

## CAPÍTULO 4

### LAS DIFICULTADES DEL FUNCIONALISMO (SELECCIÓN) \*

Ned Block

#### 1.0 Funcionalismo, conductismo y fisicalismo

El punto de vista funcionalista acerca de la naturaleza de la mente es, en este momento, ampliamente aceptado.<sup>1</sup> Al igual que el conductismo y el fisicalismo, el funcionalismo pretende responder a la pregunta “¿Qué son los estados mentales?”. Me ocuparé de las formulaciones del funcionalismo *via* la tesis de la identidad. Ellas dicen, por ejemplo, que el dolor es un estado funcional, así como las formulaciones del fisicalismo *via* la tesis de la identidad dicen que el dolor es un estado físico.

Comenzaré describiendo al funcionalismo y esbozando la crítica funcionalista al conductismo y al fisicalismo. Luego argumentaré que las dificultades atribuidas por el funcionalismo al conductismo y al fisicalismo infectan también al funcionalismo.

Una caracterización del funcionalismo que es probable que sea lo suficientemente vaga como para ser aceptada por la mayoría de los funcionalistas, es la siguiente: cada tipo [*type*] de estado mental es un estado que consiste en una disposición a actuar de ciertas maneras *y a tener ciertos estados mentales*, dados ciertos *inputs* sensoriales y ciertos estados mentales. Expuesto de este modo, el funcionalismo puede verse como una nueva encarnación del conductismo. El conductismo identifica a los estados mentales con disposiciones a actuar de ciertas maneras en ciertas situaciones de *input*. Pero como han señalado sus críticos

\* “Troubles with Functionalism”, en *Perception and Cognition. Minnesota Studies in the Philosophy of Science. Vol. IX*, compilado por W. Savage, 1978. Con autorización del autor y de Minnesota University Press.

1. Véase Fodor, 1965; Lewis, 1972; Putnam, 1966, 1967, 1970, 1975a; Armstrong, 1968; quizá Sellars, 1968; quizá Dennett, 1969, 1978b; Nelson, 1969, 1975 (pero véase también Nelson, 1976); Pitcher, 1971; Smart, 1971; Block y Fodor, 1972; Harman, 1973; Grice, 1975; Shoemaker, 1975; Wiggins, 1975.

(Chisholm, 1957; Geach, 1957; Putnam, 1963), desear G como meta, no puede identificarse con, digamos, la disposición a hacer A en circunstancias de *input* en las cuales A conduce a G, puesto que, después de todo, el agente podría no *saber* que A conduce a G y de este modo podría no estar dispuesto a hacer A. El funcionalismo reemplaza a los “*inputs sensoriales*” conductistas por “*inputs sensoriales y estados mentales*”, y el funcionalismo reemplaza las “*disposiciones a actuar*” conductistas por “*disposiciones a actuar y tener ciertos estados mentales*”. Los funcionalistas quieren individuar causalmente a los estados mentales, y puesto que los estados mentales tienen causas y efectos mentales tanto como causas sensoriales y efectos conductuales, los funcionalistas individúan a los estados mentales, en parte, en términos de las relaciones causales con otros estados mentales. Una consecuencia de esta diferencia entre el funcionalismo y el conductismo es que existen organismos posibles que de acuerdo con el conductismo tienen estados mentales pero que, de acuerdo con el funcionalismo, no los tienen.

De tal modo, las condiciones necesarias de lo mental que el funcionalismo postula son, en un aspecto, más fuertes que las postuladas por el conductismo. De acuerdo con el conductismo, es necesario y suficiente para desear que G, que un sistema sea caracterizado por un cierto conjunto (quizás infinito) de relaciones *input-output*; es decir, de acuerdo con el conductismo, un sistema desea que G en el caso de que un cierto conjunto de condicionales de la forma “Emitirá O dado I” sea verdadero de él. Sin embargo, de acuerdo con el funcionalismo, un sistema podría tener esas relaciones *input-output* aunque no deseara que G; porque de acuerdo con el funcionalismo, que un sistema deseé que G depende de que tal sistema tenga estados internos que tienen ciertas relaciones causales con otros estados internos (y con *inputs* y *outputs*). Puesto que el conductismo no apela al requerimiento de “estado interno”, existen sistemas posibles de los cuales el conductismo afirma y el funcionalismo niega que tengan estados mentales.<sup>2</sup> Una manera de enunciar esto es que, de acuerdo con el funcionalismo, el conductismo peca de *liberalismo*, al adscribir propiedades mentales a cosas que de hecho no las tienen.

A pesar de la diferencia entre funcionalismo y conductismo que acabamos de esbozar, no es necesario que los funcionalistas y los conductistas no compartan un mismo espíritu.<sup>3</sup> Shoemaker (1975), por ejemplo,

dice: “En una de sus interpretaciones, el funcionalismo en la filosofía de la mente es la doctrina de que los términos mentales o psicológicos son, en principio, eliminables de una cierta manera” (págs. 306-7). Los funcionalistas han tendido a tratar a los términos de estado-mental en una caracterización funcional de un estado mental, de manera muy diferente de la de los términos de *input* y de *output*. Así, en la versión más simple de la teoría, en términos de máquina de Turing (Putnam, 1967; Block y Fodor, 1972), los estados mentales se identifican con la totalidad de los estados de máquina-de-Turing, los que se definen a sí mismos *implícitamente* mediante una tabla de máquina que menciona *explícitamente* los *inputs* y los *outputs*, descriptos de manera no-mentalista.

Según la versión del funcionalismo que ofrece Lewis, los términos de estado-mental se definen por medio de una modificación del método de Ramsey, de manera tal que elimina el uso esencial de terminología mental de las definiciones pero no elimina la terminología de *input* y *output*. Es decir, ‘dolor’ se define como sinónimo de una descripción definida que contiene términos de *input* y de *output* pero no terminología mental (véase Lewis, 1972).

Además, el funcionalismo tanto en las versiones de máquina como en las que no son de máquina, insistió de modo típico en que las caracterizaciones de los estados mentales deberían contener descripciones de *inputs* y de *outputs* en lenguaje *físico*. Armstrong (1968), por ejemplo, dice:

Podemos distinguir entre ‘conducta física’, que refiere a cualquier acción o pasión del cuerpo meramente física, y ‘conducta propiamente dicha’, que implica relación con la mente... Ahora bien, si en nuestra fórmula [“estado de la persona apto para producir cierta clase de conducta”] ‘conducta’ significara ‘conducta propiamente dicha’, entonces estaríamos dando una explicación de los conceptos mentales en términos de un concepto que ya presupone la mentalidad, lo cual sería circular. De este modo, resulta claro que en nuestra fórmula ‘conducta’ tiene que significar ‘conducta física’ (pág. 84).

En consecuencia, puede decirse que el funcionalismo “ubica” a los estados mentales sólo en la periferia, es decir, mediante la especificación física, o al menos no-mental, de *inputs* y de *outputs*. Una tesis básica de este artículo es que, a causa de este rasgo, el funcionalismo no puede

---

minos mentales pueden definirse en términos no-mentales, entonces el funcionalismo es una versión del conductismo.

2. La inversa es también verdadera.

3. Ciertamente, si uno define ‘conductismo’ como el punto de vista de que los té-

evitar el tipo de problema por el cual condena correctamente al conductismo. El funcionalismo también peca de liberalismo, por razones muy similares a las del conductismo. Sin embargo, a diferencia del conductismo, el funcionalismo puede ser alterado con naturalidad para evitar el liberalismo; pero sólo al precio de fracasar de manera igualmente ignominiosa.

El fracaso del que hablo es el que el funcionalismo atribuye al *fiscalismo*. Por ‘fiscalismo’ significa la doctrina de que el dolor, por ejemplo, es idéntico a un estado físico (o fisiológico).<sup>4</sup> Como muchos filósofos argumentaron (notablemente Fodor, 1965; Putnam, 1966; véase también Block y Fodor, 1972), si el funcionalismo es verdadero, es probable que el fiscalismo sea falso. Este punto se ve más claramente en relación con las versiones del funcionalismo en términos de máquina de Turing. Cualquier máquina de Turing abstracta dada puede realizarse en una amplia variedad de dispositivos físicos; es plausible, por cierto, que dada una correspondencia putativa entre un estado de máquina de Turing y un estado de configuración física (o fisiológica), habrá una realización posible de la máquina de Turing que proporcionará un contraejemplo a esa correspondencia. (Véase Kalke, 1969; Gendron, 1971, y Mucciolo, 1974, para argumentos en contra no-convictos; véase también Kim, 1972.) En consecuencia, si dolor es un estado funcional no puede, por ejemplo, ser un estado cerebral, porque las criaturas sin cerebro pueden realizar la misma máquina de Turing que las criaturas con cerebro.

Tengo que destacar que el argumento funcionalista contra el fiscalismo no apela, meramente, al hecho de que una máquina de Turing abstracta pueda ser realizada mediante sistemas de *composición material* diferente (madera, metal, vidrio, etcétera). Argumentar de este modo

4. Estado tipo [*state type*], no estado caso [*token*]. A lo largo de este artículo, entenderé por ‘fiscalismo’ la doctrina que dice que cada tipo distinto de estado mental es idéntico a un tipo distinto de estado físico; por ejemplo, dolor (el universal) es un estado físico. El fiscalismo de casos, por otra parte, es la doctrina (más débil) de que cada dolor particular fechable es un estado físico de uno u otro tipo. El funcionalismo muestra que el fiscalismo de tipos es falso, pero no muestra que el fiscalismo de casos sea falso. Por ‘fiscalismo’ entiendo fiscalismo de *primer orden*; la doctrina de que, por ejemplo, la propiedad de tener dolor es una propiedad física de primer orden (en el sentido Russell-Whitehead). (Una propiedad de primer orden es aquella cuya definición no requiere cuantificación sobre propiedades; una propiedad de segundo orden es aquella cuya definición requiere cuantificación sobre propiedades de primer orden, y sobre ninguna otra propiedad.) La afirmación de que tener un dolor es una propiedad física de segundo orden es en realidad una forma (fiscalista) de funcionalismo. Véase Putnam, 1970.

sería como argumentar que la temperatura no puede ser una magnitud microfísica porque la misma temperatura puede ser poseída por objetos con diferentes estructuras microfísicas (Kim, 1972). Los objetos con diferentes estructuras microfísicas tal como objetos hechos de madera, de metal, de vidrio, etcétera, pueden tener muchas propiedades microfísicas interesantes en común, tal como una energía cinética molecular del mismo valor promedio. Más bien, el argumento funcionalista contra el fiscalismo es que es difícil ver cómo *podría haber* una propiedad física de primer orden no-trivial (ver la nota 4) en común con todas las realizaciones físicas posibles de un estado de máquina de Turing dado y sólo con ellas. ¡Trátese de pensar en un candidato remotamente plausible! Al menos, la prueba de cómo concebir uno recae en quienes piensan que tales propiedades físicas son concebibles.

Una manera de expresar este punto es que de acuerdo con el funcionalismo, el fiscalismo es una teoría *chauvinista*: niega propiedades mentales a sistemas que de hecho las tienen. Al decir, por ejemplo, que los estados mentales son estados cerebrales los fiscalistas excluyen injustamente a las pobres criaturas carentes de cerebro que, sin embargo, tienen mente.

Un segundo punto importante de este trabajo es que el argumento mismo que el funcionalismo usa para condenar al fiscalismo puede aplicarse con igual éxito contra el funcionalismo; ciertamente cualquier versión del funcionalismo que evite el liberalismo cae, como el fiscalismo, en el chauvinismo.

Este artículo tiene tres partes. La primera argumenta que el funcionalismo es culpable de liberalismo; la segunda, que una manera de modificar al funcionalismo para evitar el liberalismo es unirlo más firmemente a la psicología empírica, y la tercera, que ninguna versión del funcionalismo puede evitar tanto el liberalismo como el chauvinismo.

### 1.1 Algo más acerca de lo que el funcionalismo es

Una manera de ordenar la desconcertante variedad de teorías funcionalistas consiste en distinguir entre aquellas que se exponen en términos de una máquina de Turing y aquellas que no.

Una tabla de máquina de Turing lista un conjunto finito de estados de tabla-de-máquina,  $S_1 \dots S_n$ ; de *inputs*,  $I_1 \dots I_m$ ; y de *outputs*,  $O_1 \dots O_p$ . La tabla especifica un conjunto de condicionales de la forma: si la máquina está en el estado  $S_i$  y recibe el *input*  $I_j$ , emite el *output*  $O_k$  y pasa al estado  $S_l$ . Es decir, dado cualquier estado y cualquier *input*, la

tabla especifica un *output* y un estado siguiente. Cualquier sistema con un conjunto de *inputs*, *outputs* y estados relacionados de la manera especificada por la tabla, es descripto por la tabla y es una realización del autómata abstracto especificado por la tabla.

Para tener el poder de computar cualquier función recursiva, una máquina de Turing tiene que ser capaz de controlar su *input* de ciertas maneras. En las formulaciones estándar se considera que el *output* de una máquina de Turing tiene dos componentes. Imprime un símbolo sobre una cinta, luego corre la cinta y pone, así, un nuevo símbolo ante la vista del lector del *input*. Para que una máquina de Turing tenga poder completo, la cinta tiene que ser infinita, en al menos una dirección y corrible en ambas direcciones. Si la máquina no tiene control sobre la cinta, es un "transductor finito" ["finite transducer"], una máquina de Turing muy limitada. No es necesario considerar que los transductores finitos tengan una cinta. Quienes creen que el funcionalismo de máquina es verdadero tienen que suponer que la cuestión acerca del poder que tenemos como autómatas, es una cuestión empírica sustantiva. Si somos máquinas de Turing de "poder completo", el entorno [*environment*] tiene que constituir parte de la cinta...

Una versión muy simple del funcionalismo de máquina (Block y Fodor, 1972) sostiene que cada sistema que tiene estados mentales es descripto por al menos una tabla de máquina de Turing especificable, y que cada tipo de estado mental del sistema es idéntico a uno de los estados de la tabla-de-máquina [*machine table*]. Consideremos, por ejemplo, la máquina de Turing descripta en el siguiente cuadro (cf. Nelson, 1975):

	$S_1$	$S_2$
$S_1$	No emite <i>output</i> Pasa a $S_2$	Emite una gaseosa Pasa a $S_1$
$S_2$	Emite una gaseosa y una moneda de \$ 0,05 Permanece en $S_1$	
moneda de \$ 0,05 [nickel] <i>input</i>		
moneda de \$ 0,10 [dime] <i>input</i>		

Se puede contar con una descripción cruda de la versión simple del

funcionalismo de máquina si se considera la afirmación de que  $S_1$  = deseo-moneda de 0,10 [*nickel-desire*], y  $S_2$  = deseo-moneda de 0,05 [*dime-desire*]. Por supuesto que ningún funcionalista sostendría que una máquina de ese tipo desee algo. Más bien, la versión simple de funcionalismo de máquina descripta arriba formula un planteo análogo respecto de una hipotética tabla de máquina mucho más compleja. Adviértase que el funcionalismo de máquina especifica explícitamente a los *inputs* y *outputs* e implícitamente a los estados internos (Putnam [1967, pág. 434] afirma: "Para decirlo una vez más, los  $S_i$  se especifican sólo *implícitamente* por la descripción, es decir, se especifican *sólo* por el conjunto de probabilidades de transición dado en la tabla de máquina"). Un dispositivo tiene que aceptar monedas de \$ 0,05 y de \$ 0,10 como *inputs* y devolver monedas de \$ 0,05 y gaseosas como *outputs*, para ser descripto por esa tabla de máquina. Pero los estados  $S_1$  y  $S_2$  pueden ser virtualmente de cualquier naturaleza (aun de naturaleza no-física), en tanto que esa naturaleza conecte a los estados entre sí y con los *inputs* y *outputs* especificados en la tabla de máquina. Todo lo que se nos dice de  $S_1$  y  $S_2$  son esas relaciones; así, puede decirse que el funcionalismo de máquina reduce la mentalidad a estructuras *input-output*. Este ejemplo debería sugerir la fuerza del argumento funcionalista contra el fisicalismo. ¡Trátese de pensar en una propiedad física de primer orden (véase la nota 4) que pueda ser compartida por todas las realizaciones de esta tabla de máquina y sólo por ella!

También se puede caracterizar a los funcionalistas según que consideren a las identidades funcionales como parte de una psicología a priori o de una psicología empírica... Los funcionalistas a priori (por ejemplo, Smart, Armstrong, Lewis, Shoemaker), son los herederos de los conductistas lógicos. Tienden a considerar a los análisis funcionales como análisis de los significados de los términos mentales, mientras que los funcionalistas empíricos (por ejemplo, Fodor, Putnam, Harman) consideran a los análisis funcionales como hipótesis científicas substantivas. En lo que sigue, referiré al primer punto de vista como 'Funcionalismo' y al último como 'Psicofuncionalismo'. (Usaré 'funcionalismo', con 'f' minúscula, como neutral entre Funcionalismo y Psicofuncionalismo. Cuando distinga entre Funcionalismo y Psicofuncionalismo usaré siempre letras mayúsculas.)

El Funcionalismo y el Psicofuncionalismo y la diferencia entre ellos pueden clarificarse en términos de la noción de la oración Ramsey [*Ramsey sentence*] de una teoría psicológica. Los términos de estado-mental [*mental-state terms*] que aparecen en una teoría psicoló-

gica pueden definirse de varias maneras mediante la oración Ramsey de la teoría... Todas las teorías de la identidad de estado funcional [*functional state identity theories*] pueden entenderse como definiendo un conjunto de estados funcionales ... mediante la oración Ramsey de una teoría psicológica, correspondiendo un estado funcional a cada estado mental. El estado funcional que corresponde a dolor se llamará 'el correlato funcional Ramsey' [*Ramsey functional correlate*] de dolor, con respecto a la teoría psicológica. En términos de la noción de correlato funcional Ramsey de una teoría, la distinción entre Funcionalismo y Psicofuncionalismo puede definirse como sigue: el Funcionalismo identifica el estado mental S con el correlato funcional Ramsey de S, con respecto a una teoría psicológica de *sentido común*; el Psicofuncionalismo identifica S con el correlato funcional Ramsey de S con respecto a una teoría psicológica *científica*.

Esta diferencia entre Funcionalismo y Psicofuncionalismo da origen a una diferencia en la especificación de los *inputs* y *outputs*. Los Funcionalistas están limitados a especificar *inputs* y *outputs* que sean una parte plausible del conocimiento de sentido común [*common-sense knowledge*]; los Psicofuncionalistas no tienen tal limitación. Si bien ambos grupos insisten en la especificación física —o al menos no-mental— de *inputs* y de *outputs*, los Funcionalistas precisan clasificaciones externamente observables (tales como *inputs* caracterizados en términos de objetos presentes en la vecindad del organismo, *outputs* en términos de movimientos de partes del cuerpo). Los Psicofuncionalistas, en cambio, tienen la opción de especificar *inputs* y *outputs* en términos de parámetros internos tales como señales en las neuronas de *input* y de *output*...

Sea T una teoría psicológica de sentido común o bien de psicología científica [*psychological theory of either common-sense or scientific knowledge*]. T puede contener generalizaciones de la forma: quien quiera que esté en el estado w y reciba el *input* x emite el *output* y, y pasa al estado z. Escribamos T como

$$T(S_1 \dots S_n, I_1 \dots I_k, O_1 \dots O_m)$$

en donde las S son estados mentales, las I son *inputs* y las O son *outputs*. Las 'S' han de entenderse como *constantes* de estado mental, tal como 'dolor', no como variables; lo mismo vale para las 'I' y las 'O'. Así, uno podría también escribir T como

T (dolor..., luz de 400 nanómetros entrando por el ojo izquierdo..., dedo gordo del pie izquierdo se mueve 1 centímetro a la izquierda...)

Para obtener la oración Ramsey de T, reemplácese por variables los términos correspondientes a estados mentales —pero *no los términos correspondientes a inputs y outputs*—, y prefíjese un cuantificador existencial para cada variable.

$$\exists F_1 \dots \exists F_n T(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m)$$

Si 'F<sub>17</sub>' es la variable que reemplazó a la palabra 'dolor' cuando se formó la oración Ramsey, entonces podemos definir al dolor, en términos de la oración Ramsey, como sigue:

$$x \text{ tiene (siente) dolor } [x \text{ is in pain}] \leftrightarrow \exists F_1 \dots \exists F_n \\ T[(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m) \& x \text{ tiene } F_{17}]$$

El correlato funcional Ramsey de dolor es la propiedad expresada por el predicado del lado derecho de este bicondicional. Nótese que este predicado contiene constantes de *input* y de *output*, pero no constantes [de términos] mentales puesto que las constantes [de términos] mentales fueron reemplazadas por variables. El correlato funcional Ramsey de dolor es definido en términos de *inputs* y *outputs*, pero no en términos mentales.

Por ejemplo, sea T la teoría acerca de que el dolor es causado por daño en la piel y es causa de preocupación y de la proferencia de "ouch", y que la preocupación causa a su vez fruncir el entrecejo. Entonces la definición Ramsey sería:

x tiene (siente) dolor  $\leftrightarrow$  Hay 2 estados (propiedades), el primero de los cuales es causado por daño en la piel y causa la proferencia de "ouch" y del segundo estado, y el segundo estado causa fruncir el entrecejo, y x está en el primer estado.

