta observación no es incompatible con la filosofía wittgensteiniana, especialmente con *Sobre la certeza*. Por mi parte, mantengo mis dudas al respecto. No creo que Wittgenstein pueda admitir un uso cognoscitivo de 'tengo dolor' sin entrar en conflicto con aspectos esenciales de su filosofía de la psicología.

propiedades', U. T. Place y J. J. C. Smart, consideraron su tesis de la identidad psicofísica como un complemento del conductismo lógico, y no como una alternativa global al mismo. Según Place,²² el análisis conductista de los conceptos de estados intencionales²³ era fundamentalmente correcto. Sin embargo, había un residuo de estados mentales de contenido fenomenológico que parecían resistirse a este tipo de análisis, conceptos como el de sensación, experiencia, conciencia e imagen mental. Place propone la hipótesis de la identidad psicofísica sólo para este último grupo de conceptos: las sensaciones, como el dolor, o las imágenes mentales, no eran disposiciones para la conducta, sino estados del cerebro. Esta restricción de la hipótesis de la identidad a los estados fenomenológicos es también aceptada por Smart.²⁴ En cualquier caso, este tipo de estados, las experiencias de calor y frío, de dolor y placer, de las cualidades sensibles de las cosas, plantean un reto a una concepción materialista del mundo, inspirada en la unidad de la ciencia, ya que, al hallarse en tales estados, un sujeto parece relacionarse con objetos no físicos, las sensaciones de calor y frío, dolor, sabor, color, etc. Estos 'objetos' no son físicos en la medida en que no se hallan en el espacio intersubjetivo, en el mundo objetivo estudiado por las ciencias físico-naturales. Como señala Place: 'Si suponemos, por ejemplo, que cuando un sujeto informa de una post-imagen verde está aseverando la presencia en su interior de un objeto que es literalmente verde, es claro que nos las tenemos con una entidad para la que no hay lugar en el mundo de la física.'²⁵ ¿Cómo interpretar entonces enunciados en primera persona como 'tengo dolor' o 'tengo una post-imagen verde' de un modo compatible con el fisicalismo? Smart se hace eco del análisis 'expresivo' de tales enunciados en la versión de Wittgenstein, pero, aunque reconoce sus ventajas para el fisicalismo, no lo considera satisfactorio: 'Aunque... soy muy receptivo a la explicación "expresiva" de los enunciados de sensación, no creo que esta explicación resuelva el problema. Quizá se deba a que no he pensado en ello lo suficiente, pero me parece que, cuando una persona dice "tengo una post-imagen", *está* haciendo un informe genuino, y que, cuando dice "tengo dolor", *está* haciendo algo más que "reemplazar una conducta de dolor" por otra y que este "más" no es simplemente decir que está en apuros.'²⁶ La sugerencia de Smart, pues, es que esos enunciados en primera persona son informes o descripciones de algo, frente a la concepción conductista.

Detengámonos brevemente en la explicación que Place y Smart ofrecen de los enunciados sobre estados fenomenológicos en primera persona, ya que, en nuestra opinión, esta explicación, un tanto olvidada hoy, tiene un gran interés y aborda la cuestión de un modo más cercano a determinadas intuiciones espontáneas que el conductismo lógico o el funcionalismo.

Frente a las reticencias del conductismo lógico, Place y Smart aceptan la existencia de estados fenomenológicos con contenido cualitativo. Las personas tienen realmente experiencias conscientes. Suponer, sin embargo, en el espíritu del cartesianismo, que estas experiencias involucran una relación inmediata con objetos y eventos no físicos significa incurrir en un error: 'Este error lógico, al que me referiré como la "falacia fenomenológica", consiste en suponer que, cuando el sujeto describe cómo le parece que son las cosas, cómo le suenan, huelen o saben, cómo

²² U. T. Place, 'Is Consciousness a Brain Process?', en W. G. Lycan (ed.), *Mind and Cognition*, Blackwell, Oxford, 1990, pp. 29-36. El artículo de Place fue publicado por primera vez en 1956.

²³ Place se refiere a ellos como conceptos cognitivos y volitivos. Cf. op. cit., p. 29.

²⁴ J. J. C. Smart, 'Sensations and Brain Processes', en D. M. Rosenthal (ed.), *The Nature of Mind*, Oxford University Press, Oxford/Nueva York, 1991, pp. 169-180. El artículo de Smart fue publicado por primera vez en 1962.

²⁵ Place, op. cit., p. 35.

²⁶ Smart, op. cit., p. 170.

las siente, está describiendo las propiedades literales de objetos y eventos en un tipo peculiar de pantalla de cine o de televisión, conocida normalmente en la bibliografía psicológica contemporánea como el "campo fenoménico".²⁷ Place sugiere que el modo en que sentimos las cosas no es una cosa adicional a éstas. Para él, no podríamos tener la capacidad de emitir informes introspectivos acerca de cómo nos parecen las cosas, de cómo las experimentamos, si no tuviéramos la capacidad de describir las propiedades reales de las cosas mismas, de los objetos físicos en el espacio público. La primera capacidad es parasitaria y derivada de la segunda. Así, describimos nuestra experiencia subjetiva de las cosas, no por referencia a propiedades fenoménicas de objetos no físicos, sino por referencia a propiedades reales de objetos físicos que normalmente dan lugar al tipo de experiencia consciente que tratamos de describir. Pensemos por ejemplo en algunos adjetivos con los que podemos describir el dolor: intenso, persistente, punzante, profundo, localizado, difuso. Estos adjetivos se aplican también, y de modo primario, a objetos y procesos físicos. Las experiencias subjetivas no son relaciones con objetos interpuestos entre el sujeto y los objetos públicos. La explicación de nuestra capacidad de describir esas experiencias en primera persona no requiere postular la existencia de tales objetos no físicos. Así, señala Place: 'Cuando describimos la post-imagen como verde no estamos diciendo que hay algo, la post-imagen, que es verde; estamos diciendo que tenemos el tipo de experiencia que normalmente tenemos al mirar una mancha de luz verde y que hemos aprendido a describir de ese modo.'²⁸ Independientemente de su incompatibilidad con el materialismo, concebir esa experiencia como una relación con un objeto no físico de color verde da lugar a una incongruencia, en la medida en que, para tener color, un objeto ha de tener extensión y ésta es, para el propio Descartes, el atributo de lo físico. Por nuestra parte, podríamos añadir, parafraseando a Place, que cuando describimos un dolor como punzante no estamos diciendo que hay algo, el dolor, que es punzante; estamos diciendo que tenemos el tipo de experiencia que normalmente tenemos cuando nos pinchamos con un objeto punzante.

Hasta aquí la explicación de Place de los enunciados en primera persona sobre la experiencia subjetiva tiene una plausibilidad independiente de la hipótesis de la identidad, y podría ser aceptada por muchos para quienes el cartesianismo no resulta aceptable pero tampoco consideran correcta la hipótesis de la identidad. Esta última, sin embargo, da a la explicación anterior una forma más comprometida. Identificar la experiencia subjetiva de una post-imagen verde con un proceso cerebral lleva a suponer que tener dicha experiencia sería hallarse en el estado cerebral que es causado normalmente por una superficie verde. Esta explicación no se compromete tampoco con la existencia de un objeto que sería la post-imagen. No es ésta la que se identifica con un proceso cerebral, lo que contravendría la llamada 'ley de Leibniz', ya que se supone que la post-imagen, a diferencia del proceso cerebral, no se halla en el espacio. Es más bien la *experiencia* de la post-imagen la que se identifica, tentativamente, con el proceso cerebral. Y no es absurdo decir que la experiencia en cuestión tiene lugar allí donde se encuentra el sujeto de la misma.

Frente a objeciones de carácter semántico o conceptual, Place insiste en que la hipótesis de la identidad es una hipótesis empírica, no descartable por razones *a priori*. Place y Smart comparan su hipótesis con enunciados como 'el agua es H₂O' o 'la luz es radiación electromagnética'. Podemos

²⁷ Place, op. cit., p. 35.

²⁸ Ibid.; cf. tb. Smart, op. cit., pp. 172, 174.

aceptar que nuestros conceptos de agua y de H₂O son distintos y que no entendemos lo mismo por 'agua' que por 'H₂O'. Pero de ello no podemos concluir que el agua no es H₂O. De hecho lo es.

Sin embargo, los conceptos de lo mental son realmente peculiares. No son simplemente distintos de los conceptos neurofisiológicos, sino que su atribución en primera y en tercera persona se rige por criterios distintos. Según Smart y Place, cuando informamos de nuestras experiencias, estamos en realidad informando de estados de nuestro cerebro. Pero, ¿cómo podría ser así si la atribución de estados neurofisiológicos a uno mismo y a los demás no presenta la asimetría propia de la atribución de experiencias? Yo no poseo especial autoridad con respecto a los estados de mi cerebro. De hecho, un neurólogo es más competente que yo para determinarlos. Pero este mismo neurólogo habrá de confiar en mis declaraciones para saber si tengo dolor cuando me implanta un electrodo en el cráneo. Una teoría que sostiene que las experiencias son estados del cerebro nos debe una explicación de esta divergencia. Smart pone una objeción semejante en boca de un posible adversario: 'Si digo sinceramente "veo una post-imagen amarillenta-naranja" y no estoy cometiendo un error verbal, no puedo estar equivocado. Pero puedo estar equivocado acerca de un proceso cerebral. El científico que mira el interior de mi cerebro podría estar sufriendo una ilusión. Además, tiene sentido decir que dos o más personas están observando el mismo proceso cerebral, pero no que dos o más personas están informando de la misma experiencia interna.'²⁹ La respuesta de Smart es la siguiente: 'Esto muestra que el lenguaje de los informes introspectivos tiene una lógica diferente del lenguaje de los procesos materiales. Es obvio que, hasta que la teoría del proceso cerebral haya mejorado mucho y sea ampliamente aceptada, no habrá *criterios* para decir "Smith tiene una experiencia de tal y tal clase" *excepto* los informes introspectivos de Smith. Así, hemos aceptado una regla de lenguaje según la cual (normalmente) lo que dice Smith vale.'³⁰ Smart admite, en su respuesta, que los términos mentales no sólo tienen un significado diferente de los términos neurofisiológicos, sino también una lógica diferente. Sin embargo, tiende a ver esta lógica como un rasgo provisional. La autoridad de la primera persona sería un expediente al que hemos de recurrir a falta de algo mejor. Puesto que no tenemos, hoy por hoy, otros criterios, hemos establecido una convención que concede esa autoridad a los informes introspectivos. Sin embargo, cuando 'la teoría del proceso cerebral haya mejorado mucho y sea ampliamente aceptada', dispondremos ya, se supone, de otros criterios, de carácter objetivo, para determinar las experiencias de un sujeto, de modo que los informes introspectivos serían sólo un criterio más y no necesariamente el más confiable. En caso de conflicto, podría suceder que atribuyéramos dolor —o ausencia del mismo— a una persona a pesar de que ésta asegurase lo contrario. Sin embargo, Smart minusvalora el carácter constitutivo de la 'lógica' de los términos mentales. Si yo afirmo sinceramente que no tengo dolor y otra persona se empeña en atribuírmelo observando mi cerebro, no estamos hablando de lo mismo. La palabra 'dolor' habría cambiado de significado en boca de esta persona. Pero en este caso la teoría de la identidad perdería interés. Que una propiedad llamada 'dolor', distinta de lo que ahora entendemos por 'dolor', fuese un estado del cerebro sería tal vez un hecho interesante, pero no una respuesta al problema de la naturaleza de la mente. Hay casos en que el descubrimiento de una identidad de propiedades genera criterios de identidad más fiables. Pensemos, por ejemplo, en el descubrimiento de que el agua es H₂O. El análisis químico resulta ahora decisivo ante un conflicto con criterios sensoriales, sabor, olor, color. Pero esto no supone

²⁹ Smart, op. cit., p. 173.

³⁰ Ibid., pp. 173-174.

que con la expresión 'H₂O' se esté hablando de otra cosa que con la expresión 'agua'. Sin embargo, el caso de los términos mentales es claramente distinto. Antes del descubrimiento de la química moderna, el concepto de agua no presentaba la asimetría característica de los predicados mentales.

En nuestra opinión, cualquier teoría acerca de la identidad de las propiedades mentales con propiedades aparentemente distintas, *P*, debe afrontar el problema de la asimetría. ¿Son las propiedades *P* objeto de una atribución asimétrica? ¿He de atribuir una propiedad *P* con criterios distintos en mi propio caso y en el de los demás? Si la respuesta es negativa, es difícil ver cómo la identidad propuesta podría ser correcta. La teoría de la identidad psicofísica no supera esta prueba de modo satisfactorio.

5. El funcionalismo.

En general, las teorías reductivas de la mente tienen problemas para dar cuenta de la asimetría y de la autoridad de la primera persona y, con ello, para dar cuenta de la subjetividad. De ahí que hallemos con frecuencia intentos de negar la realidad de tales rasgos o de rebajar su significación en el contexto de tales teorías. Podemos comprobar también esta tendencia en el funcionalismo. Esta teoría, que ha reemplazado progresivamente a la teoría de la identidad de propiedades, sostiene que las propiedades mentales son propiedades funcionales. La identidad propuesta por esta teoría es más abstracta, ya que la clasificación funcional puede agrupar bajo una misma categoría estados u objetos muy diversos desde el punto de vista de sus propiedades físicas. Así, por ejemplo, ser un reloj o ser un carburador son propiedades funcionales. Clasificar objetos bajo las categorías 'reloj' y 'carburador' es llevar a cabo una clasificación funcional, no una clasificación física. Podemos fácilmente suponer que en el cajón con el rótulo 'relojes' encontraremos objetos enormemente diversos en sus propiedades físicas. Igualmente, señalar las siete en punto es un estado que pueden compartir muchos de estos objetos. Se trata asimismo de un estado funcional, cuyos casos particulares pueden no tener apenas características físicas relevantes en común.

El funcionalismo permite la atribución de propiedades mentales a seres muy diversos desde el punto de vista físico. Este es un grado de libertad del que carecía la teoría de la identidad.³¹ Sistemas físicos tan diversos como los organismos humanos, posibles seres extraterrestres carentes de sistema nervioso e incluso ordenadores o robots pueden tener estados mentales con tal que posean la complejidad funcional adecuada. Aunque el funcionalismo rechaza la reducción de los conceptos mentales a conceptos conductuales y a conceptos neurofisiológicos, su inspiración es también reductiva y naturalista. En una definición funcionalista de una propiedad mental no deben aparecer términos mentales. Se habla en ella únicamente de 'estados' caracterizados exclusivamente por sus relaciones con causas y efectos descritos en términos no mentales y con otros estados caracterizados también de este modo.³² La psicología se reduce así a una teoría funcional. No hay

³¹ Este aspecto es subrayado por Hilary Putnam, el primer filósofo que formuló con claridad la hipótesis del funcionalismo. Cf. H. Putnam, 'The Nature of Mental States', en D. M. Rosenthal (ed.), *The Nature of Mind*, cit., pp. 197-203. Este artículo fue publicado por primera vez en 1967. Cf. tb. su 'Mentes y máquinas', en A. M. Turing, H. Putnam, D. Davidson, *Mentes y máquinas*, ed. Tecnos, Madrid, 1985, pp. 63-101. Este artículo fue publicado originalmente en 1960 bajo el título 'Minds and Machines'.

³² Cf. N. Block, 'An Excerpt from "Troubles with Functionalism"', en W. G. Lycan (ed.), *Mind and Cognition*, cit., pp. 444-468, esp. pp. 448-450.

nada especial en las propiedades mentales que no pueda ser recogido en una caracterización funcional.

Notemos que la concepción funcionalista de la mente revela con claridad su origen en una reflexión sobre la atribución de estados mentales en tercera persona. Para David Armstrong, uno de los precursores de esta doctrina, concebimos los estados mentales, en el marco de la vida cotidiana, en términos causales. Según Armstrong, 'el concepto de un estado mental incluye *de modo esencial y exhaustivo* [subr. mío] el concepto de un estado que es apto para ser *la causa de ciertos efectos o apto para ser el efecto de ciertas causas*.'³³ Así, un estado mental es un estado interno que atribuimos a un sujeto para entender y explicar su conducta, dadas las influencias causales del entorno en que se halla y dados otros estados internos. Este es el concepto de estado mental que subyace al funcionalismo, aun en sus versiones más refinadas, como la inspirada en la inteligencia artificial. De acuerdo con este concepto, las atribuciones de estados mentales descansan en inferencias basadas en la constatación de un contexto causal. Parece claro que esta concepción no se compadece con el carácter aparentemente no inferencial de las autoatribuciones. Pensemos por ejemplo en el dolor. Vemos que una persona tropieza y comienza a quejarse. Explicamos entonces su conducta atribuyéndole dolor. Aquí puede parecer que inferimos un estado interno al que llamamos 'dolor' para dar cuenta de su conducta (aunque, en nuestra opinión, no es esto lo que hacemos). Sin embargo, si yo soy esa persona y me atribuyo dolor, no lo hago infiriendo ese estado a partir de mis quejas. La concepción funcionalista no se adapta bien a esta asimetría.

En realidad, en el tratamiento de las autoatribuciones, el funcionalismo ofrece soluciones paralelas a las que encontramos en el conductismo lógico. Por una parte, puede concebirlas como parte de los efectos característicos de un estado mental. Así, el dolor es aquel estado que daría lugar, típicamente, entre otras cosas, a la emisión de 'tengo dolor'. Las dificultades de esta concepción son similares a las que presenta la concepción expresiva de las autoatribuciones. Por otra parte, puede optar por una solución semejante a la de Ryle: no hay realmente asimetría ni especial autoridad de la primera persona; nos atribuimos estados mentales sobre la misma base sobre la cual los atribuimos a otros, a saber, la interpretación y explicación de la conducta. Algo muy cercano a esto es lo que, en nuestra opinión, sostiene Daniel Dennett cuando escribe: 'Postulamos todas esas aparentes actividades y procesos mentales *con vistas a entender* la conducta que observamos —con vistas, de hecho, a entender la conducta tanto como podamos, especialmente *cuando la conducta que observamos es la nuestra* [subr. mío]. Los filósofos de la mente solían... insistir en que el acceso al propio caso en tales cuestiones es bastante distinto del acceso al de los demás, pero a medida que sabemos más sobre diversas formas de psicopatología e incluso sobre las flaquezas de las personas aparentemente normales, se hace más plausible suponer que, aunque todavía hay pequeños rincones de privilegio no cuestionado, algunos asuntos sobre los que nuestra autoridad es invencible, cada uno de nosotros es, en la mayoría de aspectos, una especie de autopsicólogo inveterado, que *inventa* sin esfuerzo interpretaciones intencionales de sus propias acciones en una mezcla inseparable de confabulación, autojustificación retrospectiva y (sin duda, en ocasiones) buena teorización.'³⁴ Dennett se refiere aquí, predominantemente, a estados intencionales. Sin embargo, en otro trabajo³⁵ insiste, en una línea cercana a Wittgenstein, en que

³³ D. M. Armstrong, 'The Causal Theory of the Mind', en W. G. Lycan (ed.), *Mind and Cognition*, cit., p. 40.

³⁴ D. C. Dennett, 'Making Sense of Ourselves', en W. G. Lycan (ed.), *Mind and Cognition*, cit., p. 190.

³⁵ 'Quining Qualia', *ibid.*, pp. 519-547.

las autoatribuciones de estados fenomenológicos dependen de nuestras reacciones a ciertas situaciones. Finalmente, el funcionalismo ofrece una posibilidad que no tenía el conductismo lógico para dar cuenta de la autoridad de la primera persona. Dicha autoridad resultaría del hecho de que los estados mentales tienden a causar creencias sobre ellos en los sujetos que los poseen.³⁶ En nuestra opinión, sin embargo, además de su carácter *ad hoc*, esta explicación —y en general toda explicación de la autoridad de la primera persona y de la conciencia de nuestros propios estados mentales basada en la postulación de creencias de segundo orden— sólo reproduce el problema en un nivel superior, ya que es legítimo preguntar en qué descansa a su vez nuestra autoridad sobre esas nuevas creencias. Un intento de responder a esta pregunta sin un regreso *ad infinitum* nos lleva a las dos alternativas anteriores, cuyas dificultades ya señalamos al referirnos al conductismo lógico.

6. La reacción neocartesiana.

Hemos tratado de mostrar que el naturalismo reductivo en sus distintas formas (conductismo lógico, teoría de la identidad de propiedades, funcionalismo) tiene serias dificultades para tratar con los problemas directamente vinculados con la subjetividad, como el autoconocimiento, la asimetría en la atribución de predicados mentales y la autoridad de la primera persona. El predominio de estas concepciones ha ido, pues, acompañado de cierta insatisfacción, de un vago sentimiento de que tales concepciones son, por su propia naturaleza objetivista, por su apuesta por una perspectiva de tercera persona, estructuralmente incapaces de dar cuenta de aspectos subjetivos de la mente que sólo se revelan a la perspectiva de la primera persona. Como en un movimiento pendular, esta insatisfacción ha generado un retorno al punto de partida cartesiano con el que el conductismo lógico pretendió romper. Para el neocartesianismo, el naturalismo reductivo se mueve en círculos inesenciales sin poder acceder jamás al núcleo central de la mente: la conciencia subjetiva y sus objetos, inaccesibles a un punto de vista objetivo. La reacción neocartesiana ha insistido sobre todo en lo que más arriba denominamos el aspecto sensitivo de la subjetividad, considerando este aspecto como especialmente resistente a un tratamiento objetivo. Hay varios argumentos destinados a mostrar que determinadas propiedades de la experiencia subjetiva son un obstáculo infranqueable para cualquier teoría objetiva de lo mental. Ofreceremos una caracterización relativamente sumaria de los mismos.

Preguntémonos, en primer lugar, si no es posible que otra persona y yo mismo, al mirar un determinado objeto, tengamos experiencias subjetivas de su color completamente distintas, digamos, sistemáticamente invertidas en relación con el espectro cromático. Así, aunque ambos decimos que la hierba es verde, él ve la hierba del color del que yo veo los coches de bomberos y, aunque ambos decimos que los coches de bomberos son rojos, él los ve del color del que yo veo la hierba.³⁷ Nuestra conducta verbal no podría revelar la diferencia, ya que ambos hemos aprendido los nombres de los colores aplicándolos a objetos públicos y tenemos la misma capacidad de discriminar colores (no estamos hablando del daltonismo, un defecto que es detectable porque un daltónico no tiene la misma capacidad de discriminar colores que un sujeto no daltónico). En realidad, ningún criterio intersubjetivo, ninguna información fisiológica, conductual o funcional podría ser plenamente concluyente para decidir si esta diferencia se da. Sin embargo, la situación

³⁶ Sobre esta posibilidad, cf. p. ej. P. Carruthers, op. cit., pp. 128-129.

³⁷ Sobre este tipo de argumento, cf. N. Block, op. cit., p. 458. Cf. tb. C. McGinn, *The Problem of Consciousness*, Blackwell, Oxford, 1991, pp. 196-197.

parece concebible, y ello se debe a que la proyectamos desde la perspectiva de la primera persona: podemos concebir que, a partir de cierto momento, sufriéramos esa inversión y nos preguntamos entonces si la experiencia que entonces tendríamos no será la que siempre ha tenido la otra persona. Intuiciones cartesianas se ocultan tras este tipo de elucubraciones: sólo yo puedo saber, estamos tentados de decir, a qué color me refiero cuando digo que la hierba es verde, sólo yo sé cómo percibo los colores; los demás sólo pueden hacer conjeturas. Y esto puede decirlo cada uno de sí mismo. Así, además de los objetos públicos y los colores que públicamente les atribuimos, están los colores subjetivos. Este tipo de propiedades, que ahora suelen denominarse 'qualia', son puramente subjetivas, privadas e inefables, vedadas para siempre al conocimiento intersubjetivo. Es claro cómo esta reflexión podría ser aplicada a los sabores, los olores, placeres, dolores o sonidos.³⁸

Otros argumentos, en un espíritu semejante, han sido desarrollados para mostrar que cualquier teoría materialista, funcionalista, conductista y, en general, cualquier teoría objetiva o de tercera persona acerca de la mente es necesariamente incapaz de describir adecuadamente ciertos hechos esenciales acerca de la experiencia subjetiva. Estos argumentos pueden ser generalizados con vistas a mostrar que el materialismo es falso. En un artículo ya famoso,³⁹ Thomas Nagel sostiene que no tenemos medios para saber qué es ser un murciélago, en el sentido de qué es, *para un murciélago*, tener las experiencias que tiene, experimentar el mundo como él lo hace. El carácter subjetivo de la experiencia sólo se revela plenamente a un punto de vista y el punto de vista de un murciélago es algo a lo que no tenemos acceso: 'No será de ayuda tratar de imaginar que tenemos membranas en los brazos que nos permiten volar en el crepúsculo cazando insectos con la boca; que tenemos una visión muy deficiente y percibimos el entorno por un sistema de señales sonoras reflejadas de alta frecuencia; y que pasamos el día colgados por los pies, cabeza abajo, en un ático. En la medida (que no es mucha) en que puedo imaginar esto, sólo me dice lo que sería para *mí* comportarme como se comporta un murciélago. Pero ésa no es la cuestión. Lo que quiero saber es qué es ser un murciélago *para un murciélago*. Pero si trato de imaginar esto, estoy limitado a los recursos de mi propia mente, y estos recursos son inadecuados para esta tarea.⁴⁰ Sólo alguien capaz de tener las experiencias de los murciélagos podría saber qué es ser un murciélago en el sentido indicado. Un conocimiento completo de la neurofisiología de los murciélagos nos deja muy lejos de poder responder a esa pregunta. En realidad, 'si el carácter subjetivo de la experiencia es plenamente aprehensible sólo desde un punto de vista',⁴¹ cuanto mayor sea nuestro conocimiento objetivo, físico o neurofisiológico, tanto más nos alejaremos de la naturaleza del fenómeno. Y la experiencia es esencialmente subjetiva. Nagel se pregunta retóricamente si tiene sentido inquirir qué son *realmente* mis experiencias, por contraposición a cómo se me aparecen. No entendemos la idea de que las experiencias tienen una naturaleza objetiva distinta de cómo son *para un sujeto*, de modo que no podemos entender genuinamente qué sería captar su naturaleza en una descripción

³⁸ Para un tratamiento detallado de la objeción de los qualia al funcionalismo, desde una perspectiva más favorable a dicha objeción que la que aquí adoptamos, cf. V. Sanfélix, 'Panorama actual de la filosofía de la mente: funcionalismo y experiencia', en M. Torrevejano (comp.), *Filosofía analítica hoy*, Universidade de Santiago de Compostela, 1991, pp. 121-153.

³⁹ Th. Nagel, 'What is it like to be a bat?', en id., *Mortal Questions*, Cambridge University Press, Cambridge, 1979, pp. 165-180.

⁴⁰ Ibid., p. 169.

⁴¹ Ibid., p. 174.

objetiva, como pretenden las teorías que hemos analizado en apartados anteriores. En realidad, el recurso a los murciélagos como especie es más bien didáctico y dramático, ya que una argumentación análoga podría llevarnos a la conclusión de que no podemos saber qué es ser otra persona, experimentar las cosas como ella: pensemos en el argumento sobre la inversión de los qualia.⁴² La conclusión de Nagel es que hay hechos acerca de la mente, acerca de la experiencia subjetiva y de la conciencia, que ninguna teoría objetiva es capaz de apresar y describir adecuadamente.

Nos referiremos, finalmente, a un argumento desarrollado por Frank Jackson.⁴³ Jackson nos pide que imaginemos el siguiente caso. Mary es especialista —supongamos que es la mejor especialista del mundo— en la neurofisiología de la visión y posee toda la información física que se puede lograr acerca de lo que sucede cuando vemos tomates maduros, o el cielo sin nubes, y usamos términos de color. Sin embargo, vive y ha vivido siempre en un entorno de grises y toda su investigación se desarrolla a través de un televisor en blanco y negro. (Supongamos que su propio cuerpo se halla pigmentado en gris). La cuestión es ahora la siguiente, en palabras de Jackson: '¿Qué sucederá cuando Mary sea liberada de su habitación en blanco y negro o se le entregue un televisor en color? ¿Aprenderá algo o no? Parece sencillamente obvio que aprenderá algo sobre el mundo y sobre nuestra experiencia visual del mismo. Pero entonces no podemos evitar la conclusión de que su conocimiento anterior era incompleto. Ahora bien, ella tenía *toda* la información física. *Ergo*, hay que tener más información que ésta y el fisicalismo es falso.'⁴⁴ ¿Qué es exactamente lo que Mary aprende en ese momento? ¿Qué es lo que ahora sabe y antes no sabía? Jackson no da una respuesta precisa a esta pregunta y la confía a nuestra intuición. No me parece correcto decir que Mary no sabía qué eran los colores. De hecho, podría ofrecer definiciones correctas de los mismos. Podría decir, por ejemplo, que el azul es el color del cielo sin nubes y sabría perfectamente qué efectos tiene ese color en nuestros ojos, sistema nervioso, etc., de modo que podría decir: el azul es la propiedad que tiene tales y cuales efectos, ofreciendo así una caracterización funcional correcta de ese color. Podría ofrecer también una definición física, en términos de la longitud de onda de la luz reflejada por ciertos objetos. Lo que Mary no tenía, tal vez, era un conocimiento *experiencial* de los colores, no sabía qué es la experiencia visual del color. Así, hay algo que Mary no sabe acerca de los colores: cómo aparecen a nuestra experiencia. De nuevo, es la experiencia subjetiva y su cualidad lo que se opone a las concepciones objetivas de lo mental. Ningún grado de información conductual, física, neurofisiológica o funcional puede decirnos qué es la experiencia del color, del dolor, del sabor, etc. A esto sólo podemos acceder de modo subjetivo, teniendo esa experiencia.

Argumentos de este tenor jalonan en realidad la historia de la filosofía, en particular la historia del empirismo. Su renacimiento en los tiempos actuales testimonia la enorme fuerza de atracción de las intuiciones cartesianas sobre la mente.

7. Discusión de los argumentos neocartesianos.

En la base de los argumentos neocartesianos encontramos el supuesto según el cual toda caracterización de una experiencia en términos objetivos, por sus relaciones con la conducta, con

⁴² Cf. J. I. Biro, 'Consciousness and Subjectivity', en E. Villanueva (ed.), *Consciousness*, Ridgeview Publishing Company, Atascadero, California, 1991, pp. 113-133.

⁴³ F. Jackson, 'Epiphenomenal Qualia', en W. G. Lycan (ed.), *Mind and Cognition*, cit., pp. 469-477.

⁴⁴ *Ibid.*, p. 471.

el sistema nervioso o con sus causas y efectos es una caracterización accidental: no podemos saber de ese modo *qué es* dicha experiencia, cuál es su naturaleza. Así, lo único realmente esencial para saber qué es una experiencia es tenerla, sentirla; podríamos prescindir de toda esa información objetiva con tal que mantuviéramos la información subjetiva. Esta es la única realmente relevante para saber qué es dicha experiencia. Dicho de otro modo: para el cartesianismo y el neocartesianismo, la identidad de una experiencia, de un quale —la experiencia visual del rojo o la experiencia gustativa del sabor de la cerveza— es intrínseca, independiente de sus relaciones extrínsecas con otras cosas, y se revela directamente a la conciencia del sujeto. Por el contrario, para las posiciones materialistas que hemos analizado, esas relaciones extrínsecas determinan exhaustivamente esa identidad. Pensemos, sin embargo, en el argumento de la inversión del espectro cromático. *Ex hypothesi*, las relaciones extrínsecas de las experiencias visuales respectivas de los dos sujetos cuando miran la hierba son las mismas, y sin embargo, si la inversión es posible, las experiencias son distintas. ¿No da esto la razón al cartesianismo?

Diversos autores han elaborado réplicas a los argumentos neocartesianos. Daniel Dennett ha puesto en cuestión la existencia misma de los qualia, entendidos a la manera cartesiana como estados inefables y privados, dotados de una identidad intrínseca y aprehensibles directa e inmediatamente en la conciencia.⁴⁵ Dennett procede a través de ejemplos con los que trata de minar nuestra aparente confianza en la existencia de tales estados. Tomemos, pongamos por caso, un sorbo de cerveza fresca. Esa experiencia gustativa que tenemos al hacerlo es un quale, si algo lo es; es aquello que sólo conocemos al saborear la cerveza y que ninguna información objetiva podría revelarnos; una persona que lo supiera todo acerca de los procesos materiales y funcionales que tienen lugar cuando una persona bebe cerveza aprendería aún algo, y algo esencial, al probarla: cómo sabe la cerveza, cuál es su sabor. Así, si bebemos cerveza en ocasiones, se supone que sabemos cuál es su sabor, y que lo sabemos íntima y directamente. Bien, ¿y cuál es su sabor? Supongamos que nos gusta la cerveza. Recordemos ahora la primera vez que la probamos. Seguramente no nos gustó su sabor. Lo encontramos amargo y desagradable. Pero ahora nos gusta y, al tomar un sorbo, estamos en contacto inmediato con eso que se supone conocemos: esa cualidad inefable de su sabor. Ahora bien, ¿es esa cualidad, es el sabor que ahora experimentamos el mismo que experimentamos la primera vez? ¿Sabía entonces igual que ahora sólo que nuestras actitudes hacia ese sabor han cambiado o sabe ahora de modo distinto y nuestras actitudes son las mismas, es decir, encontraríamos desagradable el sabor que entonces encontramos desagradable? ¿O tal vez hay una mezcla de ambas cosas? Seguramente no podemos ofrecer una respuesta categórica, pero, si hemos de dar una respuesta, no podemos dejar de lado nuestras reacciones de entonces y de ahora, lo que entonces dijimos e hicimos y lo que hacemos y decimos ahora. ¿Dónde está, pues, esa identidad intrínseca del quale gustativo? En realidad, no necesitamos ir tan lejos: si bebemos un vaso de cerveza ahora, ¿sabe igual el primer sorbo que el último? La respuesta a esta pregunta debería depender de una comparación de qualia. Sin embargo, parece que es más bien lo que *decimos* en respuesta a esta pregunta lo que de algún modo fija los supuestos qualia.

Reflexiones de este tipo ponen en cuestión la existencia de estados dotados de las propiedades de inefabilidad, privacidad, identidad intrínseca y aprehensibilidad inmediata en la conciencia, la existencia misma de los supuestos qualia. Un partidario de éstos, sin embargo, podría objetar que es injusto que se nos exija efectuar comparaciones: la identidad del quale está fijada *ahora*. Al tomar un sorbo de cerveza, establezco contacto directo con ese quale: la cerveza sabe de

⁴⁵ D. Dennett, 'Quining Qualia', cit.

este modo. Sin embargo, la pretensión de fijar la identidad de esa sensación a través de una definición ostensiva privada nos dirige hacia observaciones inspiradas en Wittgenstein: esa pretensión sería como tratar de determinar la hora en un reloj en el que la aguja estuviera conectada a la esfera, de modo que ambas se movieran juntas o como tratar de comprobar la verdad de una información aparecida en el periódico de hoy comprando varios ejemplares.⁴⁶

No obstante, aun aceptando que la identidad de los qualia no es intrínseca, resulta difícil negar que, aun teniendo toda la información sobre los procesos neurofisiológicos, físicos y funcionales involucrados en el proceso de la visión del color, Mary, la famosa neurofisióloga, adquiere un nuevo conocimiento, aprende algo al salir de su entorno de grises, algo que no estaba contenido en esa información objetiva que suponemos completa. Distintos autores han tratado de reconciliar esta intuición con el materialismo al proponer la hipótesis según la cual el conocimiento que Mary adquiere no es el conocimiento de nuevos hechos, sino un conocimiento de tipo práctico: un *saber cómo*, no un *saber que*, en términos de Ryle. Laurence Nemirow sugiere que Mary adquiere una nueva habilidad, una capacidad para hacer ciertas cosas, en particular para 'visualizar' los colores en su imaginación. Nemirow sostiene, pues, que 'saber qué es ver los colores' debería analizarse como 'saber cómo visualizar los colores'.⁴⁷ Este análisis tiene un considerable poder explicativo, lo que habla en su favor. Da cuenta, por ejemplo, de nuestra tendencia a hablar de un nuevo conocimiento adquirido por Mary, ya que la terminología del conocimiento se aplica también a las habilidades: una persona sabe que la Tierra es redonda, pero también *sabe* nadar, montar en bicicleta o tocar el piano. Y también da cuenta de que el conocimiento que adquiere Mary lo es de algo inefable, pues no se puede aprender a nadar, a montar en bicicleta o a tocar el piano mediante una explicación verbal, ni podemos explicar a otros *cómo* lo hacemos de un modo que les ahorre la práctica y el ejercicio. Este tipo de análisis constituye una alternativa interesante a la explicación que sugieren los argumentos neocartesianos, a saber, que el conocimiento que Mary adquiere al salir de su entorno es el conocimiento de nuevos hechos o propiedades irreductiblemente subjetivas, aquello que David Lewis denomina 'información fenoménica'.⁴⁸ Lewis acepta también, en términos generales, la hipótesis de Nemirow, ya que, en su opinión, la hipótesis alternativa, la hipótesis de la información fenoménica, es incompatible con el mínimo común a todo materialismo: la tesis de la superveniencia, según la cual dos posibilidades físicamente iguales son iguales *simpliciter*. Así, si Mary conoce todos los hechos *físicos* acerca de la visión del color y, al salir de su entorno, conoce nuevos hechos, adquiere nueva información, esto supone que estos nuevos hechos no dependen de los hechos físicos, de modo que hay hechos que permiten distinguir posibilidades físicamente indistinguibles y el materialismo es falso. Cuando Mary sale de su entorno de grises, pasa a saber qué es ver los colores. Pero si 'Mary sabe qué es ver los colores' puede analizarse suponiendo que Mary adquiere nuevas habilidades, por ejemplo, habilidades para recordar, imaginar y reconocer colores, que sabe hacer cosas que antes no sabía hacer, no estamos obligados a rechazar el materialismo. No hay nuevos hechos o nuevas propiedades que no dependan de los hechos y las propiedades físicas. La adquisición de estas nuevas habilidades puede ser explicada sin recurrir al contacto de Mary con hechos fenoménicos, apelando a la nueva situación física en la que Mary pasa a hallarse y a las huellas físicas que su

⁴⁶ L. Wittgenstein, op. cit., par. 265.

⁴⁷ L. Nemirow, 'Physicalism and the Cognitive Role of Acquaintance', en W. G. Lycan (ed.), *Mind and Cognition*, cit., pp. 490-499.

⁴⁸ D. Lewis, 'What Experience Teaches', en W. G. Lycan (ed.), *Mind and Cognition*, cit., pp. 499-519.

contacto con los colores deja en ella y que le permiten ejercer esas capacidades. El tipo de hechos sobre la fisiología de la visión cromática que Mary conocía se dan ahora en ella y dan cuenta de sus nuevas habilidades.

8. La subjetividad: primera y tercera persona.

Permítasenos ofrecer, para terminar, un diagnóstico tentativo de la discusión que nos ocupa y una propuesta de solución.

Tanto los neocartesianos como sus oponentes aceptan que, cuando Mary sale de su entorno en blanco y negro, sabe algo nuevo: qué es ver los colores. Unos y otros, sin embargo, discrepan en cómo analizar este plus de conocimiento. Los neocartesianos sostienen que Mary tiene ahora acceso a hechos y propiedades reales que sólo se le revelan en la perspectiva de la primera persona. Los materialistas defienden en cambio que Mary se ha limitado a adquirir nuevas habilidades o capacidades, y que en eso consiste saber qué es ver los colores.

Por nuestra parte, vemos con simpatía el análisis en términos de habilidades, pero si reflexionamos sobre el tipo de conceptos que están en juego, a saber, conceptos psicológicos, con su peculiar asimetría de atribución, tal vez podamos ofrecer un análisis que nos permita conceder al cartesianismo que Mary adquiere algo más que habilidades sin renunciar al materialismo. En el caso de los conceptos psicológicos, la habilidad para emplearlos en primera persona es parte esencial del dominio de tales conceptos, es parte esencial del conocimiento de la naturaleza de las propiedades mentales. Recordemos que aceptábamos, con Strawson, que la asimetría de atribución era parte del significado de los predicados psicológicos. Así, si un sujeto no puede autoatribuirse correctamente, digamos, dolor, no sabe en plenitud qué significa el término 'dolor'. Otra forma de decirlo es: no sabe en plenitud qué es el dolor. Pero si un sujeto no pudiera autoatribuirse dolor sin que éste le fuera atribuido por otros y sin atribuirlo a otros, no sería posible que supiera qué es el dolor como el neocartesianismo supone que lo sabe: a través, exclusivamente, de su conciencia inmediata de una cualidad fenoménica. Por el contrario, la identidad de esta cualidad dependería necesariamente de sus relaciones con otras cosas, por ejemplo, con otros sujetos, y no sería posible fijar esa identidad en un medio puramente subjetivo.

Hay algo que Mary, la famosa neurofisióloga, no puede hacer en su acromático encierro, a saber, decir con verdad: 'Percibo algo rojo'. Y hay algo que una persona que nunca ha sentido dolor o miedo no puede hacer, a saber, decir con verdad: 'Tengo dolor' o 'tengo miedo'. Así, hay un sentido importante en que esas personas no saben bien qué es la visión cromática, el dolor o el miedo. Pero no tenemos por qué suponer que tales propiedades no dependen de propiedades físicas, aun cuando no sean ellas mismas propiedades físicas. No tenemos por qué negar lo que Lewis llama el mínimo común a todo materialismo. Por otra parte, sería erróneo decir que tales personas no saben nada de la visión cromática, el dolor o el miedo. Pueden llegar a saber muchas cosas. Pero, como predicados psicológicos, su significado incluye tanto la atribución en primera como en tercera persona, y alguien que no es capaz de llevar a cabo ambas no conoce plenamente ese significado, y, así, no sabe plenamente qué son las propiedades denotadas por esos predicados.

Quisiéramos sugerir ahora que ambos tipos de atribución se hallan en una relación de dependencia recíproca: no son posibles la una sin la otra. Tal vez esto sea afirmar demasiado, ya que Mary, o los sujetos que no sienten dolor o miedo pueden atribuir correctamente a otros visión cromática, dolor o miedo. Pero esta atribución se halla truncada: no saben del todo *qué* están atribuyendo, precisamente porque no pueden atribuirse esas propiedades a sí mismos. Sin embargo,

estos sujetos no podrían autoatribuirse visión cromática, dolor o miedo a partir de su propio caso, si nuestra sugerencia es correcta. Tratemos ahora de ampliarla y justificarla.

El aprendizaje de los conceptos mentales se lleva a cabo en una compleja interacción de atribuciones y perspectivas. Los demás nos atribuyen estados mentales sobre la base de nuestra conducta y nosotros atribuimos estos estados a los demás sobre esta misma base. Sin embargo, si pretendiéramos autoatribuirnos estados mentales sobre la base de nuestra propia conducta, esa autoatribución sería considerada, con razón, errónea. La hipótesis que quisiéramos proponer es la siguiente: la autoatribución de un estado mental como resultado de una inferencia destruye el estado mental autoatribuido y se convierte así en una empresa autodestructiva; y éste es el caso tanto si la inferencia se lleva a cabo a partir de la propia conducta como si se produce a partir de la introspección. En ambos casos, la autoatribución queda privada del necesario carácter directo e inmediato que le concede una especial autoridad frente a las atribuciones en tercera persona. La razón es que la posesión genuina de un estado mental aparta la atención de la propia conducta e incluso de las propias sensaciones y la dirige hacia otras cosas: a lo que tememos en el caso del miedo, a lo que deseamos en el caso del deseo, al objeto visto en el caso de la visión cromática, e incluso, en el caso del dolor, a la parte dolorida de nuestro cuerpo.⁴⁹ En cada uno de estos casos, si intentamos conscientemente dirigir nuestra atención a nuestras sensaciones o a nuestra conducta para descubrir cuáles son nuestros estados mentales, encontraremos que éstos se convierten en otra cosa. La dirección de la atención que conllevan determinados estados mentales tiene dos consecuencias importantes: en primer lugar, nuestra conducta, a la que no atendemos, se convierte, por su espontaneidad, en una base confiable para que los demás nos atribuyan el estado mental en cuestión y, en segundo lugar, esa dirección de la atención hace que necesitemos de los demás para llegar a identificar lo que tenemos como miedo, deseo, dolor o visión cromática. Es esta interacción de perspectivas la que da lugar a la armonía general entre las atribuciones en primera y en tercera persona y constituye así los conceptos de los diversos tipos de estado mental. La perspectiva subjetiva o de primera persona es muy distinta, pues, de como la concibe el cartesianismo: su campo de atención no abarca cualidades interpuestas entre el sujeto y el mundo, sino cualidades del mundo mismo, incluido el propio cuerpo. Las propiedades mentales son relacionales, como presuponen las diversas formas del materialismo, pero estas relaciones presuponen las que establece el sujeto desde su propio punto de vista, al dirigir su atención a ciertos aspectos del mundo común a todos. El sujeto preserva su autoridad sobre lo que percibe, siente o desea en virtud de los aspectos del mundo a los que se dirige su atención, pero no en virtud de su relación inmediata con objetos y propiedades fenoménicas, interpuestas entre él y el mundo. Sin embargo, el desarrollo de estas sugerencias nos llevaría más allá de los límites del presente trabajo.⁵⁰

⁴⁹ Tal vez, sin embargo, el caso del dolor presente peculiaridades que lo hacen difícil de integrar en la perspectiva propuesta. Por otra parte, mi propuesta no resuelve con claridad la cuestión de la autoridad del sujeto sobre sus *actitudes*, en cuanto que éstas se distinguen de sus *objetos*.

⁵⁰ He ofrecido un análisis más detallado de esta propuesta en algunos trabajos, no publicados: 'Mind and the first person', 'Intentional attitudes: first and third person' y 'Externalismo, naturalismo y autoridad de la primera persona'. Todos ellos fueron presentados a distintas audiencias.