

PANORAMA ACTUAL DE LA FILOSOFÍA ANALÍTICA DE LA MENTE: FUNCIONALISMO Y EXPERIENCIA

Vicente Sanfélix Vidarte

I. Introducción

Según reza el programa de estas sesiones, me toca a mí bosquejarles el panorama actual que la tradición analítica presenta en ese campo al que normalmente se denomina «filosofía de la mente», «psicología filosófica» o «filosofía de la psicología». Antes de nada, quisiera hacer algunas precisiones de cara a que mi tarea quede claramente circunscrita.

La primera de ellas es que no deja de resultar problemático hablar del panorama actual de la filosofía analítica en cualquier campo de investigación, dado que es harto discutible que hoy en día exista la filosofía analítica. En realidad, no estoy muy seguro de que haya existido nunca. Lo más que me parece indiscutible es que ha habido distintos filósofos - Moore, Russell, Wittgenstein, Carnap, Ryle,... y otros muchos más a los que se suele calificar de «analíticos». Y que sigue habiendo filósofos -Putnam, Davidson, Armstrong, Shoemaker... y otros muchos- cuyo pensamiento se ha construido fundamentalmente en diálogo y discusión con aquellos. Cuando hablamos, pues, del panorama actual de la filosofía analítica, debe entenderse que estamos utilizando una expresión más o menos cómoda para referirnos a las posiciones de una serie de autores que sería más exacto calificar como los descendientes de los filósofos analíticos, y que no tienen por qué compartir, ni por lo general comparten, ni un conjunto de tesis, ni un método, ni tan siquiera una concepción de lo que la filosofía sea o deba ser.

La segunda precisión es relativa al tipo de problemas que voy a tratar. Generalmente los libros cuyos propios autores califican como de filosofía de la mente (o psicología filosófica, o filosofía de la psicología) presentan una gran variedad de temas: la naturaleza de lo mental y su relación con el cuerpo, la identidad personal y el autoconocimiento, el problema de las otras mentes, de la percepción, el carácter de la acción intencional, de la comunicativa, el problema del libre albedrío... Presentar el estado de la discusión de todos estos temas es algo que no sólo sería incapaz de hacer por falta de tiempo sino también, me temo, por falta de competencia. De manera que me limitaré a hablar de la situación de uno sólo de estos temas, el más general y a la vez central de la filosofía de la mente, aquel que ningún libro que verse sobre la misma puede obviar: el de la naturaleza de lo mental, también conocido en la literatura analítica como el problema mente-cuerpo. Obviamente se trata de un tema metafísico, pero supongo que a nadie se le escapará que tiene connotaciones epistemológicas y metodológicas claras, y que la manera como se afronte está en función, entre otras cosas, de los presupuestos lógicos de los que partamos.

Hechas estas precisiones, voy a intentar presentar la manera como los filósofos descendientes de los filósofos analíticos han afrontado el problema mente-cuerpo.

II. Origen y características generales del funcionalismo

Sin embargo, así circunscrito, el problema parece muy asequible. Se trataría simplemente de discutir la viabilidad del funcionalismo, pues el funcionalismo, si no una posición universalmente compartida entre los actuales filósofos de la mente, es, sin duda, la tesis más difundida entre los mismos (Dennett, 1978d. Fodor, 1985. Putnam, 1988), de manera que la actual discusión sobre el problema mente-cuerpo parece concretarse en una discusión sobre las virtudes o/limitaciones del funcionalismo.

El problema estriba en que, como ya pasara con otras doctrinas de lo mental que en un momento u otro llegaron a convertirse en oficiales, tales como el dualismo, el conductismo, o el materialismo, bajo un rótulo general se esconden en realidad una enorme disparidad de puntos de vista (Block, 1980b), de manera que nada impide a un autor ser funcionalista en un sentido -por ejemplo, en el de reconocer que el funcionalismo es el único enfoque que suministra a la psicología una estrategia de investigación coherente- y no serlo en otro -negando, por ejemplo, que un análisis funcional de nuestros predicados psicológicos pueda hacer justicia al significado de los mismos (a mi entender, esta es probablemente la posición de Stich, 1983)-. Debiéramos empezar, pues, por aclarar lo que se entiende por «funcionalismo», y para ello, como siempre que se trata de analizar una posición filosófica históricamente concretada, nada mejor que remontarnos un poco a su génesis.

En realidad, el funcionalismo puede verse como la respuesta lógica a las insuficiencias que presentaban las doctrinas de lo mental más aceptadas antes de él: el conductismo lógico y el materialismo reductivo, o tesis de la identidad de los (tipos de) estados mentales con (tipos de) estados físicos. Estas insuficiencias son múltiples y pueden clasificarse de distintas maneras, pero para nuestros actuales propósitos bastará que digamos que eran básicamente de dos tipos: intuitivas y lógicas. Veamos.

Aunque el rótulo «conductismo lógico» se suele aplicar a las posiciones muy diferentes de autores muy distintos (Carnap, 1932; Hempel, 1935; Ryle, 1949; Wittgenstein, 1953; Strawson, 1959), muchos de los cuales, por ende, habrían rechazado tal calificativo caso de que se les hubiera propuesto, la tesis básica de esta posición puede sintetizarse diciendo que según la misma el significado de los predicados psicológicos debe ser analizado haciendo referencia no a procesos o entidades internas y privadas, sino a un conjunto de conductas públicamente observables efectivas o potenciales. Desde este punto de vista, decir de alguien que piensa que está lloviendo, o que tiene dolor de muelas, no es atribuirle ningún tipo de proceso interno o de experiencia privada, sino decir que se comporta o está dispuesto a comportarse de cierta manera, de forma análoga a como -por utilizar un ejemplo tópico de la literatura conductista- decir que el azúcar es soluble en el agua no es sino decir que si ponemos un terrón de azúcar en un vaso de agua, se disolverá.

Las objeciones intuitivas y lógicas del conductismo lógico, desarrolladas con profusión por los primeros funcionalistas (Putnam, 1965; Chichara y Fodor, 1965; Fodor, 1968; Armstrong, 1968), son obvias. Para empezar parece claro que cuando hablamos de procesos y de eventos mentales lo hacemos muy en serio. Cuando decimos que alguien está pensando, lo que decimos es exactamente eso: que está pensando; o sea que está realizando un proceso mental, y no simplemente que está dispuesto a contestar nuestras preguntas de cierta manera, o a comportarse de cualquier otro modo. Y de la misma manera, cuando decimos

de alguien que tiene dolor lo que decimos es algo muy diferente a que se queja o que está dispuesto a quejarse; lo que estamos haciendo es atribuirle algo más sustantivo: una sensación dolorosa. Que lo esencial en nuestra atribución de predicados psicológicos son, precisamente, los procesos y eventos internos que el conductista quiere dejar fuera de su significado es algo que nos parece muy claro desde el momento en que su inexistencia haría falsas nuestras atribuciones de los mismos a cualquier sujeto. Por más que un individuo se comporte como si estuviera pensando, o como si tuviera dolor, si no está realmente pensando o teniendo dolor será falso decir de él que esté pensando o teniendo dolor. Pero si la existencia de los procesos y los eventos internos es tan decisiva para las condiciones de verdad de las proposiciones en las que intervienen predicados psicológicos, difícilmente será creíble que no tenga nada que ver con el significado de los mismos. Puede resultarnos más o menos creíble la existencia de un proceso o evento mental que no tenga manifestación conductual alguna (Putnam, 1965), pero lo que es intuitivamente inadmisibles es que la corrección en la atribución de ciertos predicados psicológicos no dependa para nada de la existencia de ciertos procesos y eventos mentales.

Emparentada con esta intuición está una objeción lógica que los funcionalistas han solido hacer contra el conductismo (Armstrong, 1968; 1980), a saber: haber entendido de una manera errónea el concepto de disposición. En efecto, al pretender negar la significatividad de los procesos y eventos internos los conductistas se han visto obligados a concebir operacionalmente su naturaleza disposicional. Una disposición, según este punto de vista, no sería sino el conjunto de comportamientos que la actualizan. Pero, obviamente, es ésta una concepción muy reductiva de la naturaleza de las disposiciones, una concepción que sólo podría defenderse si adoptáramos una teoría verificacionista del significado difícilmente justificable. Y si no lo hacemos todo parece indicar como mucho más sensato optar por una concepción realista de las disposiciones, según la cual éstas son diferentes de y explican causalmente a, sus manifestaciones. Del mismo modo en que debemos entender la propiedad de la solubilidad del azúcar como una propiedad intrínseca del azúcar -su peculiar estructura molecular- responsable de su disolverse en el agua, debemos entender el dolor como un peculiar estado del sujeto causalmente responsable de su actuar de cierta manera.

Ahora bien, los estados y procesos mentales no son sólo causalmente eficaces respecto a la conducta. A veces también tienen como consecuencia la aparición o desaparición de otros estados o procesos mentales, como cuando mi llegar a creer que en Brasil es muy fácil pasárselo bien me hace desear ir allí, pero mi pensar en lo mal que lo pasaré durante el viaje en avión, me hace desistir de la idea. Y con esta interconexión causal de los eventos y procesos mentales está íntimamente relacionada la otra objeción clásica que los funcionalistas han desarrollado contra el conductismo (Fodor, 1968; 1975). Como hemos señalado, los conductistas lógicos opinaban que el significado de los predicados psicológicos podía ser adecuadamente analizado en términos de un conjunto de proposiciones que describieran conductas efectivas o potenciales del sujeto al que se adscriben. Sin embargo, no deja de ser notorio el que ningún conductista lógico haya suministrado nunca un ejemplo, ni tan siquiera aproximado, de un análisis tal para ningún predicado psíquico, por elemental que éste sea. Podría pensarse que quizás la razón de ello estribe en que, como señalaron los conductistas, el tipo de disposición que denotan los términos psicológicos es altamente

complejo. Pero ya Chisholm (1957) apuntó a una razón mucho más convincente de por qué tales reducciones de los predicados psicológicos a descripciones conductuales sería imposible, a saber: que para que una conducta cuente como expresión de determinado predicado psicológico debemos asumir que el sujeto que la lleva a cabo está en otros estados psicológicos diferentes del que pretendemos definir conductualmente (Putnam, 1962; Dennett, 1971; 1978b; Stich, 1983; Churchland, 1984; etc.). Por poner un ejemplo de nuevo típico: si yo pretendo establecer el conjunto de comportamientos que corresponden a la creencia de que llueve el problema es no sólo que la lista sería posiblemente infinita (cerrar las ventanas de casa, recoger la ropa, advertir a mi madre de que no olvide el paraguas, coger el coche para ir a la Facultad, etc.) sino que además todas y cada una de estas conductas sólo cuenta como una expresión adecuada de aquella creencia si se supone que son ciertos de mí otros predicados mentales, aquéllos por ejemplo que me atribuyen ciertos deseos (hago todas esas cosas porque, además de creer que llueve, no deseo que entre agua en mi casa, ni que se moje o ensucie la ropa tendida, ni que mi buena madre coja una pulmonía, etc.).

Hemos visto ahora los defectos que los funcionalistas han encontrado más usualmente en las tesis del conductismo lógico; este trabajo de desescombros nos permite, no obstante, avistar el terreno donde se asentarán los fundamentos del funcionalismo. A diferencia del conductismo, no negará sino que subrayará la realidad de los procesos y los eventos mentales, dotándolos de eficacia causal no sólo respecto a la conducta sino respecto a sí mismos. De esta manera el funcionalismo volverá a una forma de mentalismo, pero debemos tener bien claro que, como advierte Fodor, 1968; no por ello debe volver necesariamente a postulados dualistas.

Podemos pasar ahora a analizar la otra filosofía de la mente en contraposición a la cual nace el funcionalismo: el materialismo reductivo o la tesis de la identidad. A diferencia del conductismo lógico, el materialismo reductivo tuvo un periodo de gestación muy breve y su vigencia fue también efímera. Sin embargo, haríamos mal si menospreciáramos su importancia, pues aparte de que su intuición básica de que los procesos y eventos mentales son procesos y eventos materiales está hoy más vigente que nunca, su formulación dió origen a una intensa discusión que sirvió para plantear claramente los problemas que en la actualidad dominan el campo de la filosofía de la mente. Los trabajos pioneros de Place, 1957; Feigl, 1958; Smart, 1959; 1962; así como los hitos más importantes de esa discusión aludida: Shaffer, 1962; 1963; Cornman, 1962; Feyerabend, 1963; Nagel, 1965; Rorty, 1965; Malcolm, 1964; etc., han sido profusamente recopilados, siendo quizás las mejores selecciones la de O'Connor, 1969; y la de Borst, 1970.