El correlato funcional Ramsey de dolor con respecto a esta "teoría" es la propiedad de estar en un estado que es causado por daño en la piel y que causa la proferencia de "ouch" y otro estado que causa a su vez fruncir el entrecejo. (Nótese que las palabras 'dolor' y 'preocupación' han sido reemplazadas por variables, pero los términos de *input* y de *output* no.)

El correlato funcional Ramsey de un estado S es un estado que tiene mucho en común con S. Específicamente, S y su correlato funcional Ramsey comparten las propiedades *estructurales* especificadas por la teoría T. Pero, existen dos razones por las cuales es natural suponer que S y su correlato funcional Ramsey serán distintos. Primero, el correlato funcional Ramsey de S con respecto a T puede “incluir”, a lo sumo, aquellos aspectos de S que están relevados [*captured*] por T; los aspectos que no estén relevados por T, quedan afuera. Segundo, el correlato funcional Ramsey podría dejar también a un lado algo de lo que T releva porque la definición Ramsey no contiene el vocabulario “teórico” de T. La teoría tomada como ejemplo en el último párrafo es verdadera sólo de los organismos que sienten-dolor [*pain-feeling organisms*]; y lo es trivialmente, en virtud de su uso de la palabra ‘dolor’. Sin embargo, el predicado que expresa el correlato funcional Ramsey no contiene esa palabra (puesto que fue reemplazada por una variable), y así puede ser verdadera de cosas que no sienten dolor. Sería sencillo construir una máquina simple que tenga piel artificial, una ceja, una cinta con la grabación de “ouch” y dos estados que satisfagan las relaciones causales mencionadas, pero que no sienta dolor.

La hipótesis fuerte del funcionalismo es que para *alguna* teoría psicológica, la suposición natural de que un estado y su correlato funcional Ramsey sean distintos, es falsa. El Funcionalismo dice que existe una teoría tal que dolor, por ejemplo, es el correlato funcional Ramsey respecto de ella.

Un último punto preliminar: he dado la impresión equivocada de que el funcionalismo identifica *todos* los estados mentales con estados funcionales. Una versión tal del funcionalismo es obviamente demasiado fuerte. Sea X una réplica célula-por-célula de uno, recién creada (la cual, por supuesto, es funcionalmente equivalente a uno). Quizás uno recuerde haber celebrado su *bar-mitzva*. Pero X no recuerda haber celebrado su *bar-mitzva*, puesto que X nunca lo tuvo. Por cierto, algo puede ser funcionalmente equivalente a uno pero no saber lo que uno sabe, o [verbo], lo que uno [verbo], para una amplia variedad de verbos de “éxito” [“success”]. Peor aún, si Putnam (1975b) está en lo correcto al decir que “los significados no están en la cabeza”, sistemas funcionalmente equivalentes a uno pueden no tener, por razones similares, muchas de las demás actitudes proposicionales de uno. Supongamos que uno cree que el agua es húmeda. De acuerdo con ciertos argumentos plausibles presentados por Putnam y Kripke, una condición para la posibilidad de que uno crea que el agua es húmeda es un cierto tipo de cone-

xión causal entre uno y el agua. Nuestro “gemelo” en la Tierra Gemela que está conectado de una manera similar a XYZ pero no a H<sub>2</sub>O, no creería que el agua es húmeda.

Si el funcionalismo ha de ser defendido, tiene que ser interpretado como aplicándose solamente a una subclase de estados mentales: aquellos estados mentales “estrechos” [“narrow”] tal que las condiciones de verdad para su aplicación estén en algún sentido “dentro de la persona”. Pero aun suponiendo que una noción del carácter estrecho de los estados psicológicos pueda ser formulada satisfactoriamente, el interés del funcionalismo puede disminuir a causa de esa limitación. Menciono este problema sólo para dejarlo a un lado.

Consideraré al funcionalismo como una doctrina acerca de estados mentales “estrechos”.

### 1.2 Robots de cabeza homuncular [*homunculi-headed robots*]

En esta sección describiré una clase de estratagemas [*devices*] que, prima facie, ponen en un aprieto a todas las versiones del funcionalismo, dado que indican que el funcionalismo peca de liberalismo al clasificar sistemas que carecen de mentalidad, como teniendo mentalidad.

Consideremos la versión simple del funcionalismo de máquina ya descripto. Dice que cada sistema que tiene estados mentales es descripto por al menos una tabla de máquina de Turing de un cierto tipo y que cada estado mental del sistema es idéntico a uno de los estados de tabla-de-máquina especificados por la tabla de máquina. Consideraré que los *inputs* y los *outputs* son especificados mediante descripciones de impulsos neurales en los órganos sensoriales y mediante neuronas de *output* motor. No debe considerarse que lo que se va a decir vale para el Psicofuncionalismo y no para el Funcionalismo. Tal como señalé, toda versión del funcionalismo supone *alguna* especificación de *inputs* y de *outputs*. Una especificación Funcionalista serviría lo mismo para nuestros fines.

Imaginemos un cuerpo externamente similar a un cuerpo humano, digamos como el de uno, pero internamente muy diferente. Las neuronas asociadas a los órganos sensoriales se conectan con una hilera de luces ubicada en una cavidad vacía de la cabeza. Un conjunto de botones está conectado con las neuronas de *output* motoras. Dentro de la cavidad reside un grupo de hombrecitos. Cada uno tiene una tarea muy sencilla: implementa una “casilla” [“square”] de una tabla de máquina adecuada que lo describe a uno. Sobre una pared hay una cartelera en

la que está colocada una tarjeta de estado [*state card*]; es decir, una tarjeta que tiene un símbolo que designa a uno de los estados especificados en la tabla de máquina. He aquí lo que los hombrecitos hacen. Supongamos que la tarjeta tiene una 'G'. Esto alerta al hombrecito que implementa los casilleros G. Los hombrecitos se autodenominan 'hombres-G'. Supongamos que la luz que representa al *input*  $I_{17}$  está encendida. Uno de los hombres-G sólo tiene la siguiente tarea: cuando la tarjeta dice 'G' y la luz  $I_{17}$  está encendida, él presiona el botón de *output*  $O_{191}$  y cambia la tarjeta de estado a 'M'. Este hombre-G es llamado a ejercitarse su tarea sólo en raras ocasiones. A pesar del bajo nivel de inteligencia que se requiere de cada hombrecito, el sistema, como un todo, se las arregla para simularlo a uno, porque la organización funcional para cuya realización se los entrenó, es la de uno. Una máquina de Turing puede ser representada como un conjunto finito de cuádruplas (o quíntuplas, si el *output* es dividido en dos partes): estado actual, *input* actual, estado próximo y *output* próximo. Cada hombrecito tiene una tarea que corresponde a una única cuádrupla. A través de los esfuerzos de los hombrecitos el sistema realiza la misma (razonablemente adecuada) tabla de máquina que uno y, de tal modo, es funcionalmente equivalente a uno.<sup>5</sup>

Describiré una versión de la simulación de cabeza homuncular, que tiene más probabilidades de ser nomológicamente posible. ¿Cuántos homúnculos se requieren? Quizás un par de miles de millones sea suficiente.

Supongamos que convertimos al gobierno de China al funcionalismo y convencemos a sus funcionarios... para que realicen una mente humana durante una hora. Proporcionamos a cada una de los miles de millones de personas de China (elijo a China porque tiene un par de miles de millones de habitantes) un radiotransmisor de doble canal especialmente diseñado, que las conecta de manera apropiada con otras personas y con el cuerpo artificial mencionado en el ejemplo anterior. Reemplazamos a cada uno de los hombrecitos por un ciudadano chino y su radio-transmisor. En vez de una cartelera tenemos letras desplegadas en una serie de satélites ubicados de modo tal que puedan ser vistos desde cualquier lugar de China.

5. La idea básica de este ejemplo proviene de Putnam (1967). Estoy en deuda con Harry Field por las conversaciones mantenidas sobre este tema. El intento de Putnam de evitar al funcionalismo el problema planteado por tales ejemplos es discutido en la sección 1.3 de este trabajo.

El sistema de un par de miles de millones de personas que se comunican entre sí más los satélites, desempeña el rol de un "cerebro" externo conectado a un cuerpo artificial mediante radiotransmisión. No hay nada absurdo acerca de una persona conectada a su cerebro mediante radiotransmisión. Llegará el día, quizás, en que nuestros cerebros sean periódicamente retirados para limpieza y reparación. Imaginemos que esto se hace primero tratando a las neuronas que acoplan al cerebro con el cuerpo con una sustancia química que les permita estirarse como bandas de goma, asegurando con ello que ninguna de las conexiones cuerpo-cerebro sea interrumpida. Muy pronto, hombres de negocio inteligentes descubren que pueden atraer más clientes reemplazando las neuronas estiradas por nexos de radiotransmisión, de manera que los cerebros puedan ser limpiados sin incomodar al cliente al tener que inmovilizar su cuerpo.

No es nada obvio que el sistema corporal chino sea físicamente imposible. Podría ser funcionalmente equivalente a uno por un tiempo breve, digamos una hora.

"Pero —alguien puede objetar— ¿cómo podría algo ser funcionalmente equivalente a mí por una hora? ¿No determina mi organización funcional, digamos, cómo reaccionaría si durante una semana lo único que hiciera fuera leer el *Reader's Digest*?" Recordemos que una tabla de máquina específica un conjunto de condicionales de la forma: si la máquina está en  $S_i$  y recibe el *input*  $I_j$ , emite el *output*  $O_k$  y pasa a  $S_l$ . Estos condicionales tienen que entenderse *subjuntivamente*. Lo que le da a un sistema una organización funcional en un momento dado no es lo que *hace* en ese momento sino también los contrafácticos que son verdaderos de él en ese momento: lo que *hubiera* hecho (y lo que hubieran sido sus transiciones de estado) de haber tenido un *input* diferente o de haber estado en un estado diferente. Si es verdad de un sistema, en el tiempo  $t$ , que *obedecería* a una tabla de máquina dada sin importar en cuál de los estados esté y sin importar cuál de los *inputs* reciba, entonces el sistema es descripto, en  $t$ , mediante la tabla de máquina (y realiza en  $t$  al autómata abstracto especificado por la tabla), aun si existiera sólo por un instante. Durante la hora en la cual el sistema chino está "en funcionamiento" *tiene* un conjunto de *inputs*, *outputs* y estados de los cuales tales condicionales subjuntivos son verdaderos. Esto es lo que hace que cualquier computador realice el autómata abstracto que realiza.

Hay señales, por supuesto, a las que el sistema respondería y a las que uno no respondería, por ejemplo, a una interferencia masiva en la

radiotransmisión o a una inundación del río Yangtze. Tales eventos podrían causar un mal funcionamiento frustrando la simulación, como una bomba [*bomb*] en un computador puede hacer que el computador no realice la tabla de máquina para cuya realización fue construido. Pero así como el computador *sin* la bomba *puede* realizar la tabla de máquina, el sistema que se compone de personas y cuerpo artificial puede realizar la tabla de máquina en tanto no haya catástrofes que interfieran, tales como inundaciones, etcétera.

“Pero —alguien puede objetar— existe una diferencia entre una bomba en un computador y una bomba en el sistema chino, porque en el caso de este último (a diferencia del primero), los *inputs* especificados en la tabla de máquina pueden ser la causa del mal funcionamiento. La actividad neural inusual en los órganos sensoriales de los residentes de la provincia de Chungking ocasionada por una bomba o por una inundación del Yangtze, puede causar que el sistema se desordene.”

Respuesta: la persona que dice a qué sistema se refiere tiene que decir qué señales valen como *inputs* y *outputs*. Yo tomo como *inputs* y como *outputs* sólo a la actividad neural en el cuerpo artificial conectado mediante radiotransmisión con los habitantes de China. Las señales neurales de los habitantes de Chungking cuentan tan poco como *input* de ese sistema, como la cinta de *input* atascada por un saboteador entre los contactos de relé en las entrañas de una computadora, cuenta como *input* de esa computadora.

Por supuesto, el objeto que se compone de los habitantes de China + el cuerpo artificial, tiene *otras* descripciones de máquina de Turing bajo las cuales las señales neurales en los habitantes de Chungking *contarían* como *inputs*. Ese nuevo sistema (esto es, el objeto de esa nueva descripción de máquina de Turing) no sería funcionalmente equivalente a uno. De modo similar, cualquier computador comercial puede ser redescrito de manera que permita que la cinta atascada en su interior cuente como *input*. Al describir un objeto como una máquina de Turing, uno traza una línea entre dentro y fuera. (Si sólo consideramos a los impulsos neurales como *inputs* y *outputs*, trazamos esa línea dentro del cuerpo; si sólo consideramos a las estimulaciones periféricas como *inputs*, ...trazamos esa línea en la piel.) Al describir al sistema chino como una máquina de Turing, he trazado la línea de tal manera que satisface un cierto tipo de descripción funcional, una [descripción] que *también* uno satisface y que, de acuerdo con el funcionalismo, justifica adscripciones de mentalidad. El Funcionalismo no sostiene que todo sistema mental tenga una tabla de máquina de un tipo tal que justifique

adscripciones de mentalidad con respecto a *toda* especificación de *inputs* y de *outputs*, sino más bien, sólo con respecto a *alguna* especificación.

Objeción: el sistema chino trabajaría demasiado lentamente. El tipo de eventos y procesos con los que tenemos contacto normalmente, ocurrirían demasiado rápido para que el sistema los detectase. Así, no estaríamos en condiciones de conversar con él, jugar bridge con él, etcétera.

Respuesta: resulta difícil ver por qué la escala temporal del sistema debe importar... ¿Es realmente contradictorio o sin sentido suponer que podríamos encontrar una raza de seres inteligentes con los cuales podríamos comunicarnos sólo a través de dispositivos tales como una cámara lenta [*time-lapse photography*]? Cuando observamos a esas criaturas, parecen casi inanimadas. Pero cuando vemos las películas en cámara lenta, las vemos conversando entre sí. Por cierto, encontramos que dicen que la única manera en que ellas pueden entendernos es viendo las películas en cámara lenta. Considerar a la escala temporal como lo más importante parece crudamente conductista...

Lo que hace del sistema con cabeza-homuncular recién descrito (considérese a los dos sistemas como variantes de un único sistema) un contraejemplo posible del funcionalismo (de máquina), es que existe la duda, prima facie, de que tenga estados mentales, especialmente de que tenga lo que los filósofos han llamado, de diversos modos, “estados cualitativos”, “vivencias puras” [*“raw feels”*] o “cualidades fenomenológicas inmediatas”. (Alguien pregunta: ¿qué es lo que los filósofos han llamado estados cualitativos? Yo respondo bromeando sólo a medias: como dijo Louis Armstrong cuando le preguntaron qué es el *jazz*, “Si usted me lo tiene que preguntar, nunca podrá llegar a saberlo”.) En términos de Nagel (1974), existe la duda, prima facie, de que haya algo que sea cómo ser el sistema con cabeza-homuncular.<sup>6</sup>

### 1.3 La propuesta de Putnam

Una manera en que los funcionalistas pueden tratar de encarar el problema planteado por los contraejemplos que recurren a cabezas-homunculares, es apelando al recurso ad hoc de no darles cabida. Por ejemplo, un funcionalista podría estipular que dos sistemas no

6. Shoemaker (1975) argumenta (en respuesta a Block y Fodor, 1972) que los *qualia* ausentes son lógicamente imposibles; esto es, que es lógicamente imposible que dos sistemas estén en el mismo estado funcional y que, sin embargo, uno tenga un contenido cualitativo y el otro carezca de él.

pueden ser funcionalmente equivalentes si uno contiene partes con organizaciones funcionales características de los seres sintientes [*sentient beings*] y el otro no. En el artículo en que hipotetiza que el dolor es un estado funcional, Putnam estipula que “ningún organismo capaz de sentir dolor es susceptible de ser descompuesto en partes que separadamente posean Descripciones” (como el tipo de máquina de Turing que puede estar en el estado funcional que Putnam identifica con dolor). El propósito de esta condición es “excluir ‘organismos’ (si es que valen como tales) como los enjambres de abejas en tanto que experimentadores singulares de dolor” (Putnam, 1967, págs. 434-5).

Una manera de satisfacer el requisito de Putnam sería ésta: un organismo que es capaz de sentir-dolor no es susceptible de ser descompuesto en partes, *todas* las cuales tengan una organización funcional característica de los seres-sintientes. Pero esto no excluye mi ejemplo que apela a la cabeza-homuncular, dado que tiene partes no sintientes, tales como el cuerpo mecánico y los órganos sensoriales. No servirá irse al extremo opuesto y requerir que *ninguna* parte propia sea sintiente. De otro modo, las mujeres embarazadas y las personas con parásitos sintientes [*sentient parasites*] no podrían contar como organismos capaces de sentir dolor. Lo que parece ser importante para ejemplos como la simulación de cabeza-homuncular que he descripto, es que los seres sintientes *desempeñan un rol crucial* en dar a las cosas su organización funcional. Esto sugiere una versión de la propuesta de Putnam que requiere que un organismo capaz de sentir dolor tenga una cierta organización funcional y no tenga partes que (1) posean ellas mismas ese tipo de organización funcional y además (2) desempeñen un rol crucial en dar al sistema total su organización funcional.

Aunque esta propuesta involucra la noción vaga de “rol crucial”, es lo suficientemente precisa para hacernos ver que no funcionará. Supongamos que existe una parte del universo que contiene una materia completamente diferente de la nuestra, una materia que es infinitamente divisible. En esa parte del universo hay criaturas inteligentes de muchos tamaños, incluso criaturas semejantes a los humanos pero mucho más pequeñas que nuestras partículas elementales. En una expedición intergaláctica esa gente descubre la existencia de nuestro tipo de materia. Por razones que ellos sólo conocen deciden dedicar los próximos cientos de años a producir, partiendo de *su* materia, sustancias con las características químicas y físicas (excepto en el nivel de partículas subelementales) de *nuestros* elementos. Construyen hordas de naves espaciales de diferentes variedades remedando el tamaño aproximado de nuestros

electrones, protones y otras partículas elementales, y pilotean las naves de manera de imitar el comportamiento de esas partículas elementales. Además, las naves contienen generadores para producir el tipo de radiación que producen las partículas elementales. Cada nave tiene un equipo de expertos en la naturaleza de nuestras partículas elementales. Hacen esto para producir inmensas (de acuerdo con nuestros estándares) masas de sustancias con las características químicas y físicas del oxígeno, el carbono, etcétera. Poco tiempo después de que han logrado su objetivo, uno sale de expedición a esa parte del universo y descubre el “oxígeno”, el “carbono”, etcétera. Ignorante de su verdadera naturaleza, uno establece una colonia, y usa esos “elementos” para cultivar plantas alimenticias, proporcionar “aire” para respirar, etcétera. Dado que las moléculas de uno son intercambiadas constantemente con el entorno, uno y los demás colonizadores (en un período de pocos años) llegamos a estar compuestos principalmente de la “materia” hecha de esa gente diminuta en sus naves espaciales. ¿Sería uno menos capaz de sentir dolor, de pensar, etcétera, sólo, porque la materia de la que está compuesto (y de la que dependen sus características) contiene seres que, en sí mismos, tienen una organización funcional típica de criaturas sintientes? Creo que no. Los mecanismos electroquímicos básicos mediante los cuales se lleva a cabo la sinapsis son ahora bastante bien comprendidos. Como se sabe, los cambios que no afectan a esos mecanismos electroquímicos no afectan al funcionamiento del cerebro y no afectan a la mentalidad. Los mecanismos electroquímicos en nuestras sinapsis no serían afectados por el cambio en nuestra materia.<sup>7</sup>

Resulta interesante comparar el ejemplo de la gente-hecha-de-partícula-elemental con los ejemplos del comienzo de capítulo que apelan a la cabeza-homuncular. Una conjectura natural acerca de la fuente de nuestra intuición de que las simulaciones descriptas inicialmente que apelan a la cabeza-homuncular carecen de mentalidad, es que tienen *demasiada* estructura mental interna. Los hombrecitos podrían a veces aburrirse, a veces excitarse. Podemos imaginar aun que deliberan acerca de la mejor manera de realizar la organización funcional dada y que hacen cambios con la intención de gozar de más tiempo libre. Pero el

7. Dado que hay una diferencia entre el rol de los hombrecitos al producir su organización funcional en la situación descripta y el rol de los homúnculos en las simulaciones que apelan a la-cabeza-homuncular, con que se inicia este trabajo, cabe presumir que la condición de Putnam podría ser reformulada de modo de excluir a los segundos sin excluir a los primeros. Pero esto sería una maniobra muy ad hoc.

ejemplo de la gente-hecha-de-partícula-elemental recién descripto, sugiere que esta primera conjetura es errónea. Lo que parece importante es *cómo* la mentalidad de las partes contribuye al funcionamiento del todo.

Hay una diferencia muy notable entre el ejemplo de la gente-hecha-de-partícula-elemental y los anteriores ejemplos de homúnculos. En el primero, el cambio que se produce en uno a medