Universidad de Chile
Facultad de Filosofía y Humanidades
Departamento de Filosofía

Estructura y defensa del materialismo eliminativo

Informe de seminario de grado: "*La mente en un mundo físico*" [para optar al grado de Licenciado en Filosofía]

Autor:
Simón Palacios B.
Profesor a cargo: Guido Vallejos O

2008

Resumen . . 4

Introducción . . 5

Capítulo 1. El argumento y los supuestos del materialismo eliminativo . . 10

Capitulo 2. Las objeciones y los supuestos del materialismo eliminativo . . 12

Capítulo 3. El argumento trascendental contra el materialismo eliminativo y la reaparición de un viejo problema . . 26

Capítulo 4. Eliminando a la competencia . . 36

Conclusión. Y bosquejo de discusión epistemológica . . 42

Bibliogrtafía . . 44

Resumen

Este trabajo se divide en dos grandes bloques: 1) Se analiza el argumento del materialismo eliminativo y se establecen los supuestos fundamentales necesarios al mismo. Además se organizan algunas objeciones al eliminativismo atendiendo a qué supuestos enfrentan (o deberían enfrentar) directamente. Finalmente se trata una clase especial de objeciones contra el materialismo eliminativo. 2) Se critican tres programas en filosofía de la mente (conductismo filosófico, teoría de la identidad y funcionalismo) desde el materialismo eliminativo y suponiendo como criterio general de crítica –propiamente eliminativista- la consistencia de estos programas con el dualismo. Finalmente se bosqueja un problema epistemológico general implicado en varios de los temas tratados con anterioridad.

Introducción

Al estar finalizando mis estudios de filosofía he debido explicar a mucha gente –muchos no entendidos, otros pipiolos y otros más versados en filosofía en general- de qué trata el seminario de grado que elegí y qué es esto de "la mente en un mundo físico". Al hablarles, en primer lugar y especialmente a los menos versados en filosofía, de un supuesto problema "mente-cuerpo" se genera inmediatamente algo de sorpresa en muchos de mis interlocutores quienes sospechaban que el seminario consistiría en algo bien distinto, presumiblemente en agotadoras lecturas acerca de enredadas teorías del alma, el espíritu, los caracteres psicológicos, o cosas por el estilo. Al comentarles que muchos de los problemas estudiados son en verdad cercanos a las ciencias empíricas y al materialismo las reacciones se diversifican: unos tienden a decepcionarse por no resultar congruentes mis estudios con ciertos problemas más afines con su visión de mundo y que hubiesen gustado tratar conmigo (problemas referentes al espíritu o la inmortalidad del alma), otros se sienten muy animados y confiesan que ni siquiera sospechaban que la filosofía podía alcanzar un tratamiento de algún problema con cierto grado de precisión conceptual, sentido común y perspectivas de resolución [1].

Sea como sea llegado el momento en la conversación de comentarles qué programa específicamente había elegido estudiar dentro de las múltiples opciones en filosofía de la mente, y al agregar que posiblemente desarrollaría mi tesis al respecto, la sorpresa fue siempre común. Al expresar las tesis fundamentales del materialismo eliminativo, casi sin excepción, se me acusó de estar bromeando, se me dijo que nadie podía creer sinceramente en algo tan disparatado, o bien dejó mi interlocutor de creer que la filosofía podía tratar asunto alguno sin llevarlo al total descalabro (suponiendo que haya creído esto alguna vez).

La respuesta frente al eliminativismo fue siempre homogénea: le pareció a todo interlocutor no versado en estos temas, una doctrina descabellada, sin la menor perspectiva de poder ser siquiera algo coherente y sin una gota de sentido común. Cuento corto: he podido constatar en un estudio sociológico muy laxo que el materialismo eliminativo es visto por el sentido común como una locura sin par, una doctrina para locos o sutilísimos sofistas o una broma filosófica (y no de muy buen gusto).

Esto no es de sorprender, no es usual que alguien ande por allí aseverando que nadie nunca ha sentido dolor alguno, o que nunca nadie ha creído ni deseado nada… que todas estas cosas son compromisos teóricos suscritos al aceptar la psicología popular (cosa a la que ninguno de mis interlocutores recordaba haber suscrito nunca)… ninguna de estas expresiones resumidas y toscas del eliminativismo favorecían su aceptación popular.

Emplee, por lo tanto, varias estratagemas para suavizar el impacto de mis interlocutores con esta doctrina aberrante al sentido común; la más eficaz resultó ser la

[1] Quizás ambas clases de interlocutores exageren en su respectiva decepción y emoción. La filosofía de la mente actual, por una parte, sí parece dar cierto espacio a discusiones acerca de problemas tradicionalmente clasificados dentro de las "ciencias del espíritu" como son los caracteres mentales o la conciencia, problemas que quizás estimularían a los decepcionados, y por otra parte la filosofía de la mente no parece poseer toda la claridad, todo el sentido común y todas las perspectivas de resolución que excitaban al otro bando.

que Searle (1987) denomina “la apelación a la época heroica de la ciencia”, esto es, referir múltiples ejemplos históricos acerca de tesis que en ciertas épocas parecieron totalmente contraintuitivas y afrentosas para el sentido común, y que, sin embargo, forman hoy parte del catálogo de nuestras verdades científicas casi triviales. Otra estrategia similar, aunque menos exitosa psicológicamente, consiste en mostrar los gruesos absurdos teóricos que fueron tomados por verdades evidentísimas en diversos momentos de nuestra historia y compararlos (al menos potencialmente) con nuestra psicología popular para concluir que la fuerte complacencia de esta última al sentido común nada prueba.

Quisiera en esta introducción no repetir estas estratagemas para presentar el materialismo eliminativo al lector (sería quizás tratarlo de inexperto de antemano), por lo que comenzaré por exponer un hito relevante de esta doctrina y que sin duda hace juego con la extravagancia que frente al sentido común posee el eliminativismo en general. Tal hito es el experimento mental de los Antípodas desarrollado por Rorty (1979). Dejemos entonces que una locura explique a otra:

Suponga el lector que en algún sitio lejano (quizás fuera de nuestro planeta) existe una raza de seres muy semejantes a nosotros. Estos seres tienen una constitución física, social y moral idéntica o extremadamente semejante a la humana. Pero suponga también el lector que estos seres no poseen ningún concepto de “mente”. Una neurociencia increíblemente avanzada (aunque no completa) les permite referirse directamente a sus estados neurales, y de este modo pueden emplear correctamente casi todo nuestro lenguaje mentalista –o un equivalente neurocientífico- para referirse a sus propias creencias y otras actitudes proposicionales. Dirán por ejemplo: “se ha estimulado cierta parte específica de mi fibra C” en vez de “me he quemado el dedo”, o dirán cosas como: “Se ha estimulado C-345… ya sabes, el estado neural que me obliga a afirmar que se ha estimulado mi fibra C” en vez de “creo que me he quemado el dedo”, y otras cosas por el estilo.

Imagine el lector ahora que nosotros los humanos hemos conseguido contactarnos con estos seres fantásticos y que nuestros filósofos analíticos, haciendo chiste de la supuesta “a-mentalidad” que estos individuos parecen exhibir, han decidido bautizarlos “Antípodas” -en honor a las doctrinas monistas originarias de Australia y Nueva Zelanda-. Ya pasada la broma inicial, la falta de un lenguaje mentalista o de toda referencia a estados mentales comienza, sin duda, a preocupar a nuestros humanos filósofos. Después de insistentes interrogatorios, nuestros filósofos no consiguen que ningún Antípoda dé muestras de referirse a estados mentales en el sentido usual en el que nosotros emplearíamos estos términos. El Antípoda asegura estar refiriéndose siempre a ciertos estado de su sistema nervioso y, estando la neurociencia Antípoda muy avanzada y siendo capaces ellos mismos de manejar un lenguaje neurocientífico muy detallado y exacto, nada parece escapárseles de cuanto el humano interroga.

La discusión acerca de si los Antípodas poseen o no estados mentales (digamos, sensaciones primarias) se enciende entre los humanos. Finalmente se llega a un consenso que hace justicia a todas nuestras intuiciones acerca de lo mental: si los Antípodas poseen estados incorregibles, es seguro que estos estados son generados por sensaciones primarias y que ellos tienen, en consecuencia, mente en el sentido en que nosotros la tenemos –aun cuando pudieran ignorar todo esto por alguna extraña condición-. Pero si los Antípodas carecieran de estados incorregibles a todas luces no tendrían sensaciones primarias ni mente.

Pero al consultar a los Antípodas si ellos tenían estados incorregibles la dificultad se acrecentaba. Ellos no parecían entender con claridad el sentido de la expresión “estado incorregible”, por una parte la estimulación de las fibras-C los obligaba a entrar en el estado

C-345 (algo así como “creo que se han estimulado mis fibras-C” traducido al humano) y esto era algo, sin duda, inevitable por parte de cualquier Antípoda normal. Pero por otra parte esta clase de estados –estados de “parecer” pasaremos a llamarlos- podrían eventualmente no estar ligados causalmente a la estimulación de las fibras-C, aunque muy regularmente lo estuviesen. Y esto porque existían extraños estados concomitantes (digamos la serie de estados T) cuya activación disparaba inevitablemente estados neurales tales como C-345. Un Antípoda estaba convencido de que sus fibras-C se habían activado al sentir C-345, aun cuando C-345 fuera ocasionalmente activado por T-123 y no por la estimulación de las fibras-C. Ciertamente tenían entonces estados incorregibles, aun cuando un detallado análisis de la actividad neuronal de una Antípoda podría ponerlo, más adelante, en cuenta de que su estado C-345 no estaba siendo generado por la estimulación de las fibras-C sino por T-123.

Todo esto confundió bastante a los filósofos humanos, pues existía aun una gran diferencia entre los dolores u otras sensaciones primarias y la estimulación de las fibras-C u otros estados neuronales. Por una parte se podría (antípodamente) tener la ilusión de que se habían estimulado las fibras-C (digamos, T-123) aunque ningún Antípoda podría ponerlo en duda *in situ* –por lo que serían estados incorregibles-, pero por otra parte se suponía que no se podría tener (humanamente) la ilusión de un dolor sin tener ese mismo dolor. Ante esta disyuntiva los filósofos humanos más cercanos al materialismo comenzaron a sospechar algo desagradable para todos… quizás el caso de los Antípodas y de su “a-mentalidad” no era tan graciosos como ellos creían en un principio… quizás la broma estaba del lado de los humanos y de sus supuestos estados mentales. Si los estados incorregibles de los Antípodas cumplían las mismas funciones que las sensaciones primarias humanas, ¿no estaría abierta la posibilidad de que al hablar estos últimos de “dolor” y otras cosas semejantes estuvieran, en realidad, solamente informando acerca de sus estados neurales?

Esta inquietante cuestión volvíase cada vez más dramática al constatar que los Antípodas no se distinguían funcionalmente de los seres humanos en ningún otro punto: poseían un discurso intencional, eran autoconscientes, tenían moralidad y todos los rasgos de una cultura humanamente desarrollada (incluyendo, supongamos, una religión y una doctrina de la inmortalidad –aunque fisicalista-), además de las increíbles semejanzas biológicas… supuesto todo esto, el propio Antípoda podría preguntar molesto: “¿Por qué ustedes los humanos se creen tan distintos en base a esas entidades misteriosas que no pueden, por cierto, explicar ni probar fuera de una experiencia supuestamente privilegiada?, y si somos idénticos en todos los demás rasgos ¿por qué creen que tales cosas (los fenómenos mentales) son tan relevantes, siquiera importantes, para caracterizar “personas”?, y ya que han pretendido enseñarnos a nosotros –cuando nos consideraban personas, pues supongo que ahora dudan de esto también- que *debíamos* de tener cosas tales como sensaciones primarias, y puesto que no han podido probarnos esto, siquiera expresarlo como un problema coherente en nuestro lenguaje, ¿no deberían ustedes también aceptar que quizás nosotros los Antípodas tenemos algo que enseñarles, que quizás nosotros tenemos la razón y ustedes están confundidos por su lenguaje y sus malas teorías, que no hay tales cosas como sensaciones primarias, es decir, que todos somos *personas sin mente*?”

Aunque el filósofo humano ya sospechaba todas estas implicancias de su fracasado intento por “mentalizar” a los Antípodas, aun quiso establecer una corrección importante: “los estados incorregibles -afirmó- fueron descritos de modo demasiado laxo y ambiguo, lo cierto es que *siempre que hagamos un informe incorregible sobre un estado de nosotros*

mismos, tiene que haber una propiedad que se nos ofrezca y que nos induzca a hacer el informe[2]. Si los Antípodas insisten en que sus informes incorregibles no están fundados en una propiedad sustantiva que se les ofrece, y siguen describiendo estos estados como *pareceres*, es decir, como ciertos estados en los que el cerebro Antípoda entra inevitablemente ante ciertos hechos (probablemente otros hechos neurales), entonces ellos estarán insistiendo simple y llanamente en que no tienen mente. Esto porque toda la intuición humana acerca de la atribución de mentalidad parece sustentarse en P… es por el hecho de poseer sensaciones cualitativas que suponemos que los perros tienen alguna forma de mentalidad, y es por la ausencia de ellas que suponemos que un robot (aun el más sofisticado de los autómatas) no la tiene, y punto"

La discusión llega entonces a un punto muerto y disocia en varios bandos a humanos y Antípodas: por una parte algunos humanos creen que los Antípodas deben de tener sensaciones primarias que fundamenten sus estados incorregibles –sean estos estados de la clase que sean - por los rasgos manifiestos de su conducta (intencionalidad, por ejemplo), y esto aun cuando los propios Antípodas ignoren estas sensaciones; otros humanos creer que la discusión no puede resolverse en lo absoluto pues ninguna evidencia empírica beneficiaría más a uno que a otro bando, por lo que presumiblemente se sienten cómodos siendo escépticos respecto de la mentalidad de los Antípodas y sospechen frecuentemente en su fuero interno que los Antípodas no son más que complejísimos autómatas (una broma filosófica enviada por Dios quizás); y un tercer grupo de humanos y fundamentalmente la gran mayoría de los Antípodas llegará a pensar que la idea de que existen ciertas entidades cognoscibles de modo incorregible es una distorsión del lenguaje respecto de los *pareceres* causada por una neurociencia poco desarrollada y un intento desesperado por suplir los vacíos teóricos mediante extrañas categorías del lenguaje.

Sea como sea no hay buenas perspectivas de acuerdo aun habiendo hecho explícita la premisa P. Pero hay algo revelador al atender a la inutilidad de esta premisa para resolver el problema… ante esto cabe preguntar: ¿Qué clase de premisa es P y de donde ha salido a colación? No parece ser una premisa empíricamente testeable, y aun si lo fuera para los humanos según alguna interpretación especial, no lo sería en ningún caso para los Antípodas. Tampoco es una definición útil para los Antípodas que consiga, por ejemplo, acercarlos a la idea de una "mentalidad" en sentido humano, puesto que está constituida en términos ontológicos incompatibles con la taxonomía fisicalista de los Antípodas. Finalmente tampoco es una proposición para la cual se haya dado ninguna forma de justificación especial distinta de las instancias que la propia premisa P supone definir (instancias de sensaciones primarias o categorías mentales), ni es una premisa evidente *per se*, especialmente para los Antípodas. ¿Por qué insistirían entonces en traer a colación los humanos una premisa tan controversial y tan explícitamente inservible para resolver el problema de los Antípodas? Al parecer esta premisa es traída a colación no para jugar un rol justificativo dentro de algún argumento, sino para sentar las bases de toda la discusión. Esto parece evidente cuando notamos que P es, en verdad, el resumen de toda nuestra teoría mentalista, teoría –que según Rorty (1979)- tiene su origen histórico en Descartes y ha sobrevivido, a punta de intuiciones, en gran parte del lenguaje filosófico como una herencia tácita y pocas veces discutida. P no viene, en resumen, a defender lo mental, viene a *postularlo*.

Concluimos así la historia imaginaria de humanos y Antípodas y de sus infértiles discusiones acerca de lo mental.

[2] Esta premisa "humana" es fundamental en el ejemplo de Rorty y servirá también más adelante, por lo que pasaremos a llamarla (como hace el propio autor) "P".

Ahora bien: ¿Qué moraleja debemos sacar de esta fabulosa historia?... según creo estas son las moralejas:

Moraleja 1: que el mentalismo parece ser la elección arbitraria de un determinado lenguaje (el lenguaje fundado en la premisa P) más que una discriminación ontológica criteriosa o una buena teoría acerca de los hechos.

Moraleja 2: que las categorías y el lenguaje mentalista no aportan nada (o al menos no se vislumbra claramente por qué deberíamos creer que aporta algo) en la investigación acerca del funcionamiento de los cerebros y su relación con el conocimiento, las actividades intencionales, la autoconsciencia, los fenómenos sociales, etc., cuento corto, en la identificación de lo que llamamos "personas".

Moraleja 3: que, si se abandona el supuesto P, los estados "incorregibles" –que pretendían caracterizar definitivamente lo mental- para un determinado grupo histórico, usuario de un lenguaje (grupo que no está versado en absoluto o sólo muy superficialmente en neurociencia por ejemplo), pueden resultar "corregibles" para seres con los conocimientos teóricos suficientes (los Antípodas por ejemplo, o nosotros mismos en el futuro). En resumen, que podemos estar equivocados acerca de lo que *creemos nunca estar equivocados*, que esta es una posibilidad imaginable.

Moraleja 4: que si el experimento mental de los Antípodas es imaginable y no un sin sentido (esto último sería una prueba que caería del lado de los mentalistas) es en principio consebible que las categorías mentales y su lenguaje implícito puedan ser falsas, eliminables y reemplazables por una neurociencia más poderosa que la actual y su respectivo lenguaje.

Visto brevemente y a modo de introducción el no poco extravagante experimento mental propuesto por Rorty –que por cierto nunca osé emplear para defender al ME ante un público no acostumbrado a esta clase de bizarras argumentaciones tan comunes en filosofía de la mente- y dilucidadas ya las principales "moralejas" que de este ejemplo podemos extraer a favor del eliminativismo, pasamos a revisar tanto la estructura como la defensa del ME. Por estructura entenderé las diversas tesis que deben asumirse para que el argumento eliminativista sea concluyente, y por defensa entenderé tanto las réplicas que el ME puede esgrimir ante las objeciones más renombradas como las críticas específicas que pueden surgir desde este programa para con otros programas competidores en filosofía de la mente.

Capítulo 1. El argumento y los supuestos del materialismo eliminativo

El materialismo eliminativo (ME) es una tesis sustentada por un argumento, a primera vista, sencillo:

Premisa 1: La psicología folk es una teoría (es decir, no tiene criterios de evaluación epistemológica o metodológicamente distintos de los que pudiera presentar cualquier teoría empírica).

Premisa 2: Las mejores teorías disponibles (neurociencia, psicología empírica, psicología cognitiva) taxonomizan el mundo de forma incompatible con la psicología folk.

Conclusión: La psicología folk es una teoría falsa y debe ser abandonada (junto con la mayoría de su vocabulario).

Esto no implica que debamos de la noche a la mañana reemplazar nuestro cómodo lenguaje "folk" acerca de la psicología (tal y como no hemos reemplazado nuestro lenguaje de "física folk"[3]) en su conveniente uso cotidiano, sino que debe ser eliminado de las discusiones teóricas serias (lo cual no cierra la posibilidad de que este vocabulario "cómodo" desaparezca de hecho del lenguaje cotidiano tras popularizarse la nueva jerga más exacta).

Generalmente se apoya la posibilidad de esta eliminación poniendo vastos ejemplos presuntamente equivalentes en la historia de las ciencias naturales: fluido calórico, flogisto, éter, geocentrismo, casi todos los principios de la física aristotélica, alquimia, etc. La aparición de teorías más potentes explicativa y predicativamente, más coherentes y más simples –por nombrar algunos factores estándar en la evaluación de teorías- han desencadenado en el completo abandono de las teorías en cuestión y en el desuso de su vocabulario al menos en las discusiones especializadas (exceptuando casos de "redefinición" de algún término popular).

Pese a esto el argumento del materialismo eliminativo no es tan sencillo como lo hemos presentado. Analizaremos a continuación con algún detalle lo que considero (siguiendo a Perez (2006)) los tres supuestos fundamentales para el eliminativismo:

A) "Las visiones de mundo del sentido común son, en general, teorías y deben ser evaluadas como tal"

Ya hemos hecho explícito este supuesto en el argumento de arriba pero lo repetiremos y detallaremos por ser en extremo relevante para las tesis en cuestión. Lo que este supuesto afirma son al menos dos cosas: 1) Que todo discurso acerca de ciertos hechos del mundo que posea –aun tácitamente- ciertas leyes o enunciados lo suficientemente generales como para cumplir este rol, y que cuente con explicaciones y predicciones, es una teoría empírica[4] siguiendo el modelo nomológico-deductivo. 2) Toda teoría empírica tiene los mismos criterios generales de evaluación y es, en principio, descartable y eliminable como

[3] Esto es nuestra física macroscópica. También podríamos decir que poseemos una "química folk" cuando cocinamos o una "biología folk" al curarnos una herida.

[4] No creo que haya algún ejemplo de alguna teoría folk que sea una ciencia formal, pero quien sabe.

teoría falsa en presencia de cierta evidencia en su contra (incoherencia taxonómica con otras teorías bien establecidas, mejores teorías con el mismo poder explicativo, etc.).

B) “La referencia de un término teórico está dada por su rol inferencial dentro de una teoría dada”

Esta tesis es conocida como teoría del rol inferencial de la referencia o teoría reticular en semántica. Según ella no existe un término al que podamos dar significado fuera de una determinada teoría. Esta condición alcanza a convertirse en presupuesto para las tesis del materialismo eliminativo tanto porque permite hablar, sin caer en sin sentidos, de la desaparición (eliminación) de cierto vocabulario –por muy arraigado que esté en las prácticas- y hasta lo justifica. Es además una condición suficientemente fuerte para considerar a la psicología folk como una teoría. Además muchos de los contra argumentos esgrimidos contra A) acaban por discutirse al nivel de B). Podríamos decir que las dos primeras razones que constituyen a B) como presupuesto para el ME son “estructurales” y que el último es simplemente “justificativo” (o “defensivo”).

C) “¡No más lenguajes privilegiados!”

Con este “grito de guerra” intento resumir las conclusiones alcanzadas especialmente por Quine y Sellars (y bien expresadas en Rorty (1979)) acerca de la imposibilidad de probar el privilegio epistemológico de ningún lenguaje en particular. Este cuento es largo y bien discutido, aunque sus dos hitos fundadores y arquetípicos son señalables: 1) el ataque quieniano a la analiticidad y a la reducción propia del fundacionalismo. 2) El rechazo sellarsiano al mito de lo dado.

En filosofía de la mente, en particular, este supuesto concluye algo de gran valor para el materialista eliminativo… ¡Que los conceptos mentales en primera persona no son, *prima facie*, mejores que otros ni son en absoluto necesarios para constituir un vocabulario teórico correcto en esta clase de discusiones!... en este sentido C) también es un supuesto estructural para la tesis del ME.

A su vez este supuesto –como mostraremos- acaba siendo el último bastión de batalla contra gran parte de los argumentos que pretenden desestimar la tesis eliminativista (por lo que también es “defensivo”).

Analizamos a continuación una selección que considero estructuralmente significativa (aunque de seguro puntualmente incompleta) de argumentos en contra del ME con el fin de ver hasta qué punto consiguen su objetivo y a qué nivel se resuelven (o no).

Capitulo 2. Las objeciones y los supuestos del materialismo eliminativo

En general la mayor cantidad de argumentos contra el ME se han concentrado en tratar de rechazar A). Esto parece obvio tomando en cuenta que A) es el único supuesto que se presenta explícitamente en la formulación del argumento eliminativista. Sin embargo, como intentaré mostrar, A) es el supuesto más ligero y el más dependiente de los tres presentados. Los ataques contra A) se transportan casi siempre a una discusión referente a los otros dos supuestos y en general se resuelven únicamente en estas instancias. Intentaré probar, en general, que los supuestos eliminativistas no son independientes unos de otros, si no que se agrupan en una cadena justificativa (o "defensiva"), comenzando por A) y terminando en C).

1) Un primer y bastante ingenuo argumento es sostener que la introspección interna nos muestra claramente la existencia de actitudes proposicionales y qualia. El punto fuerte será replicar al materialista eliminativo diciéndole que tenemos experiencia directa de estados como dolores, etc. Este argumento es claramente una petición de principios, pues supone un conocimiento bien establecido de lo que se pretende discutir. Sin embargo no parece refutar A) directamente –puede suponer que nuestros estados internos son buenas instancias observacionales (aunque no públicas) para probar que la psicología folk es una teoría verdadera-, tampoco se hace cargo de B) explícitamente (salvo si especifica una semántica basada en la ostentación directa de estados mentales), e ignora en general C). Desde B) podemos argumentar que muchas veces el marco de referencia afecta los objetos de las observaciones, y para esto podemos dar bastantes ejemplos históricos (la esfera giratoria del cielo, etc.). Y C) es la refutación misma de la posibilidad de establecer estos estados epistemológicamente privilegiados como base de las teorías. Veamos esto último con más detalle.

La fuerza de una argumentación como esta está basada en la supuesta infalibilidad del conocimiento de los estados internos. Con la introspección no estamos captando una sensación o una impresión mediante alguna otra sensación o impresión (como suponemos se conocen los objetos del mundo externo), sino que estamos en relación directa y no mediada (a su vez no inferencial) con el objeto de nuestro conocimiento. De este modo, se supone, no se puede ser víctima de una impresión o una sensación falaz… no puedo "creer que sentía dolor" o "creer que creía *p*" y estar eventualmente equivocado. La distinción entre apariencia y realidad válida en el conocimiento de los eventos u objetos externos es inexistente en esta esfera. No se puede estar equivocado acerca de si se siente o no un dolor, porque un dolor, simple y llanamente, se siente o no se siente, y punto (*ser* un dolor es *ser sentido* como tal). Si estos "privilegios epistemológicos" son inobjetables parece existir una buena base para negar B): mis estados internos no juegan rol inferencial alguno dentro de una teoría, y esto, presuntamente, porque son conocidos certeramente con anterioridad a la formulación explícita de toda teoría posible. Si esto es así, tampoco A) tiene buenas perspectivas de aceptación, porque si bien la psicología folk puede eventualmente ser formulada como una teoría (ser ordenada mediante ciertas leyes, etc.), por el carácter epistemológico único de sus objetos constituyentes, la psicología popular será siempre *algo*

más que una teoría[5]. Las entidades de esta "teoría" no serían nunca eliminables, puesto que no han sido postulados en beneficio de la teoría misma, sino que han sido instanciados directamente en la introspección.

Toda la discusión se vuelca, como sospechábamos antes, hacia C). ¿Tienen en verdad un estado privilegiado los conocimientos introspectivos? Churchland (1988) da buenas razones para poner en duda todo esto.

En primer lugar esta doctrina supone que la mente es "transparente" a si misma, que se conoce a si misma con ventaja por sobre las demás cosas "externas". Esta idea sin duda va en contra de una perspectiva naturalista y evolucionista de los hechos. Los cerebros fueron seleccionados por otorgar cierta ventaja reproductiva a sus poseedores, y estas ventajas son a toda vista principalmente capacidades de distinción, reconocimiento y control del *mundo externo* (reconocer el alimento, a los depredadores, etc.). Si los cerebros han llegado a ser beneficiarios de la selección natural es precisamente por este rasgo especial. Si el autoconocimiento ayuda y privilegia el conocimiento del mundo externo (como presumiblemente lo hace) tendrá una explicación evolutiva coherente, aunque este nuevo rasgo seguirá siendo una aparición incidental destinada a mejorar el conocimiento externo, pues es éste el aspecto de mayor importancia para la supervivencia. Lo que parece fundamental en un cerebro es que pueda dar cuenta en alguna medida del mundo externo, pero si suponemos una dicotomía tan marcada entre "estados internos" y "estados externos" resulta evolutivamente misterioso que el cerebro haya desarrollado un aparato de reconocimiento y distinción de "estados internos" superior al que le brindaría mayores ventajas evolutivas. Pero si suponemos que nuestro aparato de distinción de "estados internos" no es más que una función específica y segundaria del proceso central de conocimiento del mundo externo el misterio se desvanece. Sea como sea, si ambas formas de conocimiento (sean o no sustancialmente distintas) son productos de la evolución de nuestro cerebro, no serán cosas fundamentalmente distintas y no existirían buenas razones para suponer que ninguna fuese infalible. Pero analicemos esta doctrina en su propio terreno. Ella sostiene que "no hay ninguna forma posible de error respecto de nuestros juicios acerca de nuestras propias sensaciones, impresiones, etc.", pero este "no hay" es en verdad un "no hemos imaginado aun", es decir, es un *argumentum ad ignorantiam*, pues no poder señalar ningún error en esta clase de juicios habla quizás de nuestra mediocre comprensión de estos estados o de la poca investigación práctica en este respecto o incluso, del completo error en el tratamiento del tema en general, más que de un estado de conocimiento infalible.

Pero además no parecen, en verdad, inimaginables estos estados de error en el juicio acerca de estados internos. Pensemos en la supuesta desaparición de las categorías "apariencia" y "realidad" en los juicios acerca de sensaciones. Esta desaparición puede invocarse sólo si se supone que en cada caso el error viene dado por un elemento mediador (y en general, un elemento mediador que sea una sensación[6]). Pero en verdad nada garantiza que un juicio introspectivo tal como "siento dolor" sea provocado exclusivamente por la aparición de dolores:

"Consideremos la aparición de algo bastante similar al dolor –una súbita sensación de extremo frío, por ejemplo- en una situación en la que se espera intensamente sentir dolor. Supongamos que usted es un espía que ha sido

[5] Esta objeción reaparece insistentemente, especialmente en el punto 2) de este mismo capítulo.

[6] Y claro está, se supone que no hay "sensaciones de sensaciones".

capturado y que se lo interroga extensamente con ayuda de un hierro caliente que se le aplica varias veces en la espalda a breves intervalos repetidos. Si, en el vigésimo intento, disimuladamente se le aplica en la espalda un cubo helado, su reacción inmediata diferirá poco o nada de las primeras diecinueve reacciones. Casi seguramente usted creería, por un instante, que siente dolor. El partidario de la imposibilidad de corregir estos juicios tal vez insista en que la sensación número veinte era un dolor, después de todo, a pesar de su causa benigna, sobre la base de que, si usted lo consideró como un dolor, si usted creyó que le resultaba doloroso, entonces verdaderamente era un dolor. Esta interpretación no se corresponde muy bien con el hecho de que uno puede recuperarse de esos tipos de mala identificación que acabamos de mencionar. El chillido inicial de horror da lugar a algo como: "Espere… espere… no es la misma sensación que antes. ¿Qué sucede aquí atrás?" Si la sensación número veinte realmente fue un dolor, ¿por qué nuestro juicio se invierte a los pocos segundos?" (Churchland. 1988. Pag 121-122)

Esta clase de errores se denominan *efectos de expectativa* y constituyen un fenómeno generalizado y bastante bien estudiado[7]. Estos efectos de expectativas nos brindan un marco muy amplio para generar muy diversas clases de errores de identificación tanto de estados internos como externos.

Otros efectos que brindan amplias posibilidades de errores en la identificación son los *efectos de presentación*. Los juicios correctos acerca de ciertas cualidades introspectivas parecen disminuir su cantidad si el tiempo de la sensación, a su vez, disminuye o bien si el sujeto no está en condiciones normales (digamos, está drogado o agotado).

Además de todo esto y si bien la incorregibilidad de nuestros juicios acerca de sensaciones parecía, en sentido lato y poco crítico, verosímil, esta condición no está ni cerca de ser extensible a otros estados mentales como deseos, creencias, emociones, etc. La mayoría de las personas son relevantemente malas al tratar de juzgar cada uno de estos estados mentales, por ejemplo, para saber si se sienten rencorosas, celosas, o airadas, o distinguir claramente sus deseos y expectativas de sus simples creencias[8].

Ciertamente el carácter infalible acerca de esta clase de juicios no ha sido frecuentemente defendido, pero restringir este carácter a los juicios acerca de sensaciones es problemático, ¿acaso las emociones y los deseos están más "mediados" que las sensaciones?

Las críticas descritas sin duda pueden encasillarse dentro del supuesto C), y presumiblemente existen recelos aun menos superficiales que podrían esgrimirse. Con lo expuesto, sin embargo, tenemos suficiente material para desacreditar el privilegio epistemológico supuesto a la introspección interna. Si este es el caso, tampoco habrá buenas razones para suponer que los conceptos empleados en esta clase de juicios tienen algún estatus epistemológico superior que las resguarde de ser interpretadas fuera de toda teoría. Además sabemos a qué teoría (o "prototeoría") pertenecen estos enunciados, tal es

[7] Otro ejemplo incluye la expectativa de experimentar un sabor determinado (naranja por ejemplo) y ser estimulado con otro sabor muy similar (limón por ejemplo). Los sujetos suponen estar saboreando algo con sabor a "naranja" y sólo corrigen su juicio al ser estimulados verdaderamente con este sabor.

[8] Mucha gente cree que "cree que *p*" sin saber que en verdad sólo "quiere que *p*", y mucha gente cree que "quiere que *p*" sin saber que en verdad sólo "cree que *p*". Si nada de esto ocurriese, la psicología terapéutica ni siquiera existiría.

la psicología folk. Luego, y gracias en general a C), la introspección interna (además de constituir una petición de principios como objeción) no consigue probar que sus enunciados y sus términos están fuera de toda interpretación teórica (B)) ni que no existe una teoría específica en la que juegan un rol inferencial relevante (A)).

2) Uno de los argumentos usuales contra el ME (más bien un conglomerado de argumentos) es atacar directamente A) suponiendo que nuestras intuiciones de sentido común no son una teoría. Estas intuiciones caerían dentro del estatus de las "certezas" wittgensteinianas o mooreanas. Estas "certezas" son caracterizadas por Rabossi[9] como convicciones comunes a la mayoría de los seres humanos, que son aceptadas compulsivamente, que poseen una estabilidad paradigmática y que generan una particular incosistencia ante el menor intento de abandono.

Sin entrar en los detalles propios de estas "certezas", es claro que este argumento nos lleva por camino derecho hacia la tesis de la incorregibilidad de nuestras teorías folk. Lo primero será entonces señalar que esta tesis, tomada al pie de la letra, debería suponer inamovible las intuiciones folk sobre todos los aspectos mundanos. Pero la ciencia ya ha corregido bastantes de estas "certezas inamovibles" (geocentrismo por ejemplo), suponer *ex post facto* que las antiguas *certezas* actualmente corregidas no eran en verdad *certezas* parece ser un giro falacioso en la argumentación. Además no estamos en general dispuestos a asumir de buenas y primeras algo de tan grueso calibre como que nuestras mejores ciencias (nuestras mejores teorías) no tengan en verdad nada que agregar y transformar en la visión cotidiana de las cosas.

Pero quizás hemos avanzado muy groseramente en este punto. Searle (1992) nos propone un análisis más fino del problema de la psicología popular en una serie de tesis y antítesis al respecto que pretenden ser una defensa contundente de la PP (psicología popular). Procedemos a exponer y discutir el análisis de Searle que puede sernos de gran utilidad[10]:

"Tesis: la PP (psicología popular) es una tesis empírica como cualquier otra, y como tal está sujeta a confirmación y disconfirmación empírica. Respuesta: las capacidades efectivas que las personas tienen para habérselas consigo mismas y con los demás no tienen, en su mayor parte, forma proposicional. Son, en mi opinión, capacidades de Transfondo. Cómo respondemos a las expresiones faciales, qué encontramos natural en la conducta, e incluso cómo entendemos las emisiones son, para poner un ejemplo, asuntos que tienen que ver, en gran medida, con el saber-cómo y no con teoría alguna. Se distorsionan esas capacidades si se piensa en ellas como en teorías." (Searle. 1992. pag 71.)

Esta primera respuesta no parece más que una elaboración más fina de la objeción laxamente descrita que mencionamos más arriba y encontrará, en general, idéntica réplica. ¿Por qué nuestro "Transfondo" familiar *debe* sobrevivir ante nuestras mejores teorías?, ¿Qué garantías epistemológicas y qué fuerza empírica posee la renombrada familiaridad de nuestras intuiciones –sean acerca de la mente u otras-? En general replicamos que muchos programas de investigación empíricos se han visto forzados a abandonar intuiciones bien establecidas (piénsese en la física en su estado actual) y que, en general, estos programas

[9] Las referencias a los textos de Rabossi están en Perez (2006)

[10] Al expresar Searle sus objeciones en forma de tesis (en contra de la psicología folk) y réplicas (a favor de la psicología folk), omitiré los argumentos a favor del ME que el propio Searle expresa en las tesis.

no son libres de elegir las intuiciones que rechazan mediante ningún criterio externo distinto de la compatibilidad y consistencia de esas intuiciones con la teoría.

"Tesis: sin embargo, se podría enunciar correlatos o principios teóricos que subyacen en esas capacidades. Esto constituiría una psicología popular y sería con toda probabilidad falsa, puesto que, en general, las teorías populares son falsas. Respuesta: Se puede, no sin alguna distorsión, enunciar un correlato teórico de una destreza práctica. Pero sería milagroso el que éstas sean, en general, falsas. Donde realmente son importantes, donde algo está en juego, las teorías populares tienen que ser, en general, verdaderas o, de lo contrario, no habrían sobrevivido. La física popular puede ser errónea sobre cuestiones periféricas como el movimiento de las esferas celestes o el origen de la Tierra, porque estas cosas no son demasiado importantes. Pero cuando se trata del modo en que se mueve nuestro cuerpo si uno salta desde un acantilado, o de lo que sucede si una gran roca le cae a uno encima, las teorías populares tienen que ser correctas o, de lo contrario, no habrían sobrevivido." (Searle. 1992. Pag 72)

Esta perspectiva supone de antemano la aceptación del punto establecido anteriormente. Pero aun si concediéramos esto, no parece evidente lo replicado por Searle. En primer lugar supone que las teorías folk tienen algún privilegio epistemológico de verificación en su "práctica", cuando "se pone algo en juego" y funcionan, pero no alcanzo a ver qué criterio objetivo –más allá del sentido más común imaginable- pueda darse para juzgar el éxito o fracaso de una hipótesis en cualquier teoría popular. La medicina más básicamente folk que podamos imaginar nos dirá que "no comamos cosas que nos causan dolor de estómago y náuseas". Esta guía práctica seguramente dominó vastas generaciones que pretendían evitar problemas estomacales, pero, por otra parte, muchos alimentos que causan malestares estomacales y otros síntomas intuitivamente insanos ayudan en verdad al enfermo. Las leyes de una teoría popular (puestas, de todas formas, en duda por Searle) estarían constituidas de tan gigantescos *ceteris paribus*, y de tantos "siempre y cuando" que acabarían sin duda fallando sólo "de milagro" como dice el autor, pero ¿este éxito garantizado por ambigüedad y laxitud las hacen buenas leyes de buenas teorías?[11]

Se replicará tal vez que en verdad sólo estamos describiendo el paso de una teoría folk (o Transfondo) válida en general, a través de sus excepciones, hacia una teoría sofisticada. Muchos de estos pasos hacia "teorías sofisticadas" son dados, sin embargo, mediante intuiciones tan comunes y silvestres como cualquier otra perteneciente a nuestras certezas usuales. No creo que puedan precisarse con exactitud los grados de familiaridad que tenemos en cada caso con determinados enunciados o determinadas prácticas, y además ¿no llegan a sernos familiares e intuitivas ciertas conductas, interpretaciones o prácticas por mero entrenamiento o costumbre?... las figuras geométricas tridimensionales dibujadas, por ejemplo, no nos obligan *prima facie* a determinar qué cara de la figura tomaremos como "fondo", pero ¿no podríamos eventualmente condicionar mediante estímulos a un niño para que elija siempre una alternativa fija en este respecto?, ¿y no le parecerá a él esta alternativa finalmente algo certero e intuitivo?

[11] No mencionaré mayormente la referencia de Searle a la "supervivencia de las creencias" de la PP. Quizás podríamos por nuestra parte poner la supervivencia de estas creencias en función de su utilidad a nuestra supervivencia biológica (es decir darle un enfoque evolutivo a todo el problema)… sea como sea, ambas formas de enfocar el problema me parecen muy apresuradas.

Todo lo que parece estar en juego en la tentativa de no juzgar a la psicología folk como teoría o disminuir su peso teórico es su carácter incorregible[12], pero este carácter es subjetivo y no da garantes de certeza inamovible (como pretende usárselo).

En segundo lugar el éxito de una determinada predicción no garantiza la corrección de la teoría desde donde fue implicada. Puede existir, por ejemplo, cierto rango de éxito en la práctica de la acupuntura o de las curaciones rituales mediante hierbas u otros elementos tradicionales, pero esto no nos compromete con las entidades propuestas por los usuarios de estas prácticas (energías místicas, espíritus favorables, etc.). Así mismo no estamos forzados a comprometernos con la existencia del objeto "bóveda giratoria celeste" por la utilidad náutica que pudo tener en la antigüedad. Sería, en resumen, misterioso que no existiera una relación causal entre una práctica y su éxito (supuesto que podemos definir estos términos claramente –tendría que ser dentro de una teoría probablemente-), pero no es misterioso ni inusual que ciertas prácticas exitosas sean producto de entidades postuladas e inexistentes. La supervivencia de *toda* la teoría folk es usual con sólo algunas relaciones prácticas exitosas que pueden ser de poco interés para determinar la verdad de toda o gran parte de los postulados de la teoría.

No resultaría raro, en resumen, que una teoría sea falsa, sería simplemente raro que sus prácticas funcionaran sin explicación alguna proveniente de alguna otra buena teoría.

Tesis: se convierte ahora en un asunto específico de la ciencia cognitiva (CC)[13] el decidir qué tesis de la PP son verdaderas y cuáles de sus compromisos ontológicos están justificados. La PP postula, por ejemplo, creencias y deseos para dar cuenta de la conducta, pero si resulta que la explicación de la conducta por parte de la CC es inconsistente con esto, entonces las creencias y deseos no existen. Respuesta: casi todo el contenido de esta afirmación es erróneo. En primer lugar no postulamos creencias y deseos para dar cuenta de nada. Simplemente experimentamos creencias y deseos concientes. Piénsese en ejemplos de la vida real. Hace un día tórrido y vas conduciendo un camión a través del desierto en las afueras de Phoenix. No tienes aire acondicionado. No puedes recordar cuando has estado tan sediento y deseas con toda tu alma una cerveza fría. Ahora bien, ¿dónde está la "postulación" del deseo? Los deseos conscientes se experimentan. No se postulan en mayor medida que los dolores conscientes. (Searle. 1992. Pag 72)

Esta réplica no aporta, sin embargo, nada a la discusión. No se distingue en verdad de la objeción 1) en contra del ME. Se establece que deseos y creencias no son postulados por no presentársenos de este modo en nuestra vida mental ordinaria, pero esto sólo podemos traerlo a colación por ostención interna de nuestros estados mentales. Luego hemos presupuesto que hay algo así como esos estados y que nuestra forma epistémica de acceder a ellos es lo suficientemente privilegiada como para esgrimir esta instancia como prueba decisiva. Se prueba que la psicología popular no es una teoría apelando al modo particular y no teórico en el que instanciamos ciertas entidades, sin caer en cuenta que si el ME pretendía considerar a la psicología folk como una teoría empírica más, era precisamente para negar la existencia de tales entidades.

Hasta donde hemos llevado la discusión es seguro que no "postularemos" actitudes proposicionales o qualias, quizás porque simplemente nos acaecen de forma incorregible,

[12] Es decir al mismo eje en el cual desencadena la discusión propuesta por Rorty en su experimento metal de los Antípodas.

[13] ***Un materialista eliminativo agregaría "o bien neurociencia avanzada –o completa-"***

privada y subjetivamente, o bien porque *no tenemos en verdad tales cosas*. ¡Pero éste es precisamente el punto a decidir! Al materialista eliminativo poco le importa como se sienta o se deje de sentir la aparición de una creencia o un dolor (o si se "postula" o no subjetivamente)… ¡él ni siquiera cree que estas cosas existan realmente!

Si se quiere discutir con el ME en algún sentido relevante se debe evaluar la psicología folk como teoría empírica o bien probar, sin círculo en la prueba –esto es, sin apelar a sensaciones primarias- que la psicología popular no es en verdad una teoría. Esto último no tiene muy buenas perspectivas de éxito, pero la primera vía de discusión está abierta. Para poder llevar a cabo este análisis, en cualquier caso, es preciso contar a las creencias y deseos como entidades teóricas, es decir, como postulados dentro de una red explicativa y predictiva. Si esto último se descarta de antemano no hay discusión ni defensa relevante de la psicología popular.

"Tesis: sea lo que sea, se postulen o no, es muy poco probable que haya una reducción adecuada de las entidades de la PP a la ciencia más básica de la neurobiología, de modo que parece que la eliminación es la única alternativa. Respuesta: ya he dicho hasta qué punto este es un mal argumento. Muchos tipos de entidades reales desde los chalets adosados a las fiestas de cumpleaños, desde los tipos de interés a los partidos de fútbol, no soportan una reducción apropiada a entidades de alguna teoría fundamental. ¿Por qué deberían soportarla? Sospecho que tengo una "teoría" de las fiestas de cumpleaños – por lo menos en la medida en que tengo una teoría de la "psicología popular"- y las fiestas de cumpleaños consisten ciertamente en movimientos de moléculas; pero mi teoría de las fiestas de cumpleaños no es, ni por aproximación, tan buena como mi teoría física molecular, y no hay una reducción de las fiestas de cumpleaños a la taxonomía de la física. Pero a pesar de todo, las fiestas de cumpleaños existen realmente. La cuestión de la reductibilidad de tales entidades es irrelevante para la cuestión de su existencia" (Searle. 1992. Pag 73)

Ciertamente nos parece que algo se nos escapa al dar una descripción molecular o física en general de un partido de fútbol, una fiesta de cumpleaños o las carcajadas producidas por un chiste obsceno (suponiendo que esta clase de reducción fuera posible), pero todas estas cosas tienen algo en común: lo que se nos escapa en las descripciones puramente físicas, químicas o biológicas de estos fenómenos son características ligadas con la intencionalidad, la conducta inteligente, la conciencia… en una palabra, lo mental. Tomemos para ejemplificar esto un objeto simple, una silla, ¿qué puede escapársenos acerca de una silla en una descripción puramente física? Sin duda su función como lugar de relativo descanso para los seres humanos. Esta "función" de la silla es sin duda una característica disposicional únicamente señalable por algún ser conciente que entienda para qué diantres sirven las sillas. Luego resulta natural que sólo un ser capaz de tener estados intencionales pueda dar cuenta de aquello que a las teorías de las ciencias naturales se le escabulle con respecto de una silla[14].

Lo sospechoso de este argumento es que todo lo que se supone irreductible (y además indiscutiblemente real) tiene la característica común de no ser reductible exclusivamente en relación a esta clase de propiedades disposicionales, es decir, a las propiedades que necesitan estar en relación *directa y no-teórica* con estados intencionales. Podríamos decir que esta respuesta de Searle supone una línea horizontal bajo la cual estarían las

[14] Putnam (1987) tiene una opinión similar y hasta más radical que discutiremos más adelante en el capítulo 3.

propiedades reductibles por las ciencias naturales y sobre la cual se asentarían propiedades disposicionales sólo caracterizables desde alguna forma de intencionalidad… la línea misma será, sin lugar a dudas, los estados mentales intencionales, digamos, la conciencia. Esto nuevamente comete la impostura de reintroducir lo mental de antemano, si bien este argumento tiene al menos la fuerza de desafiar al ME a mostrar cómo reducirá toda la fauna de objetos cotidianos a las ciencias naturales.

Pero lo cierto es que el ME no está obligado a realizar *de hecho* esta reducción. Si el ME está en lo correcto muchas de las propiedades ligadas a la mentalidad supuesta por la psicología folk pueden desaparecer. Esto no implica que todos los objetos ordinarios deban desparecer también. Muchos de ellos podrían encontrar eventualmente una reducción coherente con las mejores teorías naturales, pero para esto tendríamos que contar al menos con una teoría natural de los fenómenos perceptivos, conductuales, sociales, etc, dependiendo del caso. Así un partido de fútbol puede parecernos hasta la fecha algo inexpresable en términos de las ciencias naturales por propiedades tales como "el deseo de ganar de los participantes", pero supuesto que no hay tal cosa como "deseos" aun podría existir una explicación neurobiológica de por qué "algunos organismos con tales y tales sistemas nerviosos corren tras un balón intentando hacerlo traspasar una línea que ellos mismos trazaron" (digamos, una explicación evolutiva). Suponer que sólo desde la psicología folk podemos dar cuenta de *todas* las disposiciones presuntamente ligadas a la conciencia, la intencionalidad, las actitudes proposicionales, etc, es cantar victoria antes de tiempo. Es probable que, si la tesis de la eliminación propuesta por el ME fuera verdadera, muchas entidades que nombramos usualmente (como el cielo azul) y muchas propiedades de entidades (como la "comodidad" de una silla) serían drásticamente eliminadas de nuestro vocabulario, pero esto no puede aplicarse *a priori* a todas ellas (entidades y propiedades). Muchas sí podrían ser eventualmente caracterizadas en términos no disposicionales al profundizar nuestro conocimiento neurobiológico para así encontrar la ligadura entre los fenómenos social y culturalmente determinados (fiestas de cumpleaños, partidos de fútbol, etc.) y las ciencias más básicas. El que esta ligadura entre estas dos clases de objetos presuntamente incompatibles esté hoy en día realizada –defectuosamente vale decir- por las categorías mentales no demuestra la imposibilidad de una teoría más poderosa de hacer el mismo trabajo o, al menos, buena parte de éste.

Esta clase de "argumentación futurista" por parte del ME es, a mi juicio, la forma más débil de argumentos a favor de la tesis de la eliminación, sin embargo me parece suficiente al menos para defender el estatus teórico de la psicología popular.

En resumen no estamos seguros de si: 1) Todas las entidades cotidianas ("como un concierto de piano", "el asombro de una multitud", "un partido de fútbol" o "un fraude electoral") o sus propiedades deban seguir teniendo el mismo estatus ontológico supuesta una eliminación de las categorías mentales, y no sabemos tampoco si esto perjudica en verdad a la ontología (sabemos que perjudica a la ontología de sentido común, pero esto es trivial y no viene al caso para la discusión). Ni estamos seguros de si: 2) Todas las entidades y propiedades identificadas mediante categorías mentales son de hecho irreductibles al vocabulario de las ciencias naturales sin apelar a disposiciones relacionadas con estados intencionales o mentales en general.

Si bien la argumentación del ME es futurista y tiene pruebas parciales, la de de Searle no es más que un *argumentum ad ignorantiam*[15].

[15] La caracterización de este argumento que será expuesta por Putnam y que trataremos más adelante puede brindar más apoyo a esta tesis.

"Tesis: sí, pero lo que estás diciendo pide la cuestión. Estás diciendo que creencias y deseos, al igual que las fiestas de cumpleaños y los chalets adosados, no son entidades teóricas –su base evidencial no se deriva de teoría alguna-. ¿Pero no es este uno de los puntos en disputa?" Respuesta: pienso que es obvio que creencias y deseos se experimentan como tales y que, ciertamente, no se "postulan" para explicar la conducta porque no se postulan en absoluto. Sin embargo, ni siquiera las "entidades teóricas" alcanzan su legitimidad de la reductibilidad. Considérese la economía. Los tipos de interés, la demanda efectiva, la propensión marginal al consumo son todas ellas cosas a las que se hace referencia en economía matemática. Pero ninguno de los tipos de entidades en cuestión soporta una reducción apropiada a, por ejemplo, la física o la neurobiología. Y de nuevo, ¿por qué habrían de soportarla? La reductibilidad es, en cualquier caso, una exigencia extraña para la ontología, puesto que clásicamente una manera de mostrar que una entidad no existe realmente ha sido reducirla a algo distinto. Así, las puestas de Sol se reducen a movimientos planetarios del sistema solar, lo cual mostraba que, como se concebía tradicionalmente, las puestas de Sol no existen. La apariencia de que el Sol se pone viene causada por algo distinto, esto es: por la rotación de la Tierra en relación con el Sol. (Searle. 1992. Pag 74)

Este argumento adolece del mismo defecto que el anterior. Searle vuelve a presentar un caso de supuesta imposibilidad de reducción –aunque esta vez en una teoría bien desarrollada como es la economía- en un ámbito del conocimiento que se sitúa por encima de la "línea" de lo mental. *Probar* la irreductibilidad de la economía a las ciencias naturales más básicas no es posible hasta probar la completa incorrección del afán de eliminación de las categorías mentales por parte del ME. Lo único que Searle llega a demostrar en estas disquisiciones es que los fenómenos (sean actualmente considerados teóricos o no) que más se alejan de las explicaciones de la física elemental presentan un panorama de reducción más complejo –cosa archisabida-, pero esto no demuestra que: 1) Todas las teorías o fenómenos constituidos actualmente mediante alguna forma intencional de lenguaje sean irreductibles a las ciencias naturales más básicas. Ni 2) Que la psicología popular no sea una teoría. Ni 3) Que cualquier fenómeno que esté "por arriba de la línea de lo mental" debe ser necesariamente eliminado según el programa del ME.

Por otra parte, declara el autor constatar cierta extrañeza en las pretensiones ontológicas de la reducción. Searle se guía, con buen garante académico, por el viejo *definire est eliminare*, y si bien esto es cierto, no lo es necesariamente su viceversa. Uno de los objetivos generales de una reducción interteórica es seguir empleando un determinado término o un determinado vocabulario (generalmente ambiguo o relativamente inexacto) con la certeza de poder traducirlo cuando fuere necesario a un segundo vocabulario o términos bien establecidos y más exactos. Generalmente la comodidad y simplicidad del lenguaje reducido le permite sobrevivir, y las características inversas por parte del lenguaje reductor nos obligan a traducirlo a expresiones más maleables. De este modo, Searle acierta en lo siguiente: no hay puestas de Sol, y probablemente tampoco haya entidades de noventa minutos llamados "partidos de fútbol"… en este sentido *no existen* las entidades reducidas al lenguaje de las ciencias naturales. Pero la eliminación propuesta por cualquier programa eliminativista[16] no es de esta clase… es una eliminación teórica

[16] Podríamos sostener que el abandono de la alquimia por parte de la química moderna fue un movimiento "eliminativista" como sugiere Churchland (1989)

directa en función de la "incompetencia" de la teoría en cuestión -en el caso del ME la psicología folk-. En este sentido las puestas de Sol y los partidos de fútbol sí existen (serían términos reductibles a una teoría, es decir, términos aun empleables), aun cuando sean cosas bastante diferentes de las que pensábamos en un principio; pero a su vez el flogisto, los espíritus animales, las actitudes proposicionales, etc, no son reductibles a ninguna teoría, por lo que no existen simple y llanamente. Un uso de los términos[17] relacionado con la reducción interteórica afirmará que no existen los "partidos de fútbol" puesto que resultan ser en verdad cosas completamente distintas de las que pensábamos –digamos, movimientos de moléculas con cierto grado de organización muy compleja, que involucran capitalmente procesos realizados por sistemas nerviosos muy sofisticados- y otro uso de los términos relacionado con la eliminación directa de teorías afirmará (y en otro sentido) que no existen los "pegasos", es decir, que no hay ninguna teoría disponible y verosímil para hacer referencia a estas entidades.

"Tesis: con todo, es posible hacer una lista de afirmaciones de la psicología popular y ver que muchas de ellas son dudosas. Respuesta: si se mira a las listas que se dan efectivamente, hay algo sospechoso en ellas. Si tuviera que hacer una lista de algunas proposiciones de la PP, incluiría cosas como las siguientes: 1. Las creencias pueden ser, en general, verdaderas o falsas. 2. Algunas veces las personas tienen hambre, y cuando tienen hambre desean a menudo comer algo. 3. Los dolores no son, a menudo, placenteros. Por esta razón, la gente trata a menudo de evitarlos. Es difícil imaginar qué género de evidencia empírica podría refutar estas proposiciones. La razón es que, de acuerdo con una interpretación natural, no son hipótesis empíricas, o no son sólo hipótesis empíricas. Son más semejantes a los principios constitutivos de los fenómenos en cuestión. La proposición 1, por ejemplo, es más semejante a la "hipótesis" de que un touchdown en el fútbol norteamericano vale seis puntos. Si a uno se le dice que un estudio científico ha mostrado que los touchdowns valen efectivamente sólo 5,9999999999 puntos, nos daríamos cuenta enseguida que hay alguien aquí que está seriamente confundido. Es parte de nuestra definición ordinaria de touchdown el que vale seis puntos. Podemos cambiar la definición pero no podemos descubrir un hecho diferente. De forma similar, es parte de la definición de "creencia" el que las creencias son candidatos para la verdad o la falsedad. No podríamos "descubrir" que las creencias no son susceptibles de ser verdaderas o falsas. (…) Puesto que son constitutivas y no empíricas (la proposiciones centrales de la PP), la única manera de mostrar que son falsas sería mostrar que no tienen rango alguno de aplicación. Por ejemplo, los principios constitutivos de la brujería no se aplican a nada porque no hay brujas. Pero no se puede mostrar que los deseos conscientes y los dolores no existen de la manera en que se puede mostrar que las brujas no existen porque aquellas son experiencias conscientes y no se puede hacer la distinción usual entre apariencia y realidad para las experiencias conscientes." (Searle. 1992. Pag 75-76)

Hay en esta réplica una comparación desafortunada y dos puntos centrales que no comparto. Desconozco los espantapájaros que los materialistas de cualquier índole puedan haber montado con respecto a las sentencias básicas de la psicología popular para

[17] Términos tales como "existir" y "eliminar".

emplearlas de puchingball[18], pero ciertamente no es posible comparar tales sentencias con definiciones normativas, sean de la índole que sean (reglas de juegos, definiciones lógicas, etc.). Las reglas de los juegos o las reglas de un sistema formal, aunque puedan ser ambiguas (como las de algunos juegos) y nada triviales ni caprichosas como aspiran a ser los enunciados de la psicología folk, estas reglas digo son estrictamente normativas porque son modelos fabricados *a priori* por el hombre y no descubiertos empíricamente. Las leyes empíricas no funcionan del mismo modo, si bien las empleamos *como si* fueran normativas y estrictas en la práctica, son hipótesis generales frecuentemente muy regulares, pero que juegan un rol dentro de alguna teoría empírica y son susceptibles, en mayor o menor medida, de comparecer al juicio de la experiencia. Los enunciados de la psicología folk son en cualquier caso enunciados empíricos puesto que nadie los ha postulado para modelar ninguna actividad determinada, lo cual a lo sumo las dejaría en el estatus de "leyes generales" o "hipótesis universales", pero no como "enunciados constitutivos". Pero esto reafirmaría la posición de que la psicología popular es una teoría empírica más.

Claro que esto no se aceptará tan fácilmente. Se supone que hay una diferencia entre los hechos empíricos constatados en tercera persona y los hechos que, si bien empíricos en algún sentido, se instancian en primera persona. De los segundos poseemos un conocimiento privilegiado y directo, nos están dados, son *nuestros* dolores, miedos o creencias y podemos corroborar directamente (aunque no públicamente) que nos duelen, aterran o convencen. Pues bien, la cosa queda tal y como estaba antes: para refutar la condición teórica de la psicología folk nos vemos forzados a emplear categorías mentales privadas, las cuales constituían el hecho que pretendía discutirse por parte de ME.

Ahora bien, Searle asevera que sólo la inaplicabilidad de estos enunciados podrían hacerlos falsos, pero todos sus ejemplos implican sensaciones primarias o actitudes proposicionales supuestamente conocidas de manera infalible con lo que se sentencia desde el comienzo una aplicabilidad garantizada –siempre y cuando no se ponga en duda el valor epistemológico de tales estados como hace el supuesto C)-.

Lo que en cada contra-argumento Searle da por sentado es que tanto B) como C) son en algún sentido falsos. En primer lugar B) pues se intenta repetidamente establecer un criterio de referencia fuera de todo marco teórico delimitado (haciendo constante uso de las entidades presupuestas por la psicología folk y otras presuntamente irreductibles) y casi por los mismos medios se esquiva constantemente la discusión con C) presuponiendo constantemente cierto acceso privilegiado a determinados conocimientos (sensaciones primarias, actitudes proposicionales, etc.).

Si estoy en lo cierto los dos supuestos de los cuales no se hace ni mención son estructuralmente más fuertes que A) y deben ser socavados anticipadamente para poder lidiar con A).

En cualquier caso la mayoría de los argumentos de Searle pretenden traficarnos de contrabando un lenguaje privilegiado basado en la ostención interna y, si no me equivoco, el punto 1 de este capítulo estableció la incorrección de esta forma de argumentación en el debate.

3) Una versión más sutil de la misma clase de argumentos expuestos en 2) es presentado por Feyerabend (1963) y sostiene que no podemos saber de qué estamos hablando si no conectamos los términos de una teoría con un lenguaje previo y familiar. Así el reemplazo radical pretendido por el ME es en verdad algo que nos deja completamente mudos. Los resultados observacionales de la neurociencia, por ejemplo, no consiguen

[18] Searle acusa a Churchland de llevar a cabo esta práctica.

comprenderse como datos relevantes dentro del marco de la discusión mente-cuerpo si hemos eliminado todo contacto con nuestro lenguaje mentalista de base. Estos lenguajes deben ser mantenidos en algún sentido si se desea conectar significativamente estas teorías. La eliminación está prohibida.

Esta forma "sutil" del argumento también toma por sentado que C) y B) son de hecho falsas. Feyerabend rechaza esta crítica apelando (tácitamente) a los dos supuestos en cuestión:

"Esta objeción asume que los términos de un punto de vista general y de un lenguaje correspondiente pueden obtener significado sólo siendo referidos (relacionados) con los términos de otro punto de vista familiar y conocido por todos. Ahora si ese es realmente el caso, ¿Cómo hizo el punto de vista último y el lenguaje último en cada caso para obtener su familiaridad? Y si pudo obtener su familiaridad sin ayuda "de otro lado", como obviamente lo hizo, entonces no hay razón para asumir que otro punto de vista no pueda hacerlo igualmente bien." (Feyerabend, 1963, pag 1)

Aquí Feyerabend simplemente apela a C) para rechazar este argumento, e inmediatamente apela a B) en una nueva réplica a la objeción:

"Más aun, los resultados observacionales siempre deben ser formulados con respecto a un marco (background) teórico (con respecto a ciertos juegos de lenguaje para usar términos a la moda). No hay razones de por qué la fisiología no sería capaz por sí misma de formarse un marco". (Feyerabend, 1963, pag 1 y 2)

Parece ser que ninguna forma directa de atacar A) puede evitar vérselas con sus guardaespaldas B) y C). Pero más en general toda tesis que implique la incorregibilidad de cualquier lenguaje debe vérselas al menos con C), por lo que podemos descartar a buenas y primeras cualquier forma de objeción contra el ME basada en la incorregibilidad (en cualquier nivel) que no se haga cargo expresamente de C).

4) Por otra parte hay algunas buenas objeciones contra la aceptación de un criterio de referencia como el establecido por B). Es posible al menos señalar tres tesis alternativas relevantes en filosofía de la mente que pretenden competir con la tesis reticular: 1) La de la definición por ostención interna. 2) El conductismo 3) La teoría causal de la referencia. Pero 1) y 2) son en verdad malos candidatos. 1) En primer lugar lleva inexorablemente a un solipsismo semántico y además no se hace cargo en absoluto de los contra argumentos que en apoyo de B) puedan venir desde su aliado C) – se da de hecho por sentada la falsedad de C) como veíamos en el punto 1) de este capítulo-. Y 2), pese a ser hasta en cierto sentido metodológicamente compatible con el ME, presenta problemas archiconocidos, tales como la casi imposibilidad de acotar estrictamente un *definiens* para cualquier término mental por expresarlos en términos disposicionales (aunque conductuales).

La teoría de la referencia como rol causal parece, por otro lado, ser un buen candidato. Por una parte –sea que se asuma un "bautismo inicial" al estilo Kripke (1972) o alguna otra forma de determinar el significado del término (Putnam (1975) tal vez tiene algunos detalles que agregar al respecto)- esta teoría nos permite conservar nuestros usos lingüísticos usuales y por otra parte deja abierta a la investigación empírica el significado exacto de estos. Puede establecerse desde la teoría causal de la referencia, en conclusión, una buena crítica a A) puesto que no estamos obligados a considerar todo nuestro vocabulario folk como perteneciente a una teoría siempre tan amenazada como las

de las ciencias empíricas (pueden haber términos más resistentes y podemos seguir empleando significativamente los mismos términos pese al avance científico –podemos ser relativamente "conservadores" en semántica-), y además no se cae, a primera vista, en una tesis de la incorregibilidad (corregimos constantemente nuestro vocabulario de sentido común, por ejemplo, ampliamos nuestro conocimiento acerca del "agua", del "calor", etc.). Por otro lado, al presentarse como un competidor directo de la teoría del rol inferencial de la referencia (o teoría reticular) no necesita hacerse cargo expresamente de B) más que siendo una buena teoría rival.

Esta objeción ha sido tratada por Pérez (2006) en los siguientes términos:

"Ahora bien, cuando descubrimos que el agua está compuesta químicamente por dos moléculas de hidrógeno y una de oxígeno, esto es, cuando descubrimos que agua = H2O, no eliminamos el agua, ni decimos que "realmente no hay agua, ya que sólo hay H2O", ni remplazamos nuestro hablar cotidiano acerca del agua por un hablar químico en términos de H2O, ni se nos ocurre que el concepto de "agua" sea eliminable. ¿Por qué? Porque la química no puede ni busca reemplazar nuestro concepto ordinario de agua por el de H2O, ya que al establecerse la identidad entre agua y H2O lo que estamos aceptando es la realidad del agua. Es a partir de nuestro concepto ordinario de agua, el que nos permitía discriminar cierta sustancia en el mundo sin tener demasiado sustento teórico para hacerlo, que fue posible que la ciencia química se desarrollara y nos permitiera entender mejor por qué estábamos discriminando ciertas porciones de sustancia en el mundo como porciones de agua. Los conceptos ordinarios así funcionan como motor, condición de posibilidad de la actividad científica, proponiendo taxonomizaciones de fenómenos ordinarios que la ciencia busca fundamentar en razones teóricas explicativas/ más profundas. Sin taxonomización ordinaria no hay ciencia. ¿Podría la química haberse desarrollado si no hubiéramos poseído el concepto ordinario de agua y el de ginebra? Si la química madura no logra explicar qué diferencia hay entre el agua y la ginebra, parece que hay un cambio de tema, que la química no está hablando de los macrofenómenos que discriminamos gracias a nuestro lenguaje ordinario". (Pérez, 2006, pag 17)

Este argumento tal vez pueda vérselas con el supuesto A) y en parte encarar a B), pero no vemos claramente si realmente a superado los problemas planteados por C)… ¿Es o no es una tesis acerca de la incorregibilidad de nuestro lenguaje ordinario?... por una parte afirma no serlo, pero por otra descarta completamente la eliminación de términos del lenguaje ordinario que han sido tratados como términos de de clase natural. La estrategia a seguir es, según esto, incorporar la mayor cantidad de términos mentales como términos de clase natural, pero aun suponiendo que se cuente con buenos programas en filosofía de la mente para llevar a cabo este proceso (tal vez el funcionalismo sea el mejor candidato), ¿podemos tildar de "corregible" una teoría que prohíbe drásticamente la eliminación teórica?, ¿por qué, en el fondo, suponen vedada esta posibilidad?, ¿por qué querríamos, en principio, suponer como designadores rígidos términos que hemos adquirido en el uso ordinario del lenguaje y no otros?, ¿qué privilegio epistemológico tienen estos términos? Según la cita transcrita más arriba porque los términos del lenguaje ordinarios modelan la actividad científica, porque es desde nuestra ordinaria discriminación entre agua y ginebra que la química comienza a investigar qué sean técnicamente estas cosas. Pero el hecho de que la química, siguiendo el ejemplo, comience por estas discriminaciones de sentido

común no implica que *deba* hacerlo. Si no presentamos algún argumento adicional que se oponga a C) no vemos por qué la química deba suscribir a algún compromiso semántico (mucho menos ontológico) con los términos ordinarios. Si la química no discriminara, en última instancia, entre el agua y la ginebra quizá no estaría "cambiando el tema", sino que estaría informándonos que nuestras "borracheras" no son un buen mecanismo de discriminación y taxonomización del mundo. ¿Existe algún argumento independiente que postule la completa corrección de nuestras discriminaciones de sentido común?... parece que un tal argumento apelaría simplemente a los viejos privilegios semánticos concedidos, por ejemplo, a nuestras sensaciones primarias (a nuestra "borrachera"). ¿Cómo puede distinguirse la objeción proveniente de la teoría causal de la referencia como una objeción distinta de las otras en cuanto a su incorregibilidad?... esta objeción parece no distinguir seriamente el nivel heurístico que nuestro lenguaje ordinario tiene con relación a las teorías científicas y el nivel justificativo. Nadie duda de que el lenguaje ordinario ha resultado útil en la vida práctica y modelador en la vida teórica, simplemente se le agradece su servicio y se revisa su justificación y su estatus dentro de las nuevas y mejores teorías (aunque él haya ayudado a modelarlas). El sentido común ha creado, posiblemente, a sus propios verdugos.

Además la formulación más generalizada de la teoría del rol causal de la referencia Kripke (1972)) asume que ciertos estados fenoménicos tienen plena validez para ser identificados como *designadores rígidos*. Pero la aceptación de estados fenoménicos como "mi sensación de dolor" como buenos candidatos para emplearse como *designadores rígidos* reivindica lo que Rorty (1979) quizás llamaría "un postulado mentalista históricamente heredado" –es decir P-. Si la teoría del rol causal quiere enfrentar en general las tesis eliminativistas debe cuidarse de no reintroducir *a priori* puntos de vista como P para no ser acusada de petición de principios en un sentido similar al que lo fue la teoría de la referencia por obtención interna.

Habiendo analizado estas objeciones podemos concluir que, si bien A) es el supuesto que juega el rol principal en la exposición del argumento eliminativista, C) es el transfondo de toda la discusión respecto de esta tesis en filosofía de la mente.

Pero discutir directamente C) sería entrar de lleno en una candente discusión epistemológica y metafísica que sobrepasa gruesamente los límites de este ensayo.

Capítulo 3. El argumento trascendental contra el materialismo eliminativo y la reaparición de un viejo problema

Existe otra línea argumentativa paralela en contra del eliminativismo. Esta clase de argumentos no intenta discutir con ninguno de los supuestos enumerados más arriba, sino que considera la propia formulación del ME como una tesis paradójica. Esta clase de contra argumentos pueden llamarse "argumentos trascendentales". Daremos una formulación provisoria y poco fina del argumento aunque puedan existir diversas variantes más sutiles y convincentes. El problema va como sigue:

"La formulación de una tesis supone de por sí cierto compromiso con la intencionalidad: supone ser una creencia o alguna actitud proposicional, justificada generalmente y expresada con intenciones de comunicación, y en general con condiciones lingüísticas intensionales. Ahora bien, si la tesis eliminativista es verdadera en algún sentido, tendrá todas estas características, pero la tesis eliminativista niega la existencia de todas estas cosas, por lo que si es verdadera debe sostener que no tiene ninguna de las condiciones que hacen verdadera una tesis. Esto la convierte en un argumento asignificativo, pues la propia tesis niega la existencia de las condiciones necesarias que hacen significativo a un enunciado."

Si esto es verdad, la hipótesis eliminativista sufre un problema paralelo al de otras tesis ampliamente desacreditadas tales como: "ningún enunciado es significativo", "ningún enunciado es verdadero", etc.

La falla de este argumento es una petición de principio análoga a la discutida en el punto 1) del capítulo anterior. Suponer que el criterio de significado está constituido sin más por caracteres mentales es sacar al ME del juego de antemano. El ME se verá obligado a sostener que no son tales los criterios de significado para una expresión, y eso es todo. Sin embargo el lenguaje significativo no intencional que el ME está obligado a presentar, al menos potencialmente, adolece de la misma debilidad discutida el punto 2 del capítulo 2 en la cuarta tesis y réplica de Searle, esto es, que es un postulado "futurista" basado en las esperanzas materialistas referentes al desarrollo de la neurociencia. El objetor podría apretar al eliminativista diciéndole cosas como: "bueno, si este lenguaje intencional mío te causa tantas paradojas y problemas para expresar tu tesis, porque no empleas tu poderoso y más sofisticado lenguaje basado en la neurociencia… ¡Ah, perdón!… lo había olvidado… porque no posees tal lenguaje"… a lo que el materialista eliminativo podría responder: "es cierto, nada quisiera más que emplear un lenguaje basado en una neurociencia completa… pero, como bien señalaste, no lo poseo… ¿Pero sabes tú por qué no está disponible este lenguaje?... ¡Porque tu multitud de programas mentalistas en filosofía de la mente y tu defensa infantil de la psicología folk despilfarran recursos, desvían esfuerzos investigativos y retardan los avances lógicos y técnicos necesarios para su realización!", presumiblemente desde este punto la discusión tenderá a resolverse mediante la portentosidad de las cuerdas vocales más que por razonables argumentos.

Sea como sea, en esta versión poco sutil del argumento trascendental el ME parece tener la razón. Suponer que los criterios de significado constituidos desde categorías mentales (categorías que pretendían estar en discusión) son de hecho *los* criterios inamoviblemente válidos para cualquier forma de lenguaje, es reivindicar *a priori* las tesis mentalistas y, por lo tanto, sabotear antes de la carrera el carro eliminativista. Churchland da un buen ejemplo de esta petición de principios:

"Considérese la teoría medieval que postula que estar biológicamente vivo significa estar animado por un espíritu vital inmaterial. Y considérese la siguiente respuesta que se le da a alguien que ha expresado que no cree en esa teoría: Mi docto amigo ha afirmado que no existe nada que sea un espíritu vital. Pero esta afirmación es incoherente. Porque si fuera verdad, entonces mi amigo no tiene espíritu vital y, por lo tanto, debe estar muerto. Pero si estuviera muerto, entonces su afirmación sólo es una retahíla de ruidos, desprovista de sentido o de verdad. Evidentemente, ¡el supuesto de que el antivitalismo está en lo cierto presupone que no puede estarlo! Q.E.D." (Churchland, 1988, pag 83-84)

Esta caracterización del argumento quizá es demasiado superficial. El argumento trascendental es capaz de apoyarse en diversas teorías e interpretaciones ontológicas, epistemológicas y referentes a la filosofía del lenguaje que podrían, eventualmente, hacerlo más plausible[19].

Daré una formulación más plausible del argumento trascendental apelando a la tesis del "realismo interno" de Putnam (1987), pese a que esta doctrina no sólo ataca al ME sino que a casi todas las versiones materialistas en filosofía de la mente. Elijo en particular esta exposición de Putnam por ser cercana y compatible con las ideas antes mencionadas de Searle además de presentar una defensa a mi juicio fuerte de la intencionalidad[20]. Demos entonces un vistazo a la defensa radical de la intencionalidad que Putnam nos brinda:

Hagamos ante todo, y siguiendo al autor, un poco de historia. Supone Putnam que ante los desvaríos filosóficos[21] del siglo recién pasado y el anterior –y del nuestro también, por qué no- los filósofos de buen sentido decidieron de buena fe adherirse al bando de los realistas y creyeron además haber hecho con esto gran justicia a nuestro sentido común. Pero al cabo de un rato, y cuando algunos de los realistas menos informados se sintieron más cómodos y distendidos, y comenzaron a soltar en la conversación frases tales como "esta silla es bastante sólida" o bien "¡Bello color rojo el de esta mesa!", empezaron a resonar, entre dientes, risas por parte de los realistas más eruditos[22]. Los realistas ingenuos preguntaban que a qué venía tanta gracia y secreto. Finalmente alguno de los Realistas les confesó lo que sigue: "Señores, es hora de decirles la verdad. Bueno, sé que han decidido apoyarnos por parecerles insólitas las doctrinas tendientes a negar que verdaderamente *hay* sillas y mesas frente a nuestros ojos… ¡pero tampoco es que sus sillas y mesas, tales y como suponían que eran, vayan a quedar intactas!... lo cierto es que no están coloreadas, ni son sólidas… ni siquiera son algo así como *una* cosa, sino que una reunión multitudinaria de invisibles *Dinge an sich*.". Ante la cara de espanto de los ingenuos, el Realista les dio una palmadita en el hombro y concluyó: "¡Tranquilo muchachos!... pueden seguir hablando

[19] Churchland, fiel a su estilo, nos ha presentado el argumento con su característica sutileza de motocierra oxidada.

[20] Está fuera de los límites de este trabajo discutir las implicancias ontológicas del realismo interno de Putnam.

[21] Léase solipsismo, idealismo, fenomenalismo, descontructivismo, etc.

[22] Siguiendo a Putnam pasamos a llamarlos Realistas con "R" mayúscula, esto es, el "realista metafísico" o el "realista científico".

de mesas y sillas en la cháchara dominical, o cuando discutan de fútbol… ¡No íbamos a abandonar el cómodo lenguaje de nuestra "imagen manifiesta" por tan poca cosa!... simplemente – les susurró a modo de consejo- hagan desaparecer esta clase de términos de sus discursos teóricamente relevantes.".

Ciertamente el realista ingenuo se siente estafado. Buscando escapar de un descalabro viene a caer en lo que parece otro… "¿Qué visión de mundo es esta que nos obliga a renunciar a nuestros objetos más usuales y a afrentar tan grandemente a nuestro sentido común?" se pregunta el realista ingenuo.

Esta "visión de mundo" del Realista, que Putnam pasa a llamar objetivista –a la visión-, consiste en suponer que existen por una parte propiedades intrínsecas de los objetos, y que además hay propiedades (como la "rojez" o la "solidez") que son secundarias y aparentes. Son en definitiva disposiciones para afectarnos, para producir en nosotros determinados datos de los sentidos, o bien determinados estados neuronales (en la variante materialista). En resumen esta "apariencia", estas propiedades que *parecieran* estar en las cosas intrínsecamente sin estarlo *realmente*, es lo que los objetivistas pasan a llamar "proyección", suponiendo que nosotros hemos "proyectado" estos estados a las cosas *como si fueran* propiedades de ellas en sí mismas.

El garrafal problema que Putnam supone haber encontrado en la visión objetivista del mundo se refiere al inconsistente tratamiento que se ha dado a las propiedades disposicionales. El realista científico debe en alguna medida hacerse cargo de los objetos que el sentido común nos presenta cotidianamente (sillas, mesas, etc.). Pero al explicarnos *qué cosas son en sí* estos objetos, acaba hablándonos, extraña y paradójicamente, de otras cosas que no nos parecen para nada aquellas de las que se pretendía hablar. Las sillas y mesas nos son trocadas por átomos, electrones, ondas, etc., objetos con los cuales –y no es que sintamos antipatía por ellos- no convivimos, *ni podemos convivir*[23], usualmente. Sin duda quien esperaba conservar el mobiliario de su casa a nivel teórico estará –llegado a este punto- profundamente decepcionado del Realismo metafísico.

Pero Putnam lo alentará a no perder la esperanza. De hecho –sostiene el autor- el Realista científico no ha podido reducir satisfactoriamente las propiedades disposicionales a la física de primer orden (a propiedades no-disposicionales)… "tener determinado color", por ejemplo, no es una propiedad equivalente sin más a alguna otra perteneciente a algún lenguaje nomológicamente estable:

"¿Qué pasa con las propiedades disposicionales? A menudo se sostiene que el color es simplemente una función de la reflexión de la luz, esto es, de la disposición de un objeto (o de la superficie de un objeto) para absorber selectivamente ciertas longitudes de onda de luz incidente y reflejar otras. Pero, verdaderamente, esto no tiene mucho que ver con la realidad de los colores. Las investigaciones recientes no sólo han mostrado que esta consideración es demasiado simple (porque cambios de reflexión en los bordes pasan a jugar un importante papel en la determinación de los colores que vemos), sino que la reflexión, en sí misma, no tiene una explicación física uniforme. De hecho, puede haber un número infinito de condiciones físicas diferentes que podrían producir la disposición de reflejar (o emitir) luz roja y absorber luz de otras longitudes de onda. Una propiedad disposicional cuya "explicación" no disposicional subyacente sea tan poco uniforme simplemente no es susceptible

[23] En sentido epistémico.

de ser representada como una función matemática de las variables dinámicas. Y éstas –las variables dinámicas- son los parámetros que esta forma de pensar considera como las "características" de los "objetos" externos." (Putnam.1987. Pag. 44-45)

Pero, ¿Acaso debemos detenernos sólo en colores, olores, o cosas de ese estilo (del "estilo sensualista")?... ¡Nada de eso!... todas las descripciones que el objetivista supone *reales* deberían ser propiedades estrictamente no disposicionales, dicho en el vocabulario de Putnam, deberían ser *disposiciones estrictas*.

Antes de analizar esto precisemos aun más la terminología del autor: 1) *Disposición estricta* es la disposición que algo tiene para hacer algo, sin importar lo que ocurra (sin ser relevante condición concomitante alguna). 2) *Disposición ceteris paribus* es la disposición que algo tiene para hacer algo *en condiciones normales*.

Como ejemplo de una disposición estricta podemos dar la disposición de los cuerpos con una masa en reposo no igual a cero de viajar a velocidades por debajo de la velocidad de la luz. Encontrar un cuerpo que no cumpla estrictamente esta disposición (cuya masa en reposo no sea igual a cero y que, sin embargo, viaje a la velocidad de la luz) es imposible, es, por así decirlo, una contradicción física.

Pero la mayoría de las propiedades (inclusive empleadas en ciencias) no comparten esta gracia; requieren, por el contrario, cierta *acotación teórica* del tipo "en condiciones normales, en condiciones favorables, etc." para encontrar algún significado estable y ser maleables a nivel técnico y usual. Son en resumen, disposiciones *ceteris paribus*. Veamos el ejemplo que Putnam nos da:

"¿Qué ocurre con la disposición del azúcar para disolverse en agua? Esta no es una disposición estricta, ya que el azúcar que se coloca en agua ya saturada con azúcar (o, incluso, con otros productos químicos apropiados) no se disolverá. ¿Es, entonces, una disposición estricta la disposición del azúcar para disolverse en agua químicamente pura? He aquí otra disposición no estricta: el primer contraejemplo que mencionaré proviene de la termodinámica. Supongamos que arrojo un terrón de azúcar en agua y que el terrón de azúcar se disuelve. Consideremos azúcar que está en agua, pero de tal forma que, aunque la situación sea idéntica a la situación que acabo de imaginar (el azúcar está disuelto en el agua) con respecto a la posición de cada partícula, todos los vectores de momento tienen una orientación exactamente opuesta respecto de los que ahora tenemos. Se trata de un ejemplo famoso: lo que ocurre en el ejemplo es que el azúcar, en lugar de permanecer disuelto, simplemente forma un terrón de azúcar que, espontáneamente, ¡salta fuera del agua! Puesto que a todo estado normal (todo estado en el cual el azúcar se disuelve) le corresponde un estado en el cual "se des-disuelve", vemos que hay infinitamente muchas condiciones físicamente posibles en las cuales el azúcar "se des-disuelve" en vez de permanecer disuelto. Por supuesto, todos son estados en los que la entropía disminuye; pero eso no es imposible, ¡sólo extremadamente improbable! ¿Diremos, entonces, que el azúcar tiene una disposición estricta a disolverse a menos que se encuentre en una condición en la cual tenga lugar una disminución de la entropía? No, porque si el azúcar se coloca en agua y se produce una congelación repentina, el azúcar no se disolverá si la congelación

tiene lugar lo suficientemente rápido… De hecho lo que podemos decir es que bajo condiciones normales el azúcar se disolverá si es colocado en agua. Y no hay ninguna razón para pensar que todas las diferentes condiciones no normales (incluyendo extraños estados cuánticos, extrañas fluctuaciones locales en el espacio-tiempo, etc.), bajo las cuales el azúcar no se disolvería si fuera colocado en agua, puedan recogerse en una fórmula cerrada del lenguaje de la física fundamental." (Putnam. 1987. Pag 51-52)

Desde aquí el proyecto objetivista, según Putnam, comienza a parecer una empresa temeraria, pues aunque el ejemplo suena trivial y hasta simpático al primer vistazo, resulta entrañar la misma dificultad expuesta para las propiedades que el Realista científico aseguraba, eran de segundo orden:

"¡Éste es exactamente el problema que hemos observado antes en conexión con la rojez y la solidez! Si son las propiedades "intrínsecas" de las cosas "externas" las que podemos representar mediante fórmulas en el lenguaje de la física fundamental, por medio de "adecuadas funciones de las variables dinámicas", entonces la solubilidad tampoco es una propiedad "intrínseca" de ninguna cosa externa. Y, de manera similar, tampoco lo es ninguna disposición "en igualdad de circunstancias"." (Putnam.1987. Pag 52)

¿Qué hará ahora el Realista?, ¿Rechazará todas las nociones disposicionales incapaces de reducirse al lenguaje de la física?... entonces constatará que el mundo de *lo que hay realmente* – para el Realista –, comienza poco a poco a despoblarse, al mismo tiempo en que el mundo de *lo que pareciera que hay y no hay realmente* sufre una inmigración abrumadora.

Quizás deba entonces el Realista reconocer la irreductibilidad de ciertas propiedades ya no tan "triviales" como "ser simpático" u "oler mal", es decir, deberá reconocer que aquellas propiedades identificables sólo *en condiciones normales* no son reductibles, para luego sumarlas a la lista de las "proyecciones" que el hombre arroja sobre la "realidad en sí".

"¿Y qué?" – Proseguirá el Realista – "¿qué importa cuantas propiedades nos queden en la realidad y cuantas sean en verdad proyecciones?".

Lo que no queda claro en medio de todo este intenso tránsito de propiedades de un lado a otro, es precisamente qué es eso de *proyectar propiedades a las cosas* que tan irrelevante y trivialmente es empleado en el discurso Realista (incluso siendo que todas sus propiedades no-estrictas son de este género). ¿Qué es una *proyección*? Ya se nos había dicho algo, es simplemente pensar que algo tiene una propiedad que no tiene en realidad, sin darnos cuenta, claro está, de que pensamos así (y sin estar, presuntamente, en lo correcto). Es en definitiva *pensar algo de algo*, ni más ni menos que tener ciertas actitudes proposicionales o, si se prefiere, tener ciertas actitudes *intencionales*. Es ni más ni menos que de la intencionalidad de la que la visión objetivista tendrá que hacerse cargo ahora para brindarnos, por fin, una visión acabada de su proyecto teórico (proyecciones incluidas).

Pero ¿Y estas proyecciones, cómo las explicará? Y es que si no logra caracterizarlas en términos no disposicionales caerá en el increíble absurdo de suponerlas a ellas mismas proyecciones, tropezando en un círculo vicioso francamente insufrible.

Deberá dar entonces, el objetivista, un programa reduccionista del fenómeno de la intencionalidad. Pero es en este punto donde Putnam asesta el golpe de gracia:

"El objetivismo moderno se ha convertido simplemente en materialismo. Y el problema central para el materialismo consiste en "explicar la aparición de la mente". Pero, si "explicar la aparición de la mente" significa resolver el problema de Brentano, esto es, decir en términos reductivos qué es "pensar que hay muchos gatos en el vecindario" y qué es "recordar dónde está París", etc., ¿por qué deberíamos pensar ahora que eso es posible? Si se ha demostrado que reducir el color, la solidez o la solubilidad a la física fundamental es imposible, ¿por qué debería ser más fácil de resolver este programa reduccionista mucho más ambicioso?" (Putnam.1987. Pag. 56)

Hay sólo dos caminos para tratar el problema de la intencionalidad: 1) Postulamos la sustancialidad de la mente como sustrato de los fenómenos intencionales. 2) Suponemos que los fenómenos mentales se explican estrictamente por fenómenos físicos, como Hobbes o Diderot lo creyeron.

No tenemos ninguna perspectiva, ni la más remota y disparatada esperanza en la actualidad, de desarrollar un programa filosófico basado en 1). Si el objetivismo se inclina por la opción 2), encontrará que también ha seleccionado un programa filosófico intensamente complicado. El objetivista quiere, según esto último, volverse un simple materialista y explicar el surgimiento de la mente, probablemente, mediante alguna reducción fisicalista. Pero si explicar la aparición de la mente es explicar los fenómenos intencionales tales y como están expuestos en la última cita de Putnam, es decir, explicar qué es "entender que Pablito quiere un chocolate" por ejemplo, si esto es lo que el proyecto objetivista pretende reducir punto por punto a un lenguaje adecuado (tendrá que ser, en última instancia, el de la física). ¿Qué pruebas aporta el objetivista para sostener que esta reducción es siquiera posible? Dado que ha fracasado al intentar reducir propiedades como "rojez" y "solidez"; incluso ha fracasado al tratar de transformar adecuadamente al lenguaje de la física fundamental propiedades como la "solubilidad"; si parece, finalmente y para colmo, imposible reducir a este lenguaje ninguna disposición no-estricta. ¿Qué nos hace pensar que el objetivista, tentado ahora por el materialismo, podrá hacerse cargo de un programa reduccionista mucho más ambicioso?

Quizás el funcionalismo propuesto a finales de los cincuenta por el propio Putnam pueda dar algún auxilio. Efectivamente este programa en filosofía de la mente parece más alentador puesto que supone que los estados intencionales de los seres pensantes son cierta organización causal para realizar una función, y no un estado físico estricto (mucho menos una sustancia aparte). Con esta tesis disponible, podemos suponer, cuando menos, que una reducción fisicalista *estricta* no es necesaria para dar cuenta de la aparición de la mente, dado que los mismos estados mentales podrían tener diversos soportes materiales. Pero tal y como las distintas bases de la "rojez" de una estrella y de una manzana no permitían su reducción al lenguaje fundamental de la física (ni siquiera como funciones de propiedades no-disposicionales), así mismo la "aparición de la mente" – fenómeno indeciblemente más complejo que la "rojez" –, a todas luces, no tiene mejores perspectivas de ser explicado y reducido al lenguaje de la física básica.

Incluso la versión computacional no permite una reducción estricta. No parece haber condiciones necesarias y suficientes en términos computacionales o computacionales-físicos para dar satisfactoria y homogénea reducción de los estados intencionales. Una condición de esta clase (suficiente y necesaria) debería ser:

"Infinitamente larga, y no construida de acuerdo con ninguna regla efectiva, ni siquiera conforme a una prescripción no efectiva que pudiéramos proponer sin usar los propios términos que han de reducirse" (Putnam. 1987. Pag 58)

> Pero vistas las dificultades de la opción 1) y 2) puede aun tomar el objetivista un tercer camino: Suponer que la intencionalidad misma es una proyección… Pero ¿tiene realmente algún sentido esta tesis?... si la noción de *proyección* presupone la de *intencionalidad*, y deseamos supuestamente ahora definir *intencionalidad* por *proyección*… ¿a qué clase de engorroso circulo estamos suscribiendo para explicar la visión objetivista del mundo?

En este engorroso círculo y en esta opción paradójica es donde Putnam sitúa al ME:

"Por extraño que parezca, y a pesar de su carácter absurdo, la idea de que el pensamiento es una mera proyección está siendo defendida por varios filósofos en los Estados Unidos y en Inglaterra. La fuerza de la tradición "objetivista" es tan intensa que algunos filósofos abandonarán las más profundas intuiciones que tenemos sobre nosotros-mismos-en-el-mundo antes que preguntarse (como hicieron Husserl y Wittgenstein) si la imagen completa no es un error. Por ello, en las últimas décadas del siglo xx nos encontramos con filósofos inteligentes que proponen que la intencionalidad misma es algo que nosotros proyectamos adoptando una "perspectiva" hacia ciertas partes del mundo[24] (¡como si "adoptar una perspectiva" no fuera, a su vez, una noción intencional!), filósofos inteligentes que pretenden que nadie tiene realmente actitudes proposicionales (creencias y deseos), que "creencia" y "deseo" son únicamente nociones de una falsa teoría llamada "psicología natural"[25], y filósofos inteligentes que defienden que no existe una propiedad como la "verdad" ni una relación como la referencia, que "es verdad" no es más que una locución que usamos para "elevar el nivel del lenguaje". Uno de ellos –Richard Rorty[26], un pensador de gran profundidad- observa que está comprometido a rechazar las intuiciones que subyacen a cualquier clase de realismo (y no sólo al realismo metafísico), aunque la mayoría de estos pensadores escriben como si estuvieran salvando el realismo (en su versión materialista), ¡abandonando la intencionalidad! ¡Es como si no hubiera ningún problema en decir "yo no niego que hay un mundo externo; lo que niego es que pensemos sobre él"! Créanlo, ésta es también la forma en que escribía Foucault. ¡La línea entre el relativismo à la française y la filosofía analítica parece ser más tenue de lo que los filósofos angloparlantes piensan! Algo bastante divertido es que la sobrecubierta de uno de los últimos ataques a la "psicología natural" lleva un anuncio entusiasta en el cual un crítico explica la importancia del libro, que está dentro de la sobrecubierta, diciendo que ¡la mayoría de la gente cree que hay cosas tales como las creencias[27]!" (Putnam. 1987. Pag 58-59)

[24] ***Se refiere a Dennett. Content and Consciusness. Atlantic High-lands, NJ: Humanities Press. 1969.***

[25] ***No lo menciona, pero es casi seguro que se refiere a Churchland o a Stich (o a ambos).***

[26] ***Rorty (1979)***

[27] ***Se refiere al libro de Stich. From Folk Psychology to Cognitive Science: The case Against Belief. Cambridge, MA: MIT Press, 1983.***

Toda esta diatriba sin embargo no constituye un argumento claro contra el ME. Precisamente el supuesto fracaso del materialismo reduccionista según Putnam, deja abierta la puerta a las posturas eliminativistas. Lo que nuestro autor termina por señalar es que esta opción parece una locura… aunque eso nadie lo ha negado: el ME parece a primera vista y especialmente bajo las categorías mentalistas un disparate de grueso calibre (ya lo mencionamos en la introducción). Pero el que exista circularidad en la formulación de la eliminación teórica de las actitudes proposicionales o una presunta condición paradójica (o refutación *a priori*) en el intento de descartar el lenguaje intencional –sea con la estrategia "proyectiva" de Dennett o la eliminativa de Churchland o Stich- sigue siendo una formulación estándar del argumento trascendental contra el ME que ya discutimos más arriba (aunque ahora desenvuelto punto por punto y decorado con una enardecida retórica).

Ahora bien, haciendo un paréntesis en el desarrollo del argumento trascendental, la exposición de Putnam ha vuelto a poner sobre la mesa y con renovada fuerza un argumento antes discutido. Es preocupante la conclusión respecto de la completa irreductibilidad de todas las propiedades disposicionales. Si esto es verdad, el ME tendrá que enfrentar nuevamente una de las objeciones de Searle. Imprevistamente y buscando dar un buen cimiento teórico al argumento trascendental, vemos reaparecer el argumento de la irreductibilidad de las entidades de sentido común y las consecuencias inaceptables que este supuesto acarrea al materialismo en general. Si no contamos con la posibilidad de una reducción (como el análisis de Putnam parece mostrar) ¿nos atreveremos a proponer una eliminación de todas las propiedades disposicionales?... esto sin duda es ir demasiado lejos, pero ¿nos llevan hasta este punto las disquisiciones de Putnam?

No discutiré, al menos, una cuestión que me parece de suma evidencia. Tal es que las propiedades disposicionales, inclusive las propiedades caracterizables exclusivamente mediante el empleo de las "*condiciones normales*" y, por supuesto, la intencionalidad, no parecen tener perspectiva alguna de ser reducidas al lenguaje de la física de primer orden (a funciones de propiedades no disposicionales por ejemplo) *por el momento*. Pero Putnam ha afirmado algo bastante más osado que esto.

Según él no hay perspectiva *posible* de realizar tal reducción. Lo ha expresado claramente en los anteriores extractos[28].

Una cosa es decir que una reducción de las propiedades disposicionales al lenguaje de la física es una empresa titánica, que requeriría un esfuerzo y un tiempo incomparable con el de otra empresa teórica, que hay más probabilidades (temporales) de que un meteorito vagabundo ponga un abrupto fin al planeta tierra y a sus habitantes antes de que veamos realizada tal aspiración… ¡de acuerdo!... es una empresa *prácticamente* imposible… pero otra cosa muy distinta es sostener que tal reducción es una empresa *simple y llanamente imposible.*

[28] Recordemos y subrayemos: "(…) **De hecho, puede haber un número infinito de condiciones físicas diferentes** que podrían producir la disposición de reflejar (o emitir) luz roja y absorber luz de otras longitudes de onda" "(…) Puesto que a todo estado normal (todo estado en el cual el azúcar se disuelve) le corresponde un estado en el cual "se des-disuelve", **vemos que hay infinitamente muchas condiciones físicamente posibles** en las cuales el azúcar "se des-disuelve" en vez de permanecer disuelto." "(…)De hecho lo que podemos decir es que bajo condiciones *normales* el azúcar se disolverá si es colocado en agua. **Y no hay ninguna razón para pensar** que todas las diferentes condiciones no normales (incluyendo extraños estados cuánticos, extrañas fluctuaciones locales en el espacio-tiempo, etc.), bajo las cuales el azúcar no se disolvería si fuera colocado en agua, **puedan recogerse en una fórmula cerrada del lenguaje de la física fundamental** ." "(…)¿ **por qué deberíamos pensar ahora que *eso* es posible ?** Si se ha demostrado que reducir el color, la solidez o la solubilidad a la física fundamental **es imposible** , ¿por qué debería ser más fácil de resolver este programa reduccionista mucho más ambicioso?"

Pero si se pretende desacreditar teóricamente la empresa reduccionista en general no basta con insistir en los problemas prácticos que acarrea esta última. Es preciso mostrar que no es una visión de mundo (en términos de Putnam) consistente, sino que es desastrosa, paradójica; cuento corto: que es una visión de mundo *imposible*. Donde el último "*imposible*" debe interpretarse como "*imposible de concebir*", o más exactamente "*imposible de concebir con la evidencia con la que actualmente contamos*". Putnam tendría que sostener, en resumen, algo similar a esto: "Mira, el Realismo científico nos pareció coherente y plausible por algún tiempo, pero con la nueva información con que contamos[29] ha cambiado su rostro, se nos ha deformado y parécenos ahora una tesis estrictamente insostenible."

Lo central es, entonces, determinar si Putnam ha demostrado suficientemente que no hay ninguna perspectiva de reduccionismo de las propiedades disposicionales a ciertas "*funciones de propiedades no disposicionales*". Para él, el trabajo parece estar hecho desde que supone que, por ejemplo, podría haber infinitas anormalidades dispuestas a estropear la condición "estricta" de una disposición como *soluble*. La irreductibilidad de las nociones intencionales parece basarse en el mismo argumento desde la variante del funcionalismo propuesta por Putnam; así podrían existir infinitas clases de soportes para las actitudes proposicionales. Con esto, a su vez, se cree que se ha demostrado en algún sentido que el reduccionismo es falso.

Lo que yo creo en este respecto es que es necesario para el argumento que las condiciones anormales que no permiten a una disposición alcanzar el grado de "estricta"[30], sean infinitas para concluir la *imposibilidad* del reduccionismo. Si fueran, digamos, muchísimas pero finitas, podríamos *en principio* expresar una en fórmula cerrada todas las anomalías y así tener una fórmula completa y estricta. Tal fórmula podrá tener la extensión que se quiera imaginar; pudieran ser las anomalías mencionadas por Putnam en referencia a la "solubilidad" las únicas anomalías, con lo que la fórmula no será en extremo extensa, o bien podrían encontrarse setecientos millones de anomalías más, con lo que la fórmula requerirá quizá una vida humana entera para leerse. Sea como sea, si las anomalías son finitas pueden expresarse en una fórmula cerrada del lenguaje de la física, por lo que el reduccionismo es *prima facie* posible.

El argumento de Putnam es entonces concluyente supuesto un conjunto infinito de anomalías para cada disposición (presuntamente) irreducible.

De este modo Putnam nos acusa de promover una empresa práctica inalcanzable, "no sigan engañándose –nos dice- hay infinitas bases de realización e infinitas anomalías, por lo que no podrán reducir nunca la intencionalidad ni las propiedades disposicionales (en ese orden)". Pero nos toca replicar preguntando: ¿cómo puede saber él que tales cosas son infinitas?… o incluso, suponiendo que las anomalías por ejemplo, son infinitas en número (por así decirlo): ¿cómo puede saber que tales anomalías (supuestamente infinitas) no comparten una característica común que no hemos elucidado y que tal característica podría dispensarnos de referirnos a todas y cada una de ellas al traducir las disposiciones a una fórmula cerrada, en un determinado programa reductivo[31]?

[29] La supuesta imposibilidad del reduccionismo.

[30] Pasaré a llamarlas "anomalías".

[31] Esto es algo parecido a la solución funcionalista que el propio Putnam proponía como programa en filosofía de la mente a mediados de los 50'.

Si no me equivoco, Putnam no puede saber estas cosas, por lo que no pudo demostrar ninguna imposibilidad en el proyecto reduccionista de los Realistas científicos. Y esto porque no veo como alguien puede conocer esta clase de hechos; hechos tales como: "dado un número de anomalías siempre habrá otra distinta". Esta clase de cosas quizás puedan saberse en ciencias formales, pero no llego a comprender como se puede estar tan seguro de la infinita variedad de ciertos eventos empíricos… ¿Cómo podemos estar justificados en creer que, por ejemplo, "siempre hay una anomalía más" aunque no podamos especificarlas?

En conclusión la exposición de Putnam nos ha permitido desenvolver más acuciosamente dos objeciones contra el ME. Por una parte pretende justificar la necesidad y constante reaparición de las categorías intencionales en el discurso teórico, presentando así como un absurdo digno de la más osada y descriteriada doctrina parisiense cualquier programa que atente contra esta categoría. Por otra parte ha vuelto a poner en juego la observación de Searle acerca de la irreductibilidad de ciertas entidades y propiedades dando más contundencia a esta objeción. Pese a todo esto, su versión de la objeción trascendental no parece aportar más evidencia que la expresión "tosca" del mismo argumento, sino simplemente un mayor y más fino desarrollo[32]. Por otra parte la discusión acerca del reduccionismo y su imposibilidad no me parece conclusiva por lo expuesto anteriormente.

Sea como sea, y suponiendo que no podemos probar que los argumentos trascendentales son peticiones de principio a buenas y primeras, podemos aun asir un experimento mental en su contra.

Tal es el experimento mental de los Antípodas desarrollado por Rorty (1979) y expuesto en la introducción. Si algún argumento trascendental es verdadero, los Antípodas son inconcebibles. Pero si Rorty está en lo cierto nada de esto ocurre. Es apremiante entonces a esta clase particular de argumentación refutar tajantemente la condición de posibilidad del experimento mental de los Antípodas. Acerca de esta potencial refutación no tengo mayor información.

[32] Es decidor en este respecto que Putnam termine sólo profiriendo acusaciones de filiación teórica por parte de los materialistas eliminativos con doctrinas bastante desacreditadas –por la filosofía anglosajona al menos- como es el postmodernismo en vez de brindar una refutación concisa.

Capítulo 4. Eliminando a la competencia

Hemos visto en general dos clases de argumentos contra el ME y podemos concluir que la discusión que cualquier filosofía de la mente puede establecer con las tesis eliminativistas desencadenará en dos puntos clave: 1) Si se ataca alguno de los supuestos del argumento básico eliminativista, se atacará necesariamente C). 2) Si se intenta un argumento trascendental (que evite de alguna forma la petición de principios) se debe refutar de alguna forma el experimento mental de los antípodas.

Si esto es correcto no existen otros bastiones donde el eliminativista pueda defender su tesis. No pretendo aquí atacar o defender ni C) ni el experimento mental de los antípodas (no se me ocurre de momento como resolver rápidamente ninguno de estos puntos –siquiera sé si son solucionables-). Lo que puede concluirse de lo anterior es que ningún programa serio en filosofía de la mente puede pretender haber refutado al ME sin haber resuelto a favor de sus propias tesis al menos uno de ambos problemas (tampoco sé si resolver uno no sería inmediatamente resolver el otro).

Lo que presentaré a continuación es una especie de "historia universal de la infamia" desde el punto de vista del ME. Esto es un análisis de algunos programas de filosofía de la mente rechazados por el ME y las razones propias del eliminativistas para desacreditarlos.

Más allá de enumerar las dificultades específicas de cada programa tratado, veré por qué el ME supone que cada uno de esos programas no consiguen hacer lo que él cree conseguir; esto es: *ser verdaderamente materialista*. Al eliminativista no parecen desagradarle por sí mismos los otros programas en filosofía de la mente, todas sus quejas parecen estar fundadas en que ninguno abandona verdaderamente el dualismo.

Según la visión del ME el dualismo (sea sustancial o de propiedades) es un viejo mañoso y remilgón que hace insoportable la convivencia en el hogar (insulta, da golpes con su bastón, escupe en la mesa, etc.). La casa, sin lugar a dudas, está habitada por sus jóvenes parientes materialistas (la teoría de la identidad, el conductismo filosófico, el funcionalismo y el eliminativismo[33]). Un buen día el materialista eliminativo ha dicho en la mesa lo que todos pensaban en cierta medida, que ya era tiempo de mandar al abuelo a un asilo permanentemente. Inmediatamente la discusión se volvió más pasional de lo necesario –según el eliminativista-. Algunos familiares, auque cansados del abuelo-dualista, todavía gustaban de algunas de sus historias y querían conservarlo en casa para no perder parte de esta rica herencia familiar; otros con lágrimas en los ojos acusaron al materialista eliminativo de cruel e inhumano "¡Sabes que el abuelo se muere si lo mandamos a un asilo!" decían "¡¿Quieres verlo, muerto verdad; eso quieres?!". A lo cual el eliminativista no quiere responder claramente ("quizá el abuelo estaría mejor muerto que en este estado de deplorable senilidad" se decía en su fuero interno). Sea como sea, nuestro héroe decidió que todo aquel que apoyara tácitamente al abuelo apoyaba la decadencia de una convivencia sana en el hogar (léase, ciencia seria en la sociedad). Finalmente puso las cosas claras: o están conmigo o están contra mí… no hay término medio. ¿Será esto cierto?... ¿Habrá quizá alguna forma en que el abuelo pueda seguir conviviendo en la casa

[33] Digamos que el epifenomenalismo está gravemente enfermo y que recibe ocasionales visitas por parte de estos familiares en el hospital.

sin estorbar las pretensiones materialistas de sus familiares? Lo que el ME pretende mostrar en definitiva es que todo materialismo que deje la puerta abierta a alguna forma de dualismo es un mal materialismo. El criterio de evaluación que parece caracterizar los argumentos del ME en contra de los demás programas en filosofía de la mente parece ser el siguiente: si usted es compatible con el dualismo, está usted fuera.

Como ya habíamos sugerido no analizaremos el epifenomenalismo, sino simplemente las tres principales doctrinas competidoras del ME: el conductismo como filosofía de la mente (aunque esté un poco viejo), la teoría de la identidad, y el funcionalismo.

Conductismo filosófico : El conductismo intentó caracterizar los estados mentales reduciéndolos a estados conductuales públicamente observables. Estos estados sin duda eran descritos como ciertas disposiciones observables. Un primer problema consistía en especificar el *definiens* de tales disposiciones en términos estrictamente fisiológicos. Una actitud proposicional como "querer comprar un sándwich" requiere de un *definiens* virtualmente interminable supuesto que no se incorporen expresiones tales como "*ceteris paribus*", "en condiciones normales", "generalmente", etc. No es probable que las disposiciones descritas por el conductista puedan ser realmente explicitadas, pero aun suponiendo esto, parece una dificultad superlativa no emplear ninguna categoría mental como *definiens* de una disposición conductual como "querer comprar un sándwich". Digamos que Juan da muestras de "querer comprar un sándwich" cuando se acerca a un local de lomitos y pide un sándwich y se apresta a pagarlo, esto podremos concluirlo siempre y cuando supongamos que Juan "no quiere jugarle una broma al vendedor de sándwiches" (otras muchas actitudes proposicionales pueden ser relevantes para el *definiens*). ¿Podemos alcanzar un *definiens* aproximado de "querer jugarle una broma a alguien" sin apelar a otros estados mentales (como el mismo "querer" que deseábamos definir)? Lo cierto es que se hizo más énfasis a la interminable cadena definitoria como objeción al conductismo que a esta reaparición de lo mental en su discurso, por lo que no estoy seguro de si existe algún programa conductista que haya pretendido ser estrictamente fisicalista en sus cadenas definitorias (probablemente se rindió antes), el punto es que, al menos *prima facie*, no sabemos si le es posible al conductista sacudirse esta infiltración dualista de su vocabulario.

Por otra parte el conductismo es en verdad una doctrina vaga acerca de los estados mentales privados. Si bien no los acepta metodológicamente no aporta ningún argumento en contra de su existencia. En resumen, y más allá del potencial mentalismo implícito en sus prácticas semánticas, el conductismo es de plano compatible con el dualismo.

Teoría de la identidad : La teoría de la identidad por su parte formula enunciados del siguiente tipo (sigo aquí a Feyerabend (1963)): "*x* es un proceso mental de la clase A ¦*x* es un proceso central de la clase α". Pero esta hipótesis de identidad en ningún caso muestra ninguna primacía del carácter físico por sobre el mental. El dualista puede felizmente leerlo de derecha a izquierda y concluir que ciertos eventos físicos tienen características mentales. Así en vez de un dualismo de eventos (como bien señala Feyerabend) obtenemos un dualismo de características. No estamos en condiciones de decidir una primacía en la identidad, aunque tácitamente hayamos querido siempre ser materialistas.

El principal problema de toda reducción interteórica es el hecho de que las condiciones de reducción suponen que el vocabulario de la teoría reductora podrá seguir refiriéndose a todas las características y rasgos del vocabulario la teoría reducida, pero el vocabulario de lo mental (como está presentado al menos en la psicología folk) no parece convenir en múltiples rasgos (algunos de ellos muy profundos, ontológicos –supóngase la no espacialidad de lo mental simplemente-) con las descripciones de las teorías fisicalistas

(neurociencia por ejemplo). Sea como sea si la psicología folk es una teoría falsa, la reducción no será posible nunca, por lo que el teórico de la identidad está en la paradójica situación de querer negar ciertos hechos ontológicamente incompatibles con las taxonomizaciones de sus mejores teorías (los hechos mentales), aunque asume en cierto sentido que la teoría que agrupa y se refiere a estos hechos (la psicología folk) es verdadera –de lo contrario, ¿cómo reducirla?-.

La teoría de la identidad es, en resumen, paradójicamente compatible con el dualismo (exceptuando algún reduccionista que no afirme tener ningún problema con las categorías y taxonomizaciones mentales –no sería un buen materialista de todos modos), por lo que no representa bien (en términos de Feyerabend) el caso que pretende defender (el materialismo).

Funcionalismo : El funcionalismo es sin duda uno de los mejores y más desarrollados programas en filosofía de la mente. Los embistes eliminativistas contra el funcionalismo son mucho más complicados y la acusación de compatibilidad con el dualismo es menos fuerte en este caso (aunque el funcionalismo es, casi por esencia y sin vergüenza, un tanto compatible con buena parte del vocabulario mentalista).

Puesto que funcionalismo supone que el rasgo esencial o definitorio de todo tipo mental es un conjunto de relaciones causales mantenidas con los efectos ambientales sobre el cuerpo, otros estados mentales y la conducta del cuerpo, y al suponer que las bases de realización de estos estados funcionales pueden ser múltiples, puede justificadamente pretender que la psicología tenga al menos cierta independencia metodológica con respecto a las ciencias físicas que estudian el cerebro, y a su vez pretender ser materialista desde el momento en que postula una dependencia ontológica. Ejemplificaremos varios de los problemas que el ME podría encontrar en la tesis funcionalista mediante el célebre experimento mental de los qualia invertidos. El funcionalismo debe sin duda hacerse cargo de este contra argumento. Para esto debe disolver la incongruencia entre las siguientes afirmaciones: 1) Que la identidad-tipo de las sensaciones visuales –por ejemplo- se infiere a partir de su rol funcional. 2) Que es concebible una inversión relativa de qualia entre dos seres sin inversión funcional.

Esta incongruencia puede ser resuelta por el funcionalista, en términos de Churchland, afirmando lo siguiente:

"1) los estados funcionales (o mejor dicho, sus realizaciones físicas) verdaderamente tienen una naturaleza intrínseca de la que depende nuestra identificación de esos estados por medio de la introspección, y también que 2) esa naturaleza intrínseca sin embargo no es esencial para la identidad-tipo de un determinado estado mental y de hecho puede variar entre un ejemplo y otro del mismo tipo de estado mental. Esto significa que el carácter cualitativo de su sensación del rojo podría ser diferente del mío, en forma leve o sustancial, y también cabe la posibilidad de que lo sea la sensación del rojo de una tercera persona. Pero en la medida en que los tres estados son provocados normalmente por objetos rojos y normalmente son la causa de que nosotros tres creamos que algo es rojo, entonces los tres estados son sensaciones del rojo, cualquiera que sea su carácter cualitativo intrínseco. Estos qualia intrínsecos simplemente constituyen rasgos prominentes que permiten una rápida identificación introspectiva de las sensaciones, así como las rayas negras sobre el color naranja constituyen un rasgo dominante para la rápida

identificación visual de los tigres. Pero no existen qualia específicos esenciales para determinar la identidad-tipo de los estados mentales, así como no hay rayas negras sobre color naranja que sean esenciales para constituir la identidad-tipo de los tigres" (Churchland, 1988, pag 70-71)

Pero el funcionalista debe reconocer la realidad de los qualia si quiere seguir esta alternativa. La pregunta es de qué modo pretende incorporarlos dentro de su programa. La única vía no dualista que parece proponerse es cierta identificación de estos qualia con el estado físico que realiza los estados mentales (funcionales) que exhiben estos estados cualitativos. Esto no supone que las bases físicas para la generación de tales qualia deba ser idéntica (como creía el teórico de la identidad), puesto que cualquier propiedad física que ejemplifica un estado funcional como los descritos es suficiente para la explicación. Así, si un mecanismo de identificación de un color determinado se relaciona con una frecuencia ondulatoria determinada, no estamos forzados a suponer que la base física que realiza tal discriminación sea físicamente isomórfica a la nuestra (podría ser un robot cuyo cerebro esté compuesto de un cableado de cobre, pero que es capaz de discriminar la frecuencia ondulatoria de 90 hertz y discriminar cualitativamente, por ende, el color rojo).

Pero si esta es la respuesta del funcionalista deberá reconocer que su tesis no difiere tanto de la del teórico de la identidad. Hay muchos ejemplos en los cuales una reducción interteórica de ciertos términos encuentra "múltiples ejemplificaciones"[34]. Pero esta multiplicidad no fundamenta en ciencias naturales nada parecido a una "independencia metodológica" (los múltiples ejemplos del concepto de "temperatura" no generan independencias metodológicas dentro de la termodinámica)… ¿por qué los múltiples ejemplos de las reducciones de las propiedades mentales (supuesto que no son tan distintas de las del teórico de la identidad) supondrían tal independencia?

El funcionalista se verá forzado a reconocer al menos tres puntos desagradables: 1) Que algunas reducciones tipo/tipo de estados mentales a estados físicos son posibles y hasta presumiblemente necesarias para completar el programa funcionalista. 2) Que no es evidente y probablemente no hay una independencia metodológica radical de la psicología. 3) Y más en general, que el funcionalismo no es tan distinto como creía de la teoría de la identidad.

A primera vista los puntos 2) y 3) parecerán los más desagradables al funcionalista, pero en verdad es el punto 1) el que genera una controversia más ostensible según el eliminativista. Si el funcionalista debe reconocer ciertas reducciones análogas a las presupuestas por el teórico de la identidad, ¿cómo podrá evitar la acusación de dualismo que hemos hecho en contra de este último? Finalmente para salvarse de la objeción de los qualia invertidos (también se aplica a los qualia ausentes -zombies-) ha tenido que aceptar al menos parcialmente una hipótesis reduccionista que implica un dualismo de caracteres según hemos mostrado más arriba.

Pero quizá el funcionalista se sienta cómodo con alguna de sus coincidencias con el reduccionismo. En rigor podría abandonar el funcionalismo clásico e intentar una "reducción funcional" al estilo de Kim (1998). Esto dejará afuera el problema de los qualia considerándolos instancias "no funcionalizables". El argumento de Kim es resumido del siguiente modo por Perez (2006):

[34] Supógase la reducción de "temperatura = promedio de energía cinético molecular", esta identidad es válida estrictamente para la temperatura de un gas. La temperatura en un sólido debe entenderse en términos algo distintos (movimientos vibratorios), así también la de un plasma (que no está constituido por moléculas), así como al vacío se adscribe una temperatura "antirradiante", etc. (La fuente de estos datos es Churchland 1988)

(i) Los qualia no se pueden funcionalizar (por inversión de especto y ausencia de qualia).

(ii) Lo que no se puede funcionalizar no se puede reducir funcionalmente

(iii) Lo que no se puede reducir funcionalmente no puede superar los problemas de causación mental.

(iv) Lo que no puede superar los problemas de causación mental es causalmente inerte.

(v) Lo que es causalmente inerte no es real (por *dictum* de Alexander: *ser real es tener poder causal*).

Conclusión: Los qualia no son reales (no existen).

Esta consecuencia es sin embargo bastante sorprendente. El argumento funcionalista de Kim parece haber conseguido, en definitiva, en un marco cercano al funcionalismo lo que el eliminativista venía pidiendo desde el principio (al menos para los qualia). Pero incluso el vocabulario mental sometido a reducción funcional parece tener un estatus dudoso dentro del programa de Kim. Pérez lo resume del siguiente modo:

"Para que la reducción funcional sea posible, dice Kim, hay que "dejar de pensar (a lo mental) como una propiedad intrínseca y constituirlo como una propiedad extrínseca caracterizada relacionalmente" (Kim 1998, p.25). Pero advierte más adelante en el último capítulo, una vez que hemos logrado la caracterización funcional de un cierto fenómeno esto es, una vez que hemos caracterizado en términos de una propiedad de segundo orden (es decir, una propiedad funcional) a un cierto fenómeno y pretendemos identificar dicha propiedad de segundo orden con una propiedad de primer orden (es decir, con el ocupante del rol funcional en cuestión, que es quien, al fin y al cabo realiza el "trabajo" causal, resolviendo los problemas de causación mental), Kim se pregunta (¿sorprendido?): ¿cómo pueden ser idénticas una propiedad de segundo orden con una de primer orden? Y se responde, y nos responde: lo que ha ocurrido es que, en sentido estricto, no tenemos que aceptar que hay en el mundo dos propiedades, una de primer orden y otra de segundo orden, en realidad lo que tenemos es una propiedad intrínseca (la de primer orden) y una caracterización extrínseca, relacional de segundo orden, de esa misma propiedad intrínseca, es decir, lo funcional, lo relacional, no son más que conceptos, maneras alternativas de describir lo que realmente hay: a saber, las propiedades de primer orden, la intrínsecas, las que hacen el trabajo causal, esto es las propiedades físicas" (Pérez, 2006, pag 9)

Parece en última instancia, como bien señala Pérez, que la variante reduccionista de Kim (aun matizada por el funcionalismo) desencadena en algo cercano a la tesis eliminativista o al menos en una eliminación por reducción. No hay algo así en el mundo como esas propiedades de segundo orden, y no corresponde identificarlas en ningún sentido con propiedades de primer orden. Si funcionalizarlas consiste en "describirlas alternativamente" no creo que haya grandes quejas por parte del ME, siempre y cuando se recalque constantemente que sólo hablamos de palabras y conceptos alternativos, quizá cómodos en la cotidianidad, pero que en definitiva no se refieren a ninguna cosa que *haya* realmente en el mundo (sólo habrán propiedades físicas) y que eventualmente puedan (estos términos) desaparecer supuesto el desarrollo de otro lenguaje más poderoso (presumiblemente el de la neurociencia).

Pero es dudoso que Kim quisiera tratar de tan poca cosa toda su reducción funcional. Es más probable que simplemente pretendiera mostrar en qué sentido una teoría "burda pero útil" acerca de ciertos fenómenos es reducible a otra teoría "fina pero poco práctica".

Sea como sea aun contamos con un argumento adicional contra las diversas empresas de tilde funcionalista, esto es la opinión de Churchland (1989). Churchland sostiene que el funcionalismo es peligrosamente tolerante a la reducción de teorías falsas. Su ejemplo es tan simple como gráfico: él imagina una reducción funcional de la alquimia (con su vocabulario de "espíritus animales, etc.") a la química moderna. Tal reducción es posible y verosímil –aun si su aceptación fuera nula-, lo cual prueba que las reducciones funcionalistas no dan garantías de estar reduciendo teorías verdaderas (a diferencia del teórico de la identidad, el funcionalista sí lograría reducir teorías falsas).

Podría argumentarse que la psicología folk no es comparable con la alquimia, que nadie aceptaría hoy en día la alquimia en presencia de una mejor teoría incompatible con esta (y que por ende no se justifica una tal reducción funcional), pero esto parece apelar simplemente a ciertas intuiciones privilegiadas que poseemos acerca de nuestra psicología folk –intuiciones que la harían más irresistible a nuestros ojos y que nos haría más tolerantes a su reducción funcional (por inútil que fuera ésta)-. Si se quiere, esta queja es análoga a la petición de principios expuesta en las objeciones al ME que llamamos trascendentales y también es un rechazo *a priori* del supuesto C). ¿No le parecería a un Antípoda una completa payasada una reducción funcional de nuestra psicología folk tanto como a nosotros nos lo parece la reducción funcional de la alquimia[35]?

Pero si la reducción funcional de Kim comparte este rasgo dudoso del funcionalismo resulta en definitiva bastante poco interesante todo el esfuerzo de la empresa. Aun si hubiese algo que rescatar dentro de las propiedades de segundo orden funcionalizadas en esta reducción, ¿cómo asegurar que no son propiedades imaginadas desde una teoría falsa y calzadas a la fuerza en el puzzle de las ciencias naturales mediante el modelo funcionalista?

Aun suponiendo que la tesis de Kim soporte las acusaciones de compatibilidad con el dualismo expuestas al funcionalismo clásico, no parece tener buenos garantes de fiabilidad ni relevancia en comparación la tesis del ME.

[35] Y por qué no podemos imaginar a algunos hombres (o a todos) con alguna clase de condición médica psiquiátrica extraña (o con una dieta alta en LSD) que tengan una relación bastante directa y hasta irresistiblemente clara y patente con algunos de los términos fundamentales de la teoría alquímica (digamos, con los espíritus animales). ¿Cambiaría esto en algo la situación?

Conclusión. Y bosquejo de discusión epistemológica

El ME es una doctrina simple si su defensor quiere dejar de lado toda forma de “intuición básica” acerca del mundo y nuestro contacto cognitivo con él. En términos más sencillos: alguien que no se comprometa con ninguna forma de sensaciones primarias, actitudes proposicionales o incluso estados intencionales o privados –estados en primera persona- considerará sin lugar a dudas al ME como un candidato tentador para resolver (disolver digamos) el problema mente cuerpo. Si nuestros análisis de objeciones contra el ME y desafíos propuestos a los otros programas en filosofía de la mente por parte de este último son correctos, quien desacredite al ME parece estar comprometido de antemano con estados tales como los que hemos agrupado bajo el apodo de “intuiciones básicas”. Si esto es verdad será quizás más provechoso discutir la validez de la tesis C) antes siquiera de intentar desenvolver cualquier programa tendiente a resolver el problema mente cuerpo.

Por mi parte no he tocado en todo este escrito es el supuesto C) que he adscrito al ME y la verdad no está dentro de mis pretensiones –como señalé antes- discutirlo en este trabajo. Pero haré al menos una acotación al respecto. Varios puntos tratados aquí muestran cierta particularidad en la relación entre la epistemología, la ontología, la filosofía del lenguaje y la filosofía de la mente a la que suelen suscribir los programas más cercanos que el ME al mentalismo y al dualismo (aunque sólo sea por compatibilidad).

Esta particularidad se hace explícita en varios casos: al tratar de decidir qué teoría de la referencia es adecuada en estas discusiones (según creo la elección de la teoría causal nos sumerge de cabeza en la incorregibilidad), a su vez se hace patente al mostrar las compatibilidades de los diversos programas de filosofía de la mente con el dualismo; también en el caso especial del funcionalismo tanto en su tolerancia según Churchland (1989) a la funcionalización de teorías falsas (que ejemplificamos con los Antípodas y la alquimia), como en el particular dilema en el que la reducción funcional de Kim cae (una especie de indecisión entre eliminativismo y funcionalismo), y finalmente y con mucha claridad, en el ejemplo mismo de los Antípodas propuesto por Rorty (1979).

Lo que todos estos casos parecen mostrar es lo siguiente: que la epistemología que es empleada en las discusiones de filosofía de la mente (excepción hecha por el ME) está constituida de antemano con términos mentalistas (mito de lo dado, actitudes proposicionales, sensaciones primarias, etc.). Esta opinión no es nueva si consideramos las críticas de Kim (1988) al proyecto quieneano de naturalización de la epistemología. En cualquier caso está constitución previamente mentalista de la epistemología seleccionada para discutir el problema mente-cuerpo parece generar la extrañeza y las situaciones paradójicas en las que entran los programas “tolerantes para con el mentalismo” al discutir con un programa esencialmente “intolerante para con el mentalismo” como es el eliminativismo.

Pero si todo esto es verdad (si los programas “tolerantes/compatibles” asumen en sus objeciones y argumentos una epistemología “mentalistamente cargada”) entonces hay un círculo en la prueba, al menos desde una visión amplia que abarca filosofía de la mente, epistemología, ontología y filosofía del lenguaje. Ninguna de estas distintivas áreas de la

filosofía alcanza una justificación cercanamente completa sin presuponer el funcionamiento acabado de las otras.

Ahora bien, quien esté comprometido con alguna premisa como C) habrá cometido más de una vez un círculo en la prueba, aunque sin ningún escrúpulo ni vergüenza (tómese el ejemplo de Quine (1986) en *epistemología naturalizada*).

En definitiva si ambos bandos cometen el mismo pecado (suponiendo que es un pecado) parece coherente la siguiente exigencia del ME y del pragmatismo en general: dejemos de vernos la suerte entre gitanos y atendamos sencillamente a los programas que presenten mayores "virtudes" epistemológicas comunes (si son capaces realmente de encontrar entre todos los bandos tales cosas): ¿Qué explicación es más simple, cual más coherente con las mejores teorías disponibles, quien tiene mayor cúmulo de evidencia empírica a favor y menor en contra?... este enfoque del problema ha sido siempre al que ha querido llegar el ME… ¿Está justificado en llevar las cosas tan lejos?

Bibliogrtafía

Churchland, P. (1988), *Matter and Consciousness*, MIT Press, Cambridge, Massachusetts.

Churchland, P. (1989), *Eliminative Materialisme and the Propositional Attitudes*, Repr. In P. Churchland, pp 1-22.

Feyerabend, P. (1963), *Mental Events and the Brain*, en Journal of philosophy LX, 11 (may 23).

Kim, J. (1988), *What is naturalized epistemology?*, en Martín Alcoff, L., editor, *Epistemology: The Big Questions* , pages 265-281. Blackwell.

Kim, J. (1998), *Mind in a Physical World,* MIT Press, London.

Kripke, S. (1971), *Naming and Necessity*, Semantics of Natural Language,

Pérez, D. (2006), *Eliminativismo, cambio conceptual y conceptos mentales*, en Rev. Int. Fil, v29, n.2, pp 707-727, jul-dez.

Putnam, H. (1979), *El significado de "significado",* en *La búsqueda del significado*, Tecnos, Madrid.

Putnam, H. (1987) *Las mil caras del realismo*, Paidos Ibérica, Barcelona.

Quine, W.V.O. (1986) *La relatividad ontológica y otros ensayos*, Tecnos, Madrid

Rorty, R. (1979), *Philosophy and the Mirror of Nature*, Princeton University Press, New Jersey.

Searle, J. (1992), *El redescubrimiento de la mente*, Crítica Grijalbo Mondadori, Barcelona.