Por contraposición a muchos de los conductistas lógicos, cuya militancia filosófica se situaba en lo que se conoce ahora como la escuela del análisis del lenguaje ordinario, los materialistas reductivos se movían prácticamente todos ellos en las coordenadas del realismo científico (Armstrong, 1983); lo que quiere decir que entendían que la ciencia natural suministra un esquema conceptual privilegiado para comprender la realidad, a la vez que constituye un polo de referencia obligado en el trabajo del filósofo. Y precisamente de la historia de la ciencia natural sacaron la idea básica que articula su posición.

En efecto, la historia nos suministra ejemplos de reducción teórica de entidades y propiedades categorizadas en el esquema del sentido común mediante la postulación de su

identidad con entidades y propiedades cuya existencia se admite en el cuerpo de una teoría científica. Todos tenemos en mente los ejemplos del agua, reducida o identificada con el H_2O , o del rayo, identificado con una descarga eléctrica, o de la temperatura de un cuerpo, que se reduce al grado de energía cinética de sus moléculas. Si estas reducciones se habían conseguido con el desarrollo de la física y la química, ¿por qué no habría que esperar que el desarrollo de la neurofisiología nos permitiera lograr reducciones semejantes de los procesos y eventos psíquicos reconocidos en nuestra psicología cotidiana?

Cuando los primeros materialistas reductivos formularon sus teorías, la neurofisiología aún distaba de ofrecer tales tipos de identificaciones reductivas -dicho sea de paso, sigue sin hacerlo y hay razones empíricas y conceptuales como para poner en duda que pueda llegar a hacerlo nunca- pero optimistas como eran respecto a las potencialidades de la ciencia, pensaron que todo era cuestión de esperar. Ya había, en su opinión, evidencias suficientes de que los procesos y los eventos mentales eran procesos y eventos neurofisiológicos o, para ser más exactos, procesos y eventos del sistema nervioso central. Siguiendo su ejemplo favorito: a la pregunta «¿qué es el dolor?» lo más sensato era responder algo como «la excitación de las -hipotéticas- fibras C».

Como ocurrió con el conductismo, los funcionalistas en su gran mayoría encuentran el materialismo reductivo intuitiva y lógicamente defectivo. Veamos por qué.

Empezaremos una vez más por las objeciones intuitivas. La más usual de todas ellas ya fue formulada claramente por Putnam, 1967. Supongamos que, en efecto, la neurofisiología llega a identificar el dolor con la excitación de cierta región neuronal de nuestro cerebro; la pregunta clave es: ¿por qué habríamos de suponer que en todos los seres que pueden experimentar dolor éste se concretará en la excitación de un mismo tipo de neuronas? En realidad, sería muy sorprendente si esto ocurriera dada la gran variedad de sistemas nerviosos estructuralmente diferentes que presentan las distintas especies de seres vivos de los que de una manera u otra podemos decir que sienten dolor. Dicho de otra forma: dado que la neurofisiología identificará básicamente los estados mentales con estados de nuestro sistema nervioso, cualquier filosofía de la mente que pretenda basarse en sus resultados para desentrañar la naturaleza de lo mental estará muy expuesta a incurrir en el pecado de chauvinismo, el pecado de tener que negar que organismos a los que por su conducta estaríamos dispuestos a atribuirles procesos y estados mentales, realmente los tengan.

La validez de ésta y otras intuiciones anti-materialistas resulta muy discutible (se pueden encontrar discusiones de este tipo de objeciones en Lewis, 1969; Kim, 1972; Wilkes, 1978; 1981; Wilson, 1979; Churchland, 1979, etc.) pero como la discusión que siguió a las tesis de los materialistas acerca de si sus propuestas identidades podrían o no cumplir la ley de Leibnitz, nos puede inducir a abrigar una sospecha que puede resultar casi ineludible a la luz de ciertas argumentaciones lógicas y epistemológicas: la de que los predicados mentales no resulten proyectables en predicados físicos (sobre la noción de proyectabilidad, Cf. N. Goodman, 1965).

Esta objeción contra el materialismo reductivo ha sido muy bien elaborada por Fodor, 1975; y también por Davidson, 1970; 1973; 1974; si bien en éste último los argumentos juegan no sólo contra el materialismo reductivo sino contra cualquier tipo de reduccionismo acerca de lo mental y, en este sentido, es muy posible que estén dirigidos también contra el mismo funcionalismo, al menos en algunas de sus versiones centrales (Cf. Lycan, 1981).

A estas alturas no nos podemos permitir el lujo de entrar en detalle en estos argumentos, pero resumiéndolos podríamos decir lo siguiente. Es obvio que cada predicado que podamos utilizar es un instrumento que nos permite clasificar las entidades del mundo. El predicado «es una moneda», por ejemplo, nos permite agrupar a ciertas entidades y distinguirlas de otras. Si quisiéramos encontrar un predicado físico que se redujera a éste predicado económico, un requisito imprescindible es que fuera co-extensivo con él. Ahora bien, ¿qué tipo de predicado físico podría cumplir esta tarea? Si consideramos que una moneda puede estar hecha de cobre, o de bronce, o de plata, o de oro, o de aluminio, o incluso de papel; que puede ser circular, octogonal, cuadrada, rectangular, etc., está claro que ese predicado físico será tremendamente complejo... y en la misma medida completamente inútil para formular con él cualquier tipo de generalización nomológica. ¿Quién se atrevería a decir, vamos a suponer, que porque todas las monedas que se hubieran encontrado hasta hoy fueran o de plata, o de bronce, o de oro, que todas las monedas con que en el futuro nos encontraremos deberán ser de planta, de bronce o de oro?. Una consideración como ésta nos sirve para ver como muy verosímil la conclusión de que resultará imposible reducir los predicados económicos a predicados físicos, y por ende la economía a física. Aún si todas las monedas son entidades físicas, de ahí no se sigue que la propiedad «ser una moneda» sea equivalente a una determinada propiedad física. ¿Por qué no habría de ocurrir lo mismo con las propiedades descritas por los predicados psicológicos?. ¿Por qué no habría de ocurrir que, aunque todo estado y proceso psicológico fuera de hecho un estado y un proceso físico, la propiedad «ser un estado o un proceso psíquico» fuera diferente a cualquier específica propiedad física?. Lo que tenemos ahora es un respaldo lógico y epistemológico para nuestra intuición de que, dado que podemos predicar los mismos tipos de estados y procesos psíquicos de organismos neurofisiológicamente muy diferentes, aquellos estados y procesos no podrán identificarse con ningún estado o proceso neurofisiológico concreto.

Como ocurriría con la revisión de los argumentos funcionalistas contra las tesis conductistas, podemos sacar ahora de la crítica al materialismo reductivo una conclusión acerca del talante del funcionalismo. Este pretenderá constituir una alternativa a la vez antichauvinista y antireductiva en filosofía de la mente. Esto es, procurará dar una caracterización de lo que son los estados y procesos mentales que no nos obliga a predicar los mismos sólo de los seres humanos (Cf. Block, 1980b; Putnam, 1967b), y a la vez defenderá para la explicación psicológica una autonomía y una irreductibilidad a cualquier otro tipo de explicación científica (Cf. Haugeland, 1978; Cummis, 1975, 1983).

Sin embargo, es preciso advertir ya mismo que estas conclusiones no serían válidas para todos los autores que suelen encuadrarse bajo el rótulo común del funcionalismo. Algunos de los considerados a veces -aunque, significativamente, no por sí mismos-, como tales -el caso de Armstrong y de Lewis (Ved. Block, 1978; Armstrong, 1983)- no estarían muy de acuerdo con la conclusividad de argumentos como los esbozados arriba, de manera que en el análisis funcional de los predicados psicológicos no verían sino un aliado metodológico y epistemológico del materialismo, que permitiría a éste soslayar alguno de los escollos con que en sus formulaciones primigenias chocó. Por supuesto, estos autores tampoco compartirían la tesis de la irreductibilidad de la psicología, y seguirían centrandos sus esperanzas en el desarrollo futuro de la neurociencia. Estas consideraciones creo que pueden servir para

explicar por qué no es del todo inusual el encontrarse con autores definidamente materialistas y a los que, en consecuencia, no se les sitúa en el terreno del funcionalismo, escribiendo, no obstante, alegatos en favor de éste (Cf. los Churchland, 1981; Wilkes, 1981).

III. Variedades del funcionalismo: La tesis del funcionalismo computacional

Lo que hasta aquí sabemos es más bien poco y algo bastante general acerca del origen y las características del funcionalismo. Vamos a entrar en un poco más de detalle. Como dejamos dicho, el funcionalismo como alternativa filosófica diferenciada fue gestándose poco a poco como un intento de corregir las deficiencias tanto del conductismo lógico como del materialismo. Una objeción muy usual que los dualistas han esbozado tanto contra uno como contra el otro es lo que se denomina a veces como la objeción del esencialismo, que viene a decir: sencillamente, no podemos identificar un estado mental con un tipo de conductas o de eventos físicos por la sencilla razón de que es concebible que se dé aquél sin que se den éstos, y al revés. Este argumento, que tiene su formulación primigenia en la obra de Descartes, ha sido reformulado de una manera u otra por los dualistas contemporáneos tales como Kripke, 1971; 1972; y Nage, 1974. Tomemos uno de los ejemplos favoritos de los dualistas: el dolor. Obviamente, es esencial al dolor el que se experimente o se perciba como doloroso; pero no es de ningún modo esencial a ninguna conducta, o a ningún evento de los que ocurren en nuestro cerebro, el que sea percibido o experimentado de este modo. El *esse* de lo mental es ser percibido de cierta manera, no así el *esse* de la conducta o de las estimulaciones nerviosas, etc. (para ver algunos argumentos contra esta intuición pueden consultarse Graham, G. & Stephens, L., 1985; Nelkin, 1987). De donde se deduce que el estado mental de tener dolor, y los eventos físicos de comportarse de cierta manera, o de sufrir ciertas estimulaciones en ciertas áreas de nuestro sistema nervioso, no tienen la misma esencia. Y difícilmente podría entenderse la identidad entre entidades esencialmente diferentes (Para una discusión de los argumentos de Kripke, Cf. Feldman, 1974; Lycan, 1974).

Por otra parte, esta diferencia ontológica entre los ámbitos de lo mental y lo físico parece explicarnos ciertas diferencias epistemológicas y lógicas que se dan en nuestro lenguaje ordinario entre el discurso psicológico y el discurso sobre lo físico. Comparemos los enunciados «Tengo miedo», «Me tiemblan las manos y la voz» y «Tengo estimuladas las fibras P». Lo que nos induce a aseverar el primero de ellos es muy diferente de lo que induce a otros a referirse a nosotros con un enunciado equivalente. Pero no así en los dos casos restantes. Decimos «tengo miedo» sólo porque sentimos miedo, lo que, obviamente, no es el caso cuando decimos «tiene miedo». Pero cuando decimos «me tiemblan las manos y la voz» o «tengo estimuladas las fibras P», las razones que tenemos para hacerlo son exactamente las mismas que pueden tener otros para decir que nos tiemblan las manos y la voz, etc. A saber, el ver u oír nuestras manos y nuestras voces temblorosas, una inspección de, o una inferencia sobre, el estado de nuestras fibras P., etc. Esta asimetría que se da entre los enunciados psicológicos en primera y tercera persona, pero no entre los enunciados que describen conductas o eventos neurofisiológicos, está conectada con otra peculiaridad epistémica de los primeros frente a los segundos. La autoatribución de estados psíquicos

parece ser incorregible, pero no así la autoatribución de conductas o de estados neurofisiológicos. Y a su vez, esta peculiaridad epistémica implica ciertas peculiaridades lógicas del discurso psicológico; en concreto, condena al sinsentido preguntas acerca de la base evidencial en que nos apoyamos para autoatribuirnos estados mentales, mientras que, por contra, las preguntas equivalentes acerca de los enunciados en que nos autoatribuimos conductas o eventos neurológicos conservan su legitimidad.

La asimetría de los enunciados psicológicos en primero y tercera persona, la incorregibilidad de unos frente a los otros, y la exclusión gramatical de ciertas cuestiones (o en términos positivos, el carácter gramatical de ciertas afirmaciones -«Sólo yo puedo sentir mi dolor»-) constituyen las coordenadas de lo que se conoce como el rompecabezas de la privacidad. El reto es encontrar una solución del mismo que no nos comprometa con posiciones dualistas. Un reto que, como es bien sabido, Wittgenstein afrontó con el famoso argumento del lenguaje privado, sin que yo tenga muy claro -debo confesarlo- si lo que intentó -no digamos ya lo que consiguió- fue encontrar esa solución o, más bien, mostrar que la alternativa dualista tampoco constituía ninguna alternativa válida para encarar el mismo.

Siete años más tarde de la publicación de la obra fundamental del segundo Wittgenstein, H. Putnam afrontó en un breve pero muy importante artículo (Putnam, 1960) el mismo reto. Su estrategia era, cuando menos, mucho más fácilmente detectable que la de Wittgenstein. Se trataba de mostrar que una máquina de Turing (Cf. Kleene, 1974; cap. XIII) que pudiera autoatribuirse estados reproduciría la situación en que nos encontramos nosotros cuando nos auto-atribuimos predicados psicológicos. Los enunciados en que la máquina se autoatribuiría estados serían tan asimétricos con respecto a los enunciados que podría producir para atribuir estados a otra máquina, tan incorregibles y tan poco susceptibles de dar pie a ciertas cuestiones, como lo son aquellos de nuestros enunciados en que nos atribuimos estados mentales. Si lo que explica las características epistemológicas y lógicas de los enunciados psicológicos es, según el dualista, la peculiaridad ontológica del sujeto que los realiza, su ser algo diferente de una realidad material, la consecuencia que debiera sacar es que las máquinas de Turing no son entidades materiales. Pero si no se está dispuesto a sacar esta conclusión, si se piensa que las máquinas de Turing son simples máquinas, artefactos mecánicos o sistemas físicos, entonces la conclusión es que la privacidad no resulta incompatible con el materialismo.

En 1960 Putnam sólo intentaba mostrar que una máquina de Turing modelaba la subjetividad de la conciencia humana -una tesis que defendió de las objeciones que le fueron planteadas en 1964- con la finalidad de defender la legitimidad del materialismo en su versión de teoría de la identidad de tipos. Su estrategia consistía en mostrar que las peculiaridades epistemológicas y lógicas del discurso sobre lo mental no tenían por qué explicarse, en contra de lo que supone el dualista que apela a argumentos esencialistas, por las tesis del acceso directo o procesos semejantes. Quizás por ello, y también porque el modelo de máquina de Turing considerado era excesivamente simplificado -por poner sólo un ejemplo: se trataba de autómatas deterministas-, aunque Putnam expresó su admiración por la analogía existente entre los estados lógicos de la máquina y los estados mentales de los seres humanos, no dejó de considerarla una mera analogía, absteniéndose de dar el paso

adicional de identificar la mente con una máquina de Turing. Sin embargo, en posteriores artículos (1964; 1967a; sobre todo) Putnam consideró la posibilidad de compliar las máquinas de Turing -concebir las como autómatas probabilísticos, antes que deterministas; como sistemas dinámicos, antes que estáticos; dotarlas de una función de preferencia racional, así como de transductores u órganos sensoriales y de un sistema motor, de manera que no sólo pudieran interactuar con el medio sino hacerlo en calidad de agentes racionales- para hacerlas susceptibles de modelar ya no sólo comportamientos psicológicamente elementales, como el de la atribución de sensaciones, sino otros mucho más complejos como los de expresar preferencias y creencias. El paso ulterior de identificar los estados mentales de cualquier organismo con los estados lógicos de un cierto tipo de máquina de Turing estaba servido. Y Putnam lo dió en 1967b formulando claramente, por vez primera, las tesis fundamentales de lo que, basándonos en Block 1978, 1980b, podemos denominar como funcionalismo mecánico o, quizás mejor, computacional. No vamos a entrar por el momento en las peculiaridades de este tipo de funcionalismo. Nos limitaremos a señalar cómo se alcanza en él la tesis básica común a todo tipo de funcionalismo.

Como hemos visto, lo propio del funcionalismo computacional consiste en identificar los estados mentales de un organismo con los estados lógicos de una máquina de Turing de cierto tipo. En consecuencia, la naturaleza de los estados mentales, su esencia, será coincidente con la naturaleza o la esencia de los estados lógicos de las máquinas de Turing. ¿Cuál es ésta?

Dado que un estado queda individualizado cuando se especifican sus relaciones con la serie de posibles inputs, outputs y con el resto de estados, bien podemos responder esta pregunta diciendo que la esencia de los estados lógicos es, por consiguiente, puramente relacional. El criterio de individuación de los estados lógicos de una máquina de Turing, o lo que es lo mismo, el criterio de individuación de los estados mentales de un organismo, estriba en el rol funcional que ese estado juega en el sistema (lógico-psicológico) completo. Dos sistemas que sean funcionalmente isomórficos serán también psicológicamente equivalentes; y dos estados que sean funcionalmente equivalentes, serán estados psicológicos tipológicamente isomórficos.

Que esto constituye una respuesta nueva al enigma de la naturaleza de lo mental es obvio. La esencia de lo psíquico no estriba en ninguna de las características que señalaron los dualistas -ni en su inextensión, ni en su carácter subjetivo o privado, ni en su accesibilidad epistemológica privilegiada- ni los materialistas -no es ninguna peculiaridad física la que hace de lo mental un fenómeno diferente-, ni tan siquiera los conductistas -pues el carácter relacional de los estados psicológicos no se reduce a sus conexiones con inputs y outputs simplemente-.

Por otra parte, a esta concepción de los estados mentales llegaron otros autores por derroteros muy diferentes de los que siguió Putnam; no basándose en la analogía de las máquinas de Turing, sino en un análisis del significado de los predicados psicológicos. A este tipo de funcionalismo -defendido sobre todo por Armstrong (1968, 1980), Lewis (1966, 1970, 1972), Smart (1971) y Shoemaker (1975b, 1981a, b, c)- lo denominaremos, siguiendo a este último autor, «Funcionalismo analítico». Veamos cómo se gestó para después analizar las peculiaridades que lo distinguen del funcionalismo computacional.

IV. Variedades del funcionalismo: las tesis del funcionalismo analítico

En el apartado anterior hicimos referencia a los argumentos esencialistas que filósofos como Nagel y Kripke han utilizado contra las posiciones materialistas. El valor de la crítica de este último es, sin embargo, independiente de la -a mi entender dudosa- validez de tales argumentos.

En efecto, lo que hace importantes los trabajos de Kripke (1971, 1972) es que la conclusión negativa acerca de los enfoques no dualistas en filosofía de la mente se encuentra como colofón de una magnífica aportación a los campos de la teoría del significado y de la lógica modal. No podemos entrar en el detalle de estas aportaciones. Para nuestros propósitos, basta tener presente que las conclusiones de Kripke fueron que, por una parte, los nombres no deben ser concebidos como abreviaturas de un grupo de descripciones definidas sino como designadores rígidos que refieren a la misma entidad en todos los mundos posibles en los que la misma exista y, por la otra y como consecuencia, que el status modal de las identidades teóricas era el de verdades metafísicamente necesarias.

Esta última conclusión afectaba directamente a los teóricos materialistas, dado que éstos siempre habían admitido que las propuestas identificaciones de estados, eventos o procesos psíquicos con estados, eventos o procesos físicos eran identificaciones contingentes. Luego o Kripke estaba equivocado, y cabe la posibilidad de identificaciones teóricas no necesarias metafísicamente, o si Kripke estaba en lo correcto, entonces las tesis de los materialistas eran falsas. Claro que esto último podía ocurrir por dos razones muy diferentes: bien porque las identificaciones propuestas no eran posibles -y esta era la opinión de Kripke-; bien porque esas identificaciones, siendo correctas, son no obstante necesarias, y éste es el punto de vista de los funcionalistas analíticos.

Lo que se requería para defender esta última posición era una nueva teoría de la manera como las identificaciones teóricas en general, y las identificaciones psicofísicas en particular, son llevadas a cabo. El mérito y la responsabilidad de esta nueva teoría fue básicamente de D. Lewis (1966, 1970, 1972. Aunque también son importantes a este respecto los trabajos de Chichara y Fodor, 1965; y Harman, 1973) y el *quid* de la misma estribaba en considerar de una manera nueva la relación entre términos teóricos y términos no teóricos. Veamos.

La epistemología de filiación positivista, anclada por el viejo dogma de la existencia de una distinción tajante entre el ámbito de lo analítico y de lo sintético, del significado y de lo empírico, había legado un concepto de teoreticidad por el que ésta se contraponía tajantemente a lo observacional. Según esta concepción las teorías científicas serían una combinación de términos categorialmente heterogéneos: teóricos los unos, observacionales los otros. Términos cuya diferencia estribaría en que mientras los primeros denotarían entidades, propiedades o procesos no observables y quasi-ficticios (recuérdese que, siguiendo básicamente a Russell, los positivistas lógicos pretendían, en un principio al menos, tratar a todos los términos no empíricos como *ficciones* lógicas), los segundos denotarían entidades, eventos o procesos empíricamente constatables.

Frente a esta concepción, la de Lewis es que la línea divisoria entre los términos teóricos y los no teóricos no estriba en su observabilidad o inobservabilidad, sino en el rol que juegan en nuestras teorías; en concreto la hipótesis de Lewis es que los términos teóricos son tales

porque en el seno de las teorías en los que se postulan los mismos son definidos funcionalmente, esto es: por el rol causal que se atribuye a sus *denotata* en relación a los eventos, propiedades o procesos denotados por los términos no teóricos.

En su trabajo de 1972, Lewis pone un ejemplo bastante pedestre en clave detectivesca. Para ilustrar su posición, particularmente prefiero acudir a un ejemplo extraído de la historia real de la ciencia. Después del descubrimiento de Urano, realizado por casualidad por W. Herschel en 1781, las observaciones del planeta recién descubierto empezaron a contradecir las expectativas que las tablas construidas para explicar su órbita hacían concebir. La corrección de las viejas tablas por parte de A. Bouvard, con el fin de ponerlas en concordancia con las nuevas observaciones, de nada sirvieron, pues al cabo de tan sólo 20 años la trayectoria efectivamente observada de Urano volvía a diverger significativamente de lo previsto por las tablas enmendadas. La opción que se presentaba era o bien concluir que la ley de Newton no era exacta, o presuponer la existencia de otra masa planetaria transuránica que, con su influjo, permitiera, preservando la ley, dar cuenta de las perturbaciones de la órbita de Urano. Esta última fue la opción por la que se decidió Le Verrier. Este predijo la existencia de un octavo planeta que sería *causalmente* responsable de las anomalías orbitales de Urano, y calculó su posición de manera que permitió a J. Galle descubrir, el 23 de Septiembre de 1846, la existencia de Neptuno.

De acuerdo con el punto de vista de Lewis, en la teoría de Le Verrier los únicos términos teóricos serían aquellos que utilizara para referirse al hipotético planeta responsable causalmente de los hechos a explicar; hipotético planeta que en la teoría de Le Verrier no se identificaría sino por criterios funcionales, o sea: sólo por aquellas propiedades que se considerasen necesarias como para hacerlo responsable causalmente de aquellos hechos. El resto de términos, esto es: todos aquellos que utilizamos en la caracterización de estos hechos, serían términos no teóricos (Lewis los denomina O-terms, pero advierte que la O no viene de «Observational» sino de «old», por lo que podríamos denominarlos, por nuestra parte, términos familiares). «Neptuno», o cualquier otro nombre con el que Le Verrier hubiera bautizado a su presunto planeta, o cualesquiera descripciones definidas que hubiera utilizado con la intención de referirse a él tales como «el planeta trans-uránico», o «el planeta más excéntrico del sistema solar», etc serían sus términos teóricos. Si Le Verrier hubiera vivido unos 100 años después podría haberse servido del procedimiento recomendado por Ramsey para formular su hipótesis, procedimiento que le habría permitido pasarse sin términos teóricos ya que estos quedan en este proceder sustituidos por variables ligadas a cuantificadores existenciales en las que recae la responsabilidad de referirse a las hipotéticas entidades cuya existencia resulta explicativamente fructífera de acuerdo con la teoría. En lugar, pues, de haber dicho algo como «Neptuno -o el planeta transuránico, etc-, cuya órbita y masa son tales y tales, según la ley de Newton, lo que hace que la órbita de Neptuno tenga tales y tales características observadas» podría haber dicho: «Existe un planeta x tal que su órbita y su masa son tales que, etc». Esta artimaña técnica resultará de la mayor importancia cuando pasemos a aplicar todas estas consideraciones al ámbito de la psicología. Por el momento sólo pretendemos hacer ver, merced a este ejemplo, la verosimilitud de la tesis de Lewis. Está ahora claro que los términos teóricos no son aquellos que denotan entidades, propiedades, eventos o procesos inobservables. De hecho Neptuno es observable

aunque para Le Verrier su nombre fuera un término teórico. No menos erróneo es considerar que los términos no teóricos o familiares denoten todos ellos entidades observables. Parte de los términos no teóricos para Le Verrier serían numerales, aquellos que caracterizarían la órbita de Urano. Y por muy platónico que se sea nadie está dispuesto a afirmar que los números son el tipo de entidades que podemos observar. La auténtica línea demarcatoria entre los términos teóricos utilizados por Le Verrier y los términos no-teóricos de su teoría pasa porque aquellos denotan entidades cuya existencia y propiedades son presupuestas para cumplir cierto rol en una explicación causal, de manera que podemos decir que son funcional e implícitamente definidos, mientras que este último no es el caso de los términos familiares o no teóricos. Por otra parte, la identidad de Neptuno con el pretendido referente de los términos teóricos de Le Verrier no fue una identidad postulada sino descubierta, fue el resultado de descubrir que Neptuno efectivamente realizaba la teoría de Le Verrier, lo que, llegado el caso, daría pie a que la considerásemos como una identidad metafísicamente necesaria.

¿Qué tiene que ver todo esto con la filosofía de la mente y, más concretamente, con el funcionalismo? La respuesta a esta pregunta empieza a perfilarse desde el momento mismo en que nos percatamos de que para los funcionalistas analíticos, como para algunos funcionalistas computacionales, nuestra utilización del lenguaje psicológico ordinario viene posibilitada por nuestra asunción de una teoría psicológica de sentido común (contra esta tesis argumentan Putnam 1964 y Wilkes, 1981). Según ellos lo que nos permitiría cuando observamos la conducta de un organismo atribuirle ciertos predicados psicológicos sería cierta teoría psicológica en la que aquellos predicados contarían como términos teóricos y, por consiguiente, tendrían su significado definido implícita y funcionalmente por esta teoría. Por poner un ejemplo, lo que según los funcionalistas analíticos nos permitiría atribuir a alguien que corre delante de un león estados psicológicos tales como los de ser conscientes de la presencia del león, sentir miedo y tener el deseo de salvarse, es una teoría psicológica de sentido común según la cual el ser conscientes de la presencia de un león, tener miedo y el deseo de sobrevivir son causas suficientemente explicativas de la conducta de correr delante de un león. Podríamos hacer explícita esa teoría formulando una serie de obviedades como la anteriormente citada, obviedades que, no obstante, nos permiten captar el significado de los términos psicológicos por su relación con otros estados psicológicos y con el conjunto de estímulos que los provocan y respuestas a que dan lugar. Una vez hecha explícita esa teoría podríamos transcribirla según la manera recomendada por Ramsey, de manera que desaparecieran de ella todos los términos teóricos, en este caso todos los términos psicológicos.

No obstante la importancia de esta conclusión -la eliminabilidad del lenguaje intencional- con la que se comprometen los funcionalistas analíticos, lo que nos interesa subrayar ahora es que, siguiendo un camino muy diferente al de los funcionalistas computacionales, han llegado a una conclusión muy parecida a la suya acerca de la naturaleza de los estados, procesos y eventos psicológicos. Los criterios para identificar éstos son, como los criterios utilizados para dotar de significado a los términos teóricos en general, puramente funcionales. *Prima facie* aquellos estados, procesos y eventos tienen una naturaleza puramente relacional, pues en tanto que estados, procesos y eventos psicológicos no son sino los estados, procesos y eventos causados por ciertos estímulos y que causan ciertas respuestas.

Empero, no conviene perder de vista las diferencias, serias y profundas, existentes entre las dos versiones del funcionalismo que hemos considerado. Veamos.

V. Diferencias entre las dos variantes del funcionalismo

Para empezar, es muy diferente el status epistemológico que los funcionalistas analíticos y los computacionales conceden a sus conclusiones. Para estos últimos, por lo general más cercanos a la concepción naturalizada de la epistemología en particular, y de la filosofía en general, sus conclusiones no tienen un carácter sino hipotético. Ellos no afirman que no quepa otro análisis sino el funcional de los estados mentales. En su versión más fuerte (la del primer Putnam) todo lo que dicen es que si los organismos dotados de una psicología pueden ser considerados como realizaciones de una máquina de Turing, entonces sus estados mentales funcionalmente definidos pueden identificarse tipológicamente con los estados lógicos de la máquina que realizan. Mientras que en su versión más débil este tipo de funcionalistas admite que los estados mentales no son estados lógicos de una máquina de Turing (Block & Fodor, 1972), pero insiste en que la consideración de los mismos como tales es teóricamente útil y está justificada en la medida en que existe alguna analogía interesante entre el funcionamiento de las máquinas de Turing y la mente (Cf. el mismo Putnam, en 1975; y Fodor, 1983). Así, de manera semejante a como los materialistas eliminacionistas fiaban al desarrollo de la neurociencia la justificación última de sus tesis, los funcionalistas computacionales fiarían la justificación de las suyas al desarrollo de una psicología que caracterice funcionalmente los estados, eventos y procesos psicológicos a los que apele en sus explicaciones.

Por contraposición, los funcionalistas analíticos no consideran que sus conclusiones tengan un carácter hipotético, ni deban esperar al desarrollo de ninguna ciencia para verse confirmadas. Más bien entienden que las mismas son el resultado de analizar el significado de los predicados psicológicos de nuestro lenguaje ordinario. Lo que defienden, por consiguiente, es que dado como utilizamos nuestro lenguaje, los eventos, procesos y estados mentales tienen una naturaleza funcional.

Esta divergencia es algo más que una mera divergencia metafilosófica, puesto que según la opción que adoptemos quedará limitado de una manera u otra el repertorio de estrategias argumentativas que podremos oponer al anti-funcionalista. Pongamos por caso que alguien objeta que el significado de nuestros predicados que refieren a procesos o eventos acompañados de contenido fenomenológico, no puede quedar agotado por un análisis funcional de los mismos (una objeción clásica del dualista). El funcionalismo analítico no tiene más remedio que tomarse la misma con toda seriedad e intentar demostrar por todos los medios que no es cierta. Por contra, el funcionalista computacional bien puede conceder la razón al objetor dualista (lo que no es conceder la razón al dualismo), pero añadir que su observación resulta totalmente irrelevante para la verosimilitud o inverosimilitud de su hipótesis. Y esta posible divergencia en las estrategias defensivas del funcionalismo analítico y del funcionalismo computacional frente al viejo argumento de los *qualia*, se repite de una forma u otra frente a todos los argumentos más típicos que contra el funcionalismo suelen hacerse.

La siguiente diferencia notable entre el funcionalismo computacional y el analítico es de índole ontológica, y se refiere a la diferente posición de uno y otro frente a la verosimilitud del materialismo, esto es: la verosimilitud de identificar los procesos psicológicos con las realizaciones físicas de los mismos. Como ya advertimos cuando hablamos de las características generales de las posiciones funcionalistas, los partidarios del enfoque computacional han apelado a una serie de intuiciones y de argumentos en contra del materialismo; argumentos e intuiciones a las que se han opuesto los partidarios del funcionalismo analítico. Lo que queremos añadir ahora es que esta diferencia no es de ninguna manera accidental al modelo de funcionalismo por el que se opte.

En efecto, si lo que conduce a las conclusiones funcionalistas es la comparación de la mente con una máquina de Turing, entonces el escepticismo respecto al materialismo parece inevitable, por la sencilla razón de que las susodichas máquinas no son sino autómatas abstractos, esto es: autómatas caracterizados no por ninguna propiedad material (por su *hardware*) sino única y exclusivamente por sus características lógicas (su *software*). Dicho de otra forma: lo único que se exige a un sistema para contar como una máquina de Turing, es que sea capaz de ejecutar el programa especificado por la tabla de la máquina, siendo irrelevante que ese sistema tenga una naturaleza material u otra o, incluso, el que tenga o no una naturaleza material (Putnam, 1975).

Si las mentes son (análogas a) máquinas de Turing, y éstas, como acabamos de decir, son autómatas abstractos, es obvio que la realización física que una mente vaya a tener, si es que, como la mayoría de funcionalistas computacionales piensan, va a tenerla, será completamente accidental. Añadamos a esto la premisa Kripkeana de que las identificaciones teóricas deben ser, caso de verdaderas, necesarias, y el funcionalismo computacional estará comprometido con la negación de la posibilidad de identificar los estados psicológicos por sus características físicas.

Justo lo contrario es lo que defiende el funcionalista analítico, y ello por la esencia misma del argumento que le conduce hasta sus conclusiones; pues recordemos que aunque en su opinión los términos psicológicos tal y como aparecen en nuestro lenguaje ordinario tienen un significado puramente funcional, esos mismos términos, en la medida en que no tengan sino una definición implícita en el contexto de la psicología del sentido común, no denotan sino entidades teóricas, entidades cuya verdadera naturaleza, o si se prefiere, su esencia, aún está por descubrir. Pero, añade el funcionalista analítico, ¿qué otra hipótesis sensata cabe hacer sino la de que esa naturaleza será física? Y si finalmente llegamos a descubrir que es así, como en su opinión ya lo estamos descubriendo, no cabría sino concluir que los estados mentales son estados físicos de una manera tan necesaria como el agua es H₂O.

Esta disidencia ontológica está, por otra parte, estrechamente relacionada con una divergencia de índole metodológica y, consecuentemente, con las expectativas que uno y otro tipo de funcionalismo tiene respecto a la psicología. Esto se hace muy claro desde el momento mismo en que nos percatamos de que lo que constituye el punto final de las explicaciones de los funcionalistas computacionales no sería para los funcionalistas analíticos sino el punto de partida.

Obviamente, si se adopta un enfoque computacional de los estados, eventos o procesos psicológicos, la tarea por excelencia del psicólogo no debiera ser otra que la de suministrar lo que Cummins (1975) denominó explicaciones analítico-funcionales. Se trataría de explicar procesos psicológicos complejos tales como la percepción, o el habla, o la resolución de problemas, etc., analizándolos o descomponiéndolos en procesos psicológicos más simples, de manera análoga a como el programador de un computador descompone una tarea compleja, tal como la de sumar, en una serie de tareas más simples tales como las de borrar, imprimir, etc. (de ahí que a veces se le denomine también con el rótulo de «funcionalismo homuncular»)... El funcionalismo computacional respalda así los enfoques cognitivos de la psicología que, como el de Neisser 1967, defienden la especificidad e irreductibilidad de ésta a disciplinas como la neurofisiología.

Por su parte, los funcionalistas analíticos debieran mostrar más entusiasmo por enfoques, como el de Lindsay y Norman, 1977, en los que las explicaciones analítico-funcionales de los procesos psicológicos son constantemente complementadas con explicaciones de la base física de esos procesos. O dicho de otra forma: los funcionalistas analíticos debieran mirar con mayor simpatía aquellas investigaciones psicológicas que no utilizaran una estrategia estrictamente top-down como la recomendada por sus colegas computacionales y que estuviera abierta a los enfoques bottom-up. Ello es así porque, recordémoslo una vez más, para este tipo de funcionalistas los análisis funcionales de los procesos psicológicos no constituirían sino el primer paso de la investigación, un primer paso que queda incompleto en tanto no se complementen con un análisis de la estructura física de los mismos. De esta manera, los funcionalistas analíticos defenderían, cuando menos, la necesidad de complementación de la psicología por la neurofisiología.

VI. Los problemas del funcionalismo

Expuestas las tesis funcionalistas en sus principales variantes, nos queda por analizar ahora los escrúpulos que dificultan la marcha del teórico funcionalista. Hay que advertir, para empezar, que muchos de esos escrúpulos han sido esbozados por los propios funcionalistas, y que las consecuencias de los mismos han sido muy diferentes en unos casos y otros. A algunos -el caso de Putnan 1988- las dificultades planteadas les han llevado a desdecirse de sus posiciones funcionalistas, mientras que a otros -Fodor es el ejemplo más notable- esas mismas dificultades les han llevado, simplemente, a rebajar las pretensiones del enfoque funcionalista, aunque siguen defendiendo que el mismo es el más adecuado.

En líneas generales podríamos decir que el reto que se presenta al funcionalismo es el de hasta qué punto su comprensión de los eventos, procesos o estados psicológicos se ajusta a nuestra concepción preteórica u ordinaria de los mismos. Un desajuste grave entre una y otra invalidaría *ipso-facto* las pretensiones del funcionalismo analítico dado que, recordémoslo una vez más, éste pretende ser el resultado de un análisis del significado de nuestro actual lenguaje psicológico en el que aquella concepción viene expresada. Pero aún si no tan drásticas, las consecuencias que tendría para el funcionalismo computacional, autoconcebido como una propuesta empírica, serían también graves. Pues si éste concluye proponiendo un enfoque de los fenómenos psicológicos que se aparte radicalmente del

sentido común, inmediatamente se le plantearán una serie de difíciles interrogantes como, por ejemplo: ¿cómo entender la relación entre el enfoque funcional y el propio del esquema conceptual ordinario? ¿Habría acaso que conceder la razón al materialismo eliminacionista (Quine, 1951; Feyerabend, 1963; Rorty, 1965; Churchland, 1984) y concluir que los conceptos con los que funciona tal esquema son radicalmente erróneos, de manera que debemos prescindir por completo de ellos? Y si es así ¿desde qué nueva óptica se estaría concibiendo la acción humana? ¿Y cuáles serían los intereses teóricos y prácticos de la misma? Por poner un ejemplo especialmente debatido en la literatura reciente (Armstrong, 1973; Loar, 1981; Lycan, 1981; Dennett, 1981, 1982; Stich, 1983; Baker, 1985; Fodor, 1987, etc.) consideraremos el caso de la creencia.

Muchas de las explicaciones psicológicas que damos de nuestro comportamiento o del de nuestros semejantes apelan a la atribución a nosotros mismos o a ellos de unas u otras creencias. Los intereses teóricos y prácticos de tal proceder explicativo son obvios: no sólo nos permite presentar nuestra conducta, o ver la de los demás como coherente, sino que también abre la posibilidad de intervención racional sobre la misma. Nuestra creencia de que el líquido que contiene el vaso que vemos es un refresco, explica, dada nuestra sed, la pretensión de bebérselo. Pero también permite a nuestros amigos hacernos desistir de tal empeño informándonos de que, en realidad, lo que el vaso contiene es cianuro.

¿Qué pasaría ahora si, como muchos creen, nuestro concepto de «creencia» se muestra refractario a un tratamiento estrictamente funcional? ¿Acaso debiéramos decir que una explicación psicológica que aspire a ser científica tendrá que prescindir de este concepto? ¿Debíramos, en consecuencia, negar toda validez a nuestras explicaciones psicológicas cotidianas y negar que es falso que nadie tenga creencias?

Y aunque no lleguemos a conclusiones tan radicales, aunque nos limitemos a decir que la explicación psicológica científica no apelará al concepto de creencia sin por ello negar validez al uso de tal concepto en las explicaciones psicológicas ordinarias, ¿bajo qué descripción de nuestra conducta explicará entonces la teoría psicológica construida en términos estrictamente funcionales la misma? ¿Qué tipo de comprensión y qué posibilidad de intervención sobre ella nos permitirá?

No es mi intención en este punto afirmar dogmáticamente que el funcionalista carecerá de respuestas satisfactorias a estos interrogantes, sino sólo justificar por qué la concordancia del análisis funcional con nuestra concepción pre-teórica del ámbito psicológico supone un desafío para cualquier variante del funcionalismo, pues caso de no darse la misma quedaría invalidada automáticamente la versión analítica del mismo y, como mínimo, claramente cuestionada la plausibilidad de la versión computacional.

Pero si éste es el problema general que el funcionalismo debe afrontar, es obvio que la gran multiplicidad de predicados psicológicos que el lenguaje ordinario pone a nuestra disposición lo hace prácticamente intratable así formulado. No podemos pronunciarnos sobre la viabilidad de un análisis funcional de todos y cada uno de los particulares conceptos psicológicos que utilizamos normalmente dado el elevado número de los mismos. Lo que sí podemos es, si conseguimos establecer una clasificación general de los estados, eventos o procesos denotados por esos conceptos, considerar la posibilidad de éxito del enfoque funcionalista en el tratamiento de los tipos que nuestros criterios taxonómicos establezcan;

una estrategia que bien puede llevarse a cabo analizando el caso de algunos conceptos que consideremos más paradigmáticos de cada uno de estos tipos.

A este respecto, inspirándonos en Wittgenstein (1967; 1980) podríamos distinguir entre aquellos predicados psicológicos a cuyos *denotata* parece consustancial un contenido fenomenológico, y aquellos a los que no. De esta manera los predicados perceptivos o sensoriales, tales como «ver», «oir», «sentir dolor», etc. se incluirían dentro del primer tipo, y muchas de las denominadas actitudes proposicionales, tales como «creer», «desear», «pensar», «juzgar», etc., en el segundo.

La problemática que rodea al funcionalismo puede redefinirse ahora como la de determinar la medida en que el mismo puede ofrecer un análisis de los estados, eventos o procesos psicológicos con contenido fenomenológico y de las actitudes proposicionales carentes del mismo.

Aún así reconducida, queda fuera del alcance de este trabajo un análisis pormenorizado de la misma. Lo que resta lo dedicaremos a considerar la viabilidad de un análisis funcional del primer tipo de conceptos psicológicos.

VII. Funcionalismo y *qualia*

Nuestros criterios para atribuir procesos sensoriales y perceptivos son múltiples. La pertinencia de las causas que los provocan, el tipo de conducta que causan, las certezas o creencias, así como otras actitudes proposicionales, que los acompañan, son relevantes a la hora de hacer tal atribución. Pero aparte de todo esto cuando atribuimos a un sujeto alguno de los predicados que se refieren a estos procesos parece que le atribuimos algo más: la conciencia de una determinada experiencia cualitativa (la conciencia de un *sense-datum*, de un *quale*, de un *raw-feel*, etc.).

Consideremos el caso del dolor. El mismo tiene unas causas típicas (una herida, la disfunción de algún órgano corporal...) y una conducta que le es propia (quejarse); el sujeto que lo sufre tiene certeza de padecerlo y, generalmente, de su localización, así como el deseo más o menos manifiesto de dejar de tenerlo. Pero además tiene una experiencia cualitativa dolorosa.

Mutatis mutandis lo mismo podríamos decir de cualquier proceso perceptivo: de la percepción visual, por ejemplo. Para decir de alguien que ha visto X tenemos en cuenta la posibilidad de que X haya sido la causa de su proceso perceptivo, el conjunto de creencias que ha adquirido mediante tal proceso sobre X, la conducta discriminativa que a consecuencia del mismo puede exhibir, etc. Pero también suponemos que ha tenido una experiencia cualitativa; una experiencia visual.

Es esta experiencia o conciencia cualitativa, que distingue a los procesos perceptivos y sensoriales de la mayoría de esos otros eventos psíquicos que son las actitudes proposicionales, la que mayores quebraderos de cabeza da al teórico funcionalista. Para empezar, ¿cómo podríamos reconstruir esta diferencia en términos estrictamente funcionales? ¿Puede el análisis funcionalista justificar una taxonomía de los estados mentales que, como ésta, nos parece tan obvia?.

Muchos han respondido negativamente a esta pregunta, y como justificación de su pesimismo han añadido la observación de que muchas veces somos capaces de discriminar

estados por su contenido fenomenológico aún antes de que seamos conscientes de ninguna diferencia funcional entre los mismos, lo que parece reforzar la conclusión de que lo que identifica al menos a algunos estados psicológicos no es su rol funcional sino algo completamente diferente del mismo, a saber: su contenido cualitativo.

La conclusión de la inapresabilidad de este carácter fenomenológico de algunos de nuestros estados mentales por el análisis funcional de los mismos se ve reforzada además por la consideración de ciertos experimentos mentales, tales como el famosísimo del espectro invertido (cif. Shoemaker, 1981c). ¿Qué ocurriría si mañana alguien se levantara y comprobase que todos los objetos que hasta hoy le producían la sensación del color rojo sistemáticamente le producían ahora la sensación de verde, y viceversa?. Lo razonable parece que sería esperar que tras un período de adaptación el sujeto afectado por esta inversión de sus sensaciones cromáticas consiguiera adaptar su conducta a la nueva situación, lo que demuestra que la sustancialidad cualitativa de ciertos estados mentales es diferente e independiente del rol funcional que los mismos puedan desempeñar. Pero si esta independencia existe ¿por qué admitir sólo la posibilidad de que dos estados funcionalmente isomórficos puedan ser cualitativamente diferentes?; ¿por qué, siendo consecuentes, no admitir también la posibilidad de que dos estados puedan ser funcionalmente equivalentes, y que, sin embargo, mientras a uno de ellos le corresponde un estado cualitativo el otro no tenga asociado ninguno? He aquí el argumento denominado del *quale* ausente.

De acuerdo con los Churchland (1981) podríamos desgranar el desafío que los *qualia* representan para el funcionalismo en cuatro problemas: el de justificar funcionalmente, o al menos explicar cómo es posible la taxonomía común de los estados mentales que distingue aquellos dotados de contenido fenomenológico de aquellos que carecen de él; el de explicar cómo es posible la discriminación de estados al margen de sus características funcionales; el problema de la inversión del espectro cualitativo y, por último, el problema de la total ausencia cualitativa. Hay que decir que los autores funcionalistas no han prestado la misma atención a todos estos problemas y que, como era de esperar dada la disparidad de posiciones que se esconde bajo el rótulo común de funcionalismo, han afrontado los mismos con estrategias harto diferentes. Veamos.

Si fuéramos funcionalistas podríamos, inspirándonos en Fodor y Block (Fodor y Block, 1972; Block, 1978; Fodor, 1981a; etc.) tomar una alternativa que cortara por lo sano. Podríamos reconocer que no cabe, apelando a criterios funcionales, construir una taxonomía de los estados mentales que se asemeje para nada a lo que hemos denominado la taxonomía común. Ni que tampoco podemos dar cuenta, en términos funcionales, de nuestra capacidad para discriminar cualitativamente algunos estados psíquicos de otros. Ni, en definitiva, dar cuenta en absoluto de la presencia o la ausencia de los *qualia*. Pero una vez admitido todo esto podríamos añadir que ello no resulta de la menor importancia, dado que las únicas teorías psicológicas fructíferas serán aquellas que caractericen a los estados psicológicos en términos estrictamente funcionales. Dicho más sencillamente, la réplica fodoriana al problema de los *qualia* consistiría en, aún reconociendo la dimensión fenomenológica de algunos de nuestros estados mentales, señalar la intrascendencia de la misma para la construcción de una teoría psicológica adecuada. Que la psicología no explique alguna de las propiedades de los fenómenos que estudia no es, por otra parte, algo sorprendente; al fin

y al cabo ninguna ciencia explica todas las propiedades de los fenómenos de su dominio, y hay propiedades de las cosas que no parecen ser de interés para ninguna ciencia. Bien pudieran ser los *qualia* uno de estos fenómenos científicamente irrelevantes.

En principio, tal estrategia defensiva del funcionalismo suena plausible, pero debemos ser conscientes del precio que nos exige pagar. Y éste no es otro que el de reconvertir el funcionalismo de una filosofía de la mente en una filosofía de la psicología. El funcionalista no nos estaría diciendo qué es la mente, sino sólo cómo la concebirá la psicología. Claro que el funcionalista partidario de esta opción podría añadir que tal precio es bastante módico, dado que eso de la filosofía de la mente sería una quimera tan inalcanzable como cualquier otra empresa metafísica. Quienes pretenden decir qué es la mente no estarían diciendo sino cómo se entienden los fenómenos psicológicos desde el punto de vista del sentido común. También ellos estarían haciendo filosofía de la psicología, sólo que no de la psicología científica sino de la cotidiana.

Sin embargo, y aún aceptando estas consideraciones metafísicas, la estrategia Fodoriana resulta discutible por razones puramente intrínsecas. Para empezar, no creo que pueda decirse sin más que el problema de los *qualia* es irrelevante para la psicología. Desde luego no lo ha sido (véase Hamlyn, 1957) y creo evidente que entre una teoría psicológica que pudiera explicarnos las peculiaridades fenomenológicas de algunos de los procesos mentales, y otra que no nos suministrara la menor idea al respecto, nos quedaríamos indudablemente con la primera. Además me parece que el tema de las propiedades fenomenológicas resulta difícilmente soslayable cuando se estudian ciertos procesos psíquicos tales como el de la percepción, aunque sea con una óptica cognitivista, dado que una explicación homuncular del mismo parece exigir una definición respecto a las posibles o efectivas interacciones de los sistemas cognitivos periféricos (transductores, módulos sensoriales o como quiera denominárselos) con los sistemas cognitivos centrales (memoria, resolución de problemas, etc.), definición que difícilmente podrá darse sin, a su vez, posicionarse sobre la naturaleza de los out-puts de aquéllos. A este respecto resulta instructivo el caso del propio Fodor (1983), quien enfrentado al problema de determinar cuál es la naturaleza de los productos de los sistemas de entrada se ve obligado a recurrir al criterio de «accesibilidad fenomenológica» para zanjar la cuestión.

Pero si la estrategia Fodoriana no nos convence -y a mí, desde luego, no me convence- aún podemos recurrir a otras estrategias de características similarmente evasivas, utilizadas por autores particularmente funcionalistas o filomaterialistas (Dennett, 1978a; los Churchland, 1981; Putnan, 1981; Davidson, 1988) para soslayar el problema de los *qualia*. En líneas generales la misma consiste en insistir en que, lejos de lo que presupone el dualista, los *qualia* distan de ser entidades perfectamente definidas y, por consiguiente, no puede apelarse a ellos para explicar el significado de los predicados psicológicos -hasta aquí el acuerdo de todos estos autores- quedando éste determinado por criterios funcionales -y ésta es la tesis específica del subgrupo de los funcionalistas-.

En efecto, el dualista parece comprometido con la tesis de que lo que determina la aplicación de un mismo predicado a dos estados o eventos psíquicos es la semejanza que pueda existir en su carácter cualitativo. Pero tal condición no se cumple ni tan siquiera en el caso de aquellos estados o eventos a los que nadie osaría negar un componente

fenomenológico. Consideremos el caso del dolor. ¿No existen acaso una gran cantidad de diferencias cualitativas entre las experiencias a las que catalogamos como dolorosas?. Y justo lo inverso podemos decir de otros fenómenos psíquicos como las emociones, por ejemplo. Muchas de ellas van acompañadas de sensaciones particulares, pero una y la misma de estas sensaciones puede acompañar a emociones muy distintas. Por otra parte, es un lugar común desde hace mucho tiempo en la filosofía de la percepción (Quinton, 1955) que, bajo condiciones normales, no prestamos la menor atención al específico carácter cualitativo de nuestras experiencias sensoriales cuando atribuimos cualidades objetivas a las cosas. Como se encarga de recordarnos Putnan, 1981, ¿podríamos asegurar que nuestras experiencias fueron semejantes cuando en dos ocasiones diferentes vimos el mismo mantel rojo?.

Las conclusiones lógicas y epistemológicas que se siguen de estas observaciones parecen claras. El contenido intencional de nuestros procesos perceptivos o sensoriales no parece poder explicarse por el contenido cualitativo de los mismos. Los criterios de significatividad de nuestros predicados psicológicos no parecen coincidir con las semejanzas y desemejanzas cualitativas de las experiencias que puedan acompañar a los estados que denotan, y lo que nos legitima a utilizarlos no es, tampoco, la consideración de la naturaleza cualitativa de tales experiencias. Proponer el rol funcional que tales estados juegan como candidato para desempeñar este rol lógico-epistémico es un paso adicional posible, aunque no obligatorio.

A mi entender, hay algo de correcto en estas observaciones, pero dudo que sea lo suficiente para librar al funcionalismo de la objeción de los *qualia*. Pues, al fin y al cabo, ¿qué es lo que todas estas observaciones demuestran?.

Volvamos al caso del dolor. Estrictamente, lo que debiéramos concluir de cuanto se nos ha dicho es que cuando nos autoatribuimos sensaciones dolorosas, por lo general no discriminamos muy sofisticadamente sus características cualitativas. Pero decir esto implica conceder que tales características existen y que, aún si de hecho no nos dedicamos a discriminarlas muy detenidamente, tales discriminaciones son posibles -de lo contrario, ¿cómo sabríamos que lo que llamamos doloroso es muy diferente en unas ocasiones y en otras?-.

Por otra parte, aún cuando sea cierto que las finas discriminaciones del carácter cualitativo de nuestras experiencias estén por lo general de más, no es menos cierto que una discriminación, por muy grosera que sea, de la naturaleza cualitativa de las mismas es absolutamente imprescindible en la aplicación de algunos de nuestros predicados psicológicos. ¿Acaso si no experimentásemos sensación alguna nos atreveríamos a decir que estamos emocionados?. Y si un estado no lo experimentásemos con ciertas propiedades ¿nos atreveríamos a calificarlo como doloroso?. ¿Y qué decir de nuestros juicios perceptivos?. Si mis sensaciones cromáticas del mantel no hubieran sido estrechamente semejantes, ¿habría podido reconocerlo en las dos ocasiones en que lo vi como el mismo mantel rojo?.

Estas últimas cuestiones me parecen imponer una matización sobre las observaciones escépticas que los funcionalistas y ciertos materialistas hacen acerca de la determinación del carácter cualitativo de nuestras experiencias psíquicas, y del papel que el mismo juega en la aplicación de nuestros predicados psicológicos. No creo que el grado de semejanza o

desemejanza cualitativa de nuestras experiencias sea igualmente irrelevante en todos los casos. Posiblemente, el mismo tenga muy poca importancia en el caso de la auto-atribución de emociones, pero es mucho más relevante a la hora de auto-atribuirnos estados dolorosos, y fundamentalmente cuando hacemos juicios perceptivos visuales.

El funcionalista puede replicar en este punto que en cualquier caso, y aún concediendo la existencia de un contenido cualitativo dotado de propiedades intrínsecas, el mismo es lógica y epistemológicamente parasitario del contenido intencional de nuestros procesos perceptivos y sensoriales. Dicho de otro modo, que sólo podemos discriminar y describir las propiedades fenomenológicas de nuestra experiencia si ya somos capaces de discriminar y describir las propiedades empíricas y objetivas de los eventos o las cosas del mundo, pues cuando decimos, por ejemplo, que experimentamos un *qualia* visual rojo lo que estamos expresando abreviadamente es, en realidad, algo como que nuestra experiencia es del tipo de las que normalmente nos causa la percepción de las cosas rojas.

Estoy por completo de acuerdo con esta observación, pero a mi entender habría que señalar también que aquí el orden ontológico es el inverso del lógico y del epistémico, y que la existencia y naturaleza de nuestra experiencia cualitativa determina los posibles contenidos intencionales de nuestros procesos perceptivos y sensoriales. Consideremos el caso del daltónico. A quien padece una enfermedad semejante le están vedados los procesos perceptivos visuales cuyo contenido intencional consista en discriminar entre el color rojo y el verde. La única hipótesis verosímil al respecto es que tal deficiencia no lo es de aprendizaje -no depende, por consiguiente, del funcionamiento de los procesos cognitivos centrales- sino que depende del funcionamiento de los transductores ópticos... de la naturaleza cualitativa del *out-put* que éstos suministran a aquéllos. La conclusión es obvia: el contenido cualitativo de nuestros procesos perceptivos y sensoriales es un condicionante causal de su posible contenido intencional. Es por ello por lo que, a pesar de su dependencia lógica y epistémica respecto al contenido intencional objetivo, un análisis del significado de nuestros predicados psicológicos encargados de referir a aquellos procesos no puede eludir la consideración de la naturaleza cualitativa de nuestra experiencia.

Pero retomemos el hilo de nuestra discusión. Como conclusión valorativa de la estrategia defensiva del funcionalismo que hemos estado considerando creo que podríamos decir que la misma fracasa por varias razones. Aún suponiendo que fuera correcta en todos sus términos, lo más que mostraría es que los *qualia* no determinan el significado ni la aplicación de nuestros predicados psicológicos ordinarios, no que los *qualia* no existan y que no caractericen a algunos de nuestros estados mentales. Podríamos formular, pues, un reproche análogo al que dirigimos contra la posición de Fodor. Admitir que el análisis funcional no puede dar cuenta del carácter cualitativo de nuestras experiencias es admitir una limitación del funcionalismo en tanto que filosofía de la mente. Y por otra parte -y esto es señalar que la tesis fundamental en torno a la cual se articulan las posiciones que estamos considerando es errónea- no parece que el análisis de algunos de los predicados psicológicos que se refieren a procesos sensoriales y perceptivos pueda pasarse sin apelar a la naturaleza cualitativa de nuestra experiencia, pues la misma parece ser una condición causalmente relevante para la existencia de aquellos procesos y, por consiguiente, un factor relevante para determinar la corrección de la atribución de aquellos predicados que los refieren. Si dos

experiencias no fueran desagradables, el individuo que las padece no estaría legitimado a calificarlas como dolorosas, y si sus sensaciones cromáticas no fueran pertinentemente semejantes mentiría cuando afirmara que está viendo que dos objetos (o uno en diferentes ocasiones) tienen el mismo color. Todo lo cual no hace sino reforzar el segundo de nuestros reproches a Fodor, a saber: que es más que dudoso que una teoría psicológica de ciertos procesos mentales pueda pasarse sin la consideración de sus características fenomenológicas.

La situación parece ser, pues, ésta. La existencia de los *qualia* resulta difícilmente disputable, y es sensato pensar que los mismos juegan un papel importante en la determinación del significado de algunos de nuestros predicados psicológicos. Por consiguiente, todas y cada una de las objeciones que se plantearon contra el funcionalismo siguen en pie, y no queda sino intentar lidiar con ellas.

Así lo han comprendido tanto algunos de quienes militan en el campo del funcionalismo computacional (el mismo Dennett, en 1978e; Rey, 1989), cuanto algunos de los que se sitúan en los dominios del funcionalismo analítico (Lewis, 1978; y, sobre todo, Shoemaker, 1975a; 1975b; 1981b; 1981c). No podemos aquí hacer justicia a los detalles de sus posiciones, y las reflexiones críticas que a continuación vamos a exponer tendrían muy diferente alcance según contra quien las dirigiéramos.

Dada la concepción puramente relacional de los eventos, estados y procesos mentales que los funcionalistas profesan, la única estrategia coherente que les cabe es la de intentar mostrar que el contenido cualitativo puede reducirse a, y explicarse en términos de, el rol funcional que el mismo desempeña. Dicho con otras palabras: la naturaleza cualitativa de nuestra experiencia está en función de los otros estados, eventos o procesos mentales con los que su procesamiento característico la relaciona, así como de los estímulos que la provocan y la conducta a la que, de manera mediatizada, da lugar.

Antes de entrar en un análisis más pormenorizado quisiera hacer dos consideraciones que pueden predisponernos escépticamente respecto a la viabilidad de esta estrategia. Recordemos que una de las críticas que los funcionalistas dirigían a los materialistas era la de que su análisis conducía indefectiblemente a una posición *chauvinista*: si el dolor debe identificarse con la estimulación de las fibras C, de cualquier organismo que careciera de las mismas habría que negar que pudiera padecer dolor. Mi pregunta ahora es si el funcionalismo no incurre en un *chauvinismo* incluso más grave.

Es obvio que la economía funcional de la mente de los animales es muy diferente de la nuestra. Los out-puts de sus órganos sensoriales, por atenernos a la diferencia más palmaria, no pueden interactuar con ningún centro de control lingüístico. Esto significa que incluso a animales muy cercanos a nosotros en la escala evolutiva, no digamos ya a los que se hallan situados muy lejos de ella, tendríamos que adscribirles contenidos cualitativos completamente diferentes de los nuestros. Y posiblemente no sólo contenidos cualitativos, lo que aún podría resultar verosímil, sino también contenidos intencionales. Pues no menos que el cualitativo, el contenido intencional de los procesos perceptivos también quedaría determinado por la economía funcional global de la que forman parte. De manera que si adoptáramos una posición funcionalista estricta no sería legítimo decir, por ejemplo, que el gato salió corriendo porque vio al perro abalanzarse contra él. Pues ver perros sería potestad de aquellos que compartieran una estructura psíquica análoga a la nuestra.

Realmente la situación para el funcionalista es incluso peor que ésta, pues una conclusión lógica de su posición es que la divergencia en los contenidos cualitativos e intencionales se produce en el seno mismo del género humano. No sólo los animales son incapaces de expresarse lingüísticamente -o al menos con un lenguaje parecido al nuestro-. Los niños no nacen hablando. Por consiguiente, también habría que decir que el contenido cualitativo de sus actos perceptivos y sensoriales es diferente del nuestro. Y lo mismo habría que decir de aquéllos que conceptualicen la experiencia de una manera alternativa a la nuestra. Dado que ante unos mismos estímulos visuales el esquimal adquiere creencias y tiene conductas discriminatorias diferentes a las nuestras, habría que concluir que el carácter cualitativo de su experiencia visual es diferente al nuestro. Y puesto que -no sé si esto es cierto, pero basta con que sea verosímil- tampoco ellos hacen distinciones entre el verde y el rojo, quizás haya que decir que más bien es semejante a la del daltónico.

En resumen pues, que si estoy en lo correcto el criterio funcional de adscripción de contenido cualitativo e intencional para los procesos perceptivos y sensoriales nos llevará a la adopción de posiciones mucho más *chauvinistas* que aquéllas con las que nos comprometía el materialismo.

La segunda observación que quiero hacer es la siguiente. Según los principios funcionalistas, el contenido cualitativo de los procesos perceptivos y sensoriales queda determinado por las relaciones (computacionales) que mantiene con otros eventos, estados y procesos mentales, así como por la conducta que (mediatizadamente) puedan causar. Hay aquí, a mi entender, una especie de reductivismo análogo a aquél del que pecaban las teorías más rudimentarias que pretendían explicar la percepción en términos de adquisición de creencias (Armstrong, 1961; Pitcher, 1971). Sin entrar en la cuestión de la existencia de un percibir no epistémico (Cf. Dretske, 1969), es duro llegar a aceptar que sólo vemos aquello de lo que nos formamos una creencia merced al uso de nuestros sentidos, y esta misma intuición subsiste aún cuando compliquemos cuanto queramos el número y la naturaleza del resto de procesos psíquicos con los que interrelacionar nuestra experiencia. Si la riqueza de ésta no puede reducirse a las creencias que en ella se originan, tampoco pueden reducirse a esas creencias los recuerdos a corto o largo plazo, los *out-puts* hacia la unidad de control lingüístico, etc. Sencillamente no procesamos todo lo que vemos.

Vamos a pasar por alto estas observaciones y a suponer que el teórico funcionalista tiene réplicas convincentes a las mismas. ¿Cómo se resolverían entonces los problemas relacionados con los *qualia*?

El criterio funcional de determinación de contenido sugiere, para empezar, una respuesta obvia al problema de la total ausencia de carácter cualitativo de la experiencia (Cf. Shoemaker, 1975b; Churchland, 1981). Recordemos que éste se planteaba por la posibilidad de que un sistema que fuera funcionalmente isomórfico con nosotros no fuera, sin embargo, consciente de *qualia* alguno. Pero si presuponemos la validez de aquel criterio tal posibilidad se desvanece de manera inmediata. Pues para que un sistema fuera funcionalmente isomórfico con nosotros sería necesario no sólo que se comportara análogamente a como nosotros lo hacemos, sino que también el resto de sus operaciones psíquicas resultaran equivalentes a las nuestras; entre ellas la introspección. Para que se diera el isomorfismo funcional estricto no bastaría, pues, que ante estímulos semejantes el otro sistema fuera

capaz de hacer discriminaciones objetivas análogas a las nuestras, sino que también debiera ser capaz de dar informes introspectivos sobre la naturaleza cualitativa de su experiencia como nosotros somos capaces de hacerlo, pues esta operación forma parte esencial del conjunto de eventos, procesos y estados psíquicos que dotan de contenido cualitativo a nuestras experiencias. Y si fuera así, ¿qué más necesitaríamos para conceder la conciencia de *qualia* a ese sistema?.

Es muy importante darse cuenta de que, como Shoemaker subrayó (1975b), este argumento no es de corte verificacionista, y que más bien lo único que asume es una teoría causal del conocimiento y de la referencia. El argumento no afirma que no tiene sentido dudar de la existencia de un *quale* cuando éste da lugar a determinadas creencias sobre su carácter cualitativo o/y a determinada conducta, porque la evidencia introspectiva y la conductual son los únicos métodos de verificación de los que disponemos para aseverar la existencia de los *qualia*. Más bien el argumento es de este tipo: si la conciencia introspectiva sobre el carácter cualitativo de nuestros estados mentales merece algún crédito epistemológico, o si los términos con que expresamos las creencias a las que tal conciencia da lugar se refieren a algo, entonces el carácter cualitativo de nuestros procesos perceptivos y sensoriales debe ser causalmente responsable de aquella conciencia. Si un sistema es funcionalmente isomórfico con nosotros, entonces entre sus estados, procesos y eventos psicológicos deberá darse una relación causal semejante a la que se da entre los nuestros; en concreto, sus creencias introspectivas también deberán, como en nuestro propio caso, estar causadas por el carácter cualitativo que la experiencia tenga. Quiero que se retenga este punto porque -ya volveré sobre ello- la asunción de esta teoría causal del conocimiento y de la referencia y, en general, de la mente, creo que nos permite construir un argumento que muestra la inviabilidad de mantener una posición estrictamente funcionalista en torno al problema de los *qualia*.

Está claro que esta solución al problema de los *qualia* ausentes sugiere respuestas obvias a los problemas de taxonomía y discriminación. Recordemos que el primero consistía en dar una caracterización funcional de la distinción entre los estados mentales con contenido fenomenológico y los estados mentales carentes del mismo, mientras que el segundo nos pedía una explicación de cómo es posible la discriminación de ciertos estados mentales aún antes de que seamos conscientes de ninguna de sus peculiaridades funcionales.

A esto último podríamos responder que una cosa tal es radicalmente imposible. Si somos capaces de discriminar entre dos estados mentales, como mínimo debemos ser capaces de tener creencias introspectivas diferentes acerca de la naturaleza cualitativa de los mismos, pero dado que las creencias introspectivas a la que los estados mentales puedan dar lugar ya son características funcionales de los mismos, de ahí se sigue que no hay discriminación posible de estados mentales que no sea ya una discriminación en virtud de sus características funcionales. Podemos discriminar entre estados mentales siendo ignorantes de algunas de sus características funcionales (por ejemplo, siendo ignorantes de sus relaciones con *inputs* estimuladores o *outputs* conductuales), pero no siendo ignorantes de todas sus características.

Por lo que hace al problema taxonómico estará ahora claro que la clasificación de los estados mentales por su posesión o carencia de características cualitativas puede ser construida funcionalmente, como una clasificación de los estados por su capacidad para dar lugar a estados introspectivos de cierto tipo.

Es más, esta solución funcionalista nos permite explicar en un cierto estilo Wittgensteiniano algunas de las peculiaridades epistemológicas de nuestra experiencia: la aparente autoridad de la primera persona sobre la misma, por ejemplo. Si nuestro testimonio acerca de las propiedades de nuestras experiencias perceptivas o sensoriales es fiable ello obedece al simple hecho de que la explicitación de su carácter cualitativo forma parte esencial de los procesos que determinan su naturaleza. Como el ¡Ay! con respecto al dolor, habría que entender la relación entre los informes en primera persona y las características de la experiencia perceptiva o sensorial más como causal y expresiva que como descriptiva.

En cualquier caso, es obvio que el tema es bastante diferente con respecto al argumento del espectro invertido. Supongamos que yo tengo una creencia introspectiva acerca del color del que tengo experiencia ahora, mientras miro el mar, y que considero que la manera correcta de manifestar mi creencia es afirmando que soy consciente de algo azul. Mi vecino puede en este momento tener una creencia introspectiva que expresa con los mismos términos. Admitamos que si ambos tenemos conciencia introspectiva del carácter cualitativo de un estado mental, entonces este estado debe tener un tal carácter, pero ¿qué garantiza que un mismo informe introspectivo haya de ser causado en todas las ocasiones por estados cualitativamente idénticos?. Los defensores del argumento del espectro invertido insisten en que es una posibilidad lógica que si yo fuera a tener conciencia del estado de mi vecino, entonces estimara que la descripción correcta del carácter cualitativo del mismo fuera otra que la que estimamos lo es cuando somos conscientes del nuestro. Y si esto es una posibilidad lógica entonces lo que se demuestra es una limitación del funcionalismo. Si bien podríamos dar una caracterización funcional de la existencia de los *qualia*, paradójicamente no podríamos definir funcionalmente sus relaciones de semejanza y desemejanza.

Podríamos, como estrategia defensiva, insistir en que la posibilidad del espectro invertido es una mera posibilidad lógica empíricamente muy difícil de concretarse (Cf. Harrison, 1973; 1984), pero no creo que la misma nos llevara muy lejos dado que sabemos de la existencia de ciertos casos que vienen a apuntar a la misma conclusión a la que el argumento de la inversión espectral apunta, tales como el de algunos de los pacientes afectados de dicromatismo.

En efecto, la merma en su capacidad discriminatoria de los colores, en estas personas en las que algún tipo de conos retinianos ha sufrido un proceso de deterioro, puede corregirse mediante la aplicación a las mismas de lentes de contacto coloreadas que actúan como filtros de luz, de manera que los enfermos dicromáticos pueden volver a distinguir las cosas rojas de las verdes... no obstante lo cual en sus informes introspectivos aseguran que su experiencia de estos colores es cualitativamente diferente de lo que lo era antes. Este fenómeno bien contrastado (Cf. Huvich, 1981) viene a indicar lo mismo que el argumento del hipotético espectro invertido: cierta autonomía del contenido cualitativo respecto al contenido intencional de nuestros procesos perceptivos y sensoriales. Pero si ello es así ya no podemos pensar que los criterios de identidad del contenido cualitativo de nuestros procesos perceptivos y sensoriales puedan reducirse a criterios puramente funcionales.

Para que esta reducción fuera factible debiéramos suponer que la identidad de roles funcionales de dos experiencias determinaría la identidad cualitativa de las mismas. Y precisamente esto es lo que pone en tela de juicio el caso de los enfermos dicromáticos.

¿Acaso no se sirven de experiencias cualitativamente diferentes de las de los receptores normales (incluidas las de ellos mismos cuando lo eran) para hacer los mismos juicios que éstos sobre los colores de las cosas?

Sin embargo, el funcionalista estricto aún puede protestar a estas alturas, y argumentar que en el caso de los dicromatos o, para el caso es lo mismo, en el de los hipotéticos afectados por la inversión del espectro, la verosimilitud de atribuirles una diferencia en el carácter cualitativo de sus experiencias es explicable en términos funcionales. Al fin y al cabo los dicromatos hacen juicios introspectivos («mi experiencia del color es ahora diferente de la experiencia que recuerdo tenía de los mismos») que los receptores normales no hacemos, y es de suponer que las diferencias no sólo en los informes introspectivos sino también en el plano conductual serían mucho más severas en el caso de los afectados por la inversión espectral. Luego las diferencias en el contenido cualitativo de la experiencia se corresponden con diferencias en el rol funcional que desempeñan.

Esta réplica no deja de ser verdadera, pero podemos complicar nuestros ejemplos hasta hacer perder al funcionalista toda esperanza de poder mantener dignamente su posición. Supongamos que a uno de los pacientes afectados de dicromatismo, para instalarle las lentes coloreadas que le van a permitir recobrar sus capacidades discriminativas, lo anestesiemos totalmente, momento que aprovechamos para inyectarle una sustancia que borre su memoria cualitativa (si esto suena muy inverosímil conviene tener en cuenta que los afectados de acromatopsia cerebral adquirida no sólo pierden su capacidad para percibir el color de las cosas, sino también para imaginárselo, o para soñarlo. Cf. Wasserman & Sacks, 1988). Cuando nuestro pobre dicromato despierte su experiencia cualitativa será diferente a la que tenía antes de que sus conos se deteriorasen... pero lo ignorará. No sólo su conducta discriminatoria sino sus propios informes introspectivos serán de todo punto indistinguibles de los nuestros.

Si el funcionalista se empeña en mantener su posición, lo único que le queda decir es que dado que la experiencia juega el mismo rol funcional antes del advenimiento de la enfermedad y después del sometimiento a la operación, en uno y otro momento tiene el mismo carácter cualitativo. Mientras la experiencia del dicromato operado con nuestra técnica no sufriría alteración alguna respecto a la experiencia que tenía cuando estaba sano, no podría decirse lo mismo respecto a la experiencia del dicromato al que se le ha instalado las lentes de contacto por el método habitual. Pero ¿cómo una mera diferencia en la memoria puede explicar una diferencia en la naturaleza de la experiencia? Si el dicromato tiene buena memoria, el contenido cualitativo de sus procesos perceptivos y sensoriales será diferente; si la tiene mala será idéntica a antes de caer enfermo.

Parece que el funcionalista estricto está aquí poniendo la carreta delante de los bueyes. Pues es conceptualmente lícito defender que la naturaleza de los eventos determina, si éste es fiel, el recuerdo que tengamos de ellos, pero no veo cómo se pueda defender que el falso recuerdo o, aún peor, la falta de recuerdo, determina la naturaleza de los eventos. ¿Se atrevería el funcionalista estricto a defender algo semejante para el mundo físico? ¿Se atrevería a defender, por ejemplo, que dado que no recordamos que el coche del vecino era de color rojo hace cinco días, éste no puede haberlo mandado pintar con el color azul que luce ahora? Desde luego que si el vecino fuera funcionalista y malmemoriado se podría

ahorrar la factura del planchista.... Y si no se atreve a defender esto -¡esperemos que no se atreva!- ¿será acaso que mientras es sensatamente realista respecto al mundo físico es fenomenista respecto al mental? ¿Defenderá acaso que en el mundo mental el *esse est percipi*, de manera que, como defendió Hume, las cosas son como nos parecen, y nos parecen tal como son?

Mucho me temo que la posición dramática que el funcionalista se ve obligado a tomar no sólo es intuitivamente indefendible -el sentido común nos dice que los dicromatos, independientemente de cómo anden de memoria, tienen una experiencia cualitativamente diferente de la de los perceptores normales- sino también conceptualmente absurda.

Volvamos al lecho rocoso de las intuiciones. El sentido común nos empuja a decir que los dicromatos con lentes coloreadas tienen una experiencia cualitativamente diferente a la de los perceptores normales. ¿Por qué?. Bueno, para empezar tenemos su testimonio introspectivo, una diferencia funcionalmente caracterizable. Pero esto no explica el caso del dicromato cualitativamente amnésico. ¿Por qué decimos de éste que su experiencia es semejante a la de sus compañeros de enfermedad y diferente de la de los perceptores sanos? La respuesta obvia es: porque el estado de sus órganos sensoriales es semejante al de sus compañeros de patología y diferente al de los perceptores normales.

Supongo que adonde quiero ir a parar ya lo verán claro. Lo que quiero indicar es que, sin negar que los criterios funcionales y conductistas (que no son sino un subgrupo de los funcionales) forman parte de los criterios que utilizamos para determinar las relaciones de semejanza o diferencia de los contenidos cualitativos de los procesos perceptivos y sensoriales, los criterios físicos no son menos ineludibles. Es el presupuesto de una naturaleza biológica común, y no sólo de una conducta o una economía psico-funcional común, el que nos justifica a adscribir contenidos cualitativos semejantes a los procesos perceptivos y sensoriales de todos nuestros congéneres.

Podríamos extendernos ahora sobre cómo este reconocimiento de la dimensión materialista que tienen los criterios ordinarios de adscripción de predicados psicológicos dotados de contenido cualitativo o fenomenológico permite soslayar, o por lo menos debilitar, el peligro de *chauvinismo* que denunciábamos acechaba al funcionalismo estricto; cómo, también, permite dar una respuesta que suene menos tautológica que la dada por el funcionalista al problema taxonómico, discriminatorio y del *quale* ausente; o cómo, en contra de las apariencias, no tiene por qué suponer una grave amenaza contra la autoridad de la primera persona; pero preferimos utilizar el poco tiempo que nos queda en construir otro argumento que conduce a la misma conclusión: la necesidad de completar los criterios funcionales con criterios materialistas.

Volvamos al argumento funcionalista contra la objeción de la ausencia de carácter cualitativo de nuestras experiencias. Recordemos que el mismo pendía de la asunción de una teoría causal del conocimiento y de la referencia. Si se tienen creencias acerca del carácter cualitativo de los estados psíquicos, y aquellas creencias tienen alguna fiabilidad epistémica, entonces las mismas deben ser causadas por estados dotados de ciertas características cualitativas. Ahora bien, un problema común tanto a las teorías causales del conocimiento, como a las teorías causales de la referencia, de la acción, etc.; o, en general, un desafío usual a cualquier teoría causal de la mente, lo constituye la posibilidad de la

existencia de cadenas causalmente desviadas. Esta posibilidad es la que obliga a matizar que una creencia sólo es conocimiento, o una acción sólo es intencional, o una expresión sólo refiere, etc., si está causalmente conectada de la manera habitual con la evidencia que la sustenta, o con los deseos y creencias que la provocan, o con la entidad que designa, etc. (véase a este respecto, Goldman, 1967; Martin & Deutscher, 1966 y Grice, 1961 y, en general, los trabajos recopilados en Davis, 1983). Pues bien, este mismo problema afecta a la solución que el funcionalista da al argumento del *quale* ausente.

¿Qué ocurriría en el caso de que alguien tuviera acceso al out-put de alguno de los sistemas modulares de entrada de otra persona?. Supongamos, por ejemplo, que la ceguera al color que alguien padece es total y no se debe a una deficiencia de sus conos retinianos sino a alguna lesión de su cortex visual. Puestos a suponer, sigamos suponiendo que los neurofisiólogos han descubierto alguna manera de transmitir la información que el ojo de este paciente suministra al nervio óptico de un perceptor normal donde la misma es procesada de la manera habitual por el módulo sensorial visual, y que el output de éste vuelve a ser transmitido al cerebro del perceptor lesionado, en el que puede interactuar con los sistemas cognitivos centrales. Este paciente podría no sólo hacer discriminaciones acerca de las propiedades empíricas objetivas que ve, sino también informes introspectivos acerca del carácter cualitativo de una experiencia que es causa de aquellas discriminaciones y de estos informes... ¡pero de una experiencia que él no tiene!

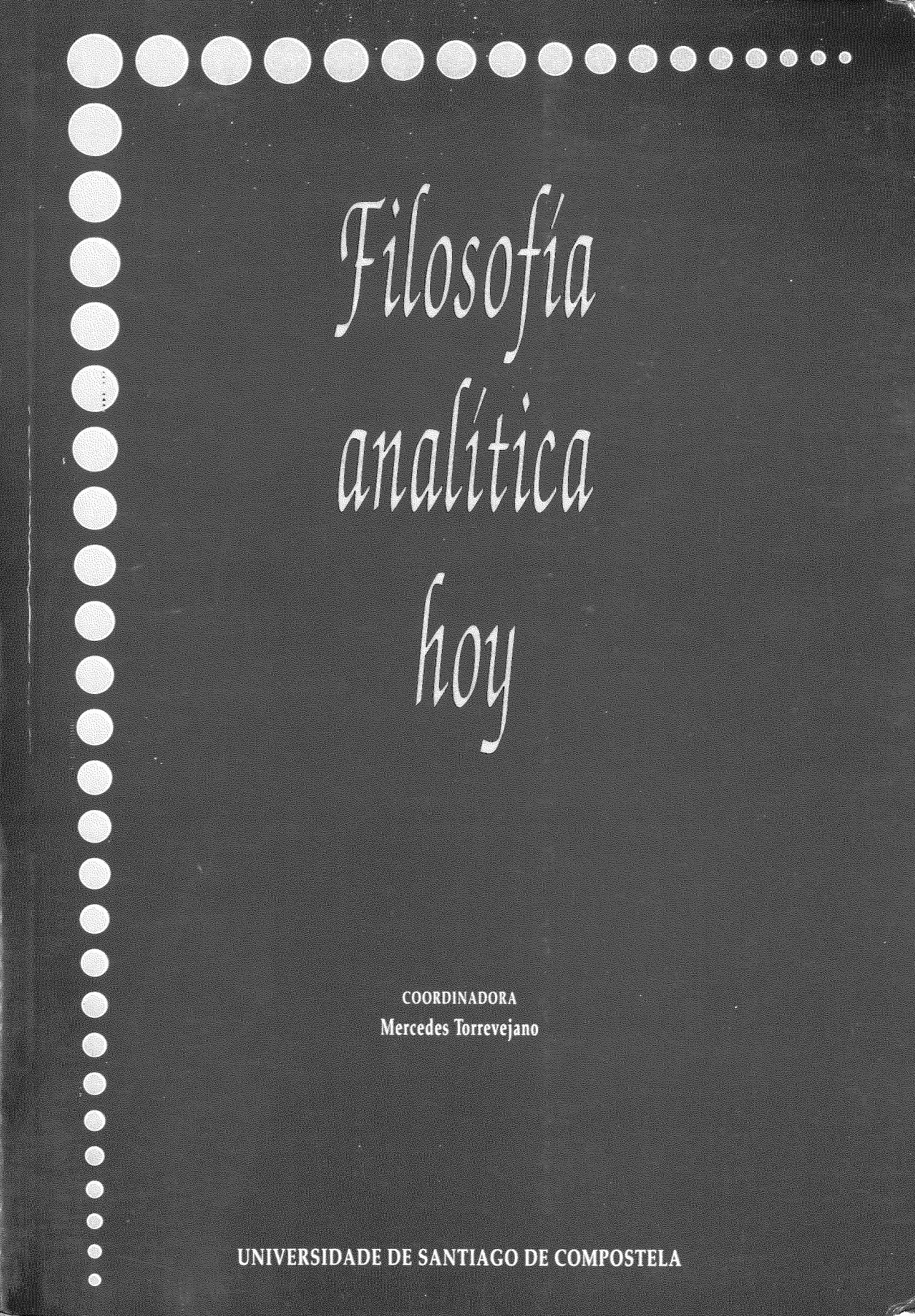
Lo que este experimento mental muestra, a mi entender, es que no basta para asegurar la posesión de estados mentales cualitativamente caracterizados el que tengamos creencias acerca de sus características cualitativas causadas por los mismos. Además debemos suponer que tal causación se produce en la manera habitual; lo que no significa sino que los procesos, eventos y estados pertinentes tienen una instanciación física normal. La explicación funcional sólo funciona, una vez más, si asumimos la normalidad física o biológica del sujeto al que atribuimos predicados psicológicos.

Si estoy en lo correcto, pues, no sólo para resolver el problema de la semejanza cualitativa de los procesos perceptivos y sensoriales, sino también para resolver el problema mismo de los *qualia* ausentes debemos completar las definiciones funcionales con intuiciones materialistas.

Por otra parte, creo que este tipo de problemas que se plantean al funcionalismo está relacionado con una característica muy básica del mismo que bien puede ser completamente errónea: su enfoque atomista del sujeto psicológico. El funcionalismo pretende explicar a este sujeto como el producto del ensamblaje de distintas estructuras funcionales, dando prioridad lógica a las partes sobre la unidad que conforman. Es este presupuesto, a la vez metodológico y ontológico, el que lo hace vulnerable a las paradojas de las realizaciones extravagantes (las habitaciones chinas de Searle, 1980, 1984; las naciones pensantes de Block, 1978; el robot con cabezas de hidra de Putnam, 1981, etc.), paradojas que pueden resolverse si, más de acuerdo con nuestros criterios preteóricos de adscripción de predicados psicológicos, consideramos que el sujeto propio de atribución de tales predicados son los organismos, los seres vivos a los que resulta consustancial su unidad, y no los sistemas cuya unidad es el resultado de un ensamblaje accidental de estructuras funcionales.

Podemos ahora concluir nuestra exposición. Las posibilidades de que el funcionalismo pueda suministrar un tratamiento adecuado de los procesos psicológicos dotados de características cualitativas son más bien escasas. No creo que una estrategia estrictamente funcionalista pueda dar cuenta de los problemas relacionados con los *qualia*. Para enfrentarse a los mismos será necesario atender también a los criterios materialistas en los que se apoya nuestra adscripción de los predicados que denotan aquellos procesos. Por otra parte, estoy convencido de que las cosas no irán mucho mejor por lo que respecta a las actitudes proposicionales. Sólo que aquí para superar las limitaciones del funcionalismo nos vemos obligados a la consideración del sujeto psicológico no sólo como un organismo sino como un organismo que interacciona con el mundo natural y, en el caso de las personas, con el social. Nuestra conclusión se acerca, pues, a la alcanzada por Putnan, 1973. Si las consideraciones biológicas son relevantes para entender la naturaleza de ciertos procesos psíquicos, las sociológicas pueden serlo no menos esenciales para la del resto.

Vicente Sanfélix Vidarte
Septiembre, 1989



*Filosofía
analítica
hoy*

COORDINADORA
Mercedes Torrejano

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FILOSOFIA analítica hoy : Encuentro de tradiciones / Coordinadora: Mercedes Torre-vejano. — Santiago de Compostela : Universidade. Servicio de Publicacións e Intercambio Científico, 1991. — 284 p. ; 24 cm. — (Cursos e Congresos da Universidade de Santiago de Compostela ; 67). — Índice: p. 5. — D.L.: C. 582-1991. — ISBN 84-7191-722-X.

I. Torrevejano, Mercedes, coord.; II. Universidade de Santiago de Compostela. Servicio de Publicacións e Intercambio Científico, ed.

140

© Universidade de
Santiago de Compostela, 1991

EDITA: Servicio de Publicacións
e Intercambio Científico
Casa Gradín
Campus Universitario

IMPRIME: Imprenta Universitaria
Pavillón de Servizos
Campus Universitario

Dep. Leg.: 582-1991
ISBN: 84-7191-722-X

INDICE

Presentación	7
Después del análisis: significado, comprensión e intencionalidad	9
<i>Juan José Acero</i>	
La filosofía analítica y su concepción del método de la filosofía	27
<i>Ernesto Tugendhat</i>	
Sobre el pretendido realismo básico del Tractatus	35
<i>Alfonso G. Suárez</i>	
El juego de saber y el lenguaje de la certeza	49
<i>Angel Alvarez Gómez</i>	
La diferencia entre explicar e interpretar según Davidson	83
<i>Manuel García-Carpintero</i>	
Panorama actual de la filosofía analítica de la mente: funcionalismo y experiencia	121
<i>Vicente Sanfélix Vidarte</i>	
Escritura y "phone": el desplazamiento derridiano de la tradición analítica	155
<i>Jorge Pérez de Tudela Velasco</i>	
Putnam, o las perplejidades de la filosofía	177
<i>Mercedes Torrevejano</i>	
El "realismo" en la tradición analítica. El caso de Hilary Putnam	193
<i>J.M. Sagüillo</i>	
Antología de texto de Hilary Putnam	217
<i>Luis Villegas</i>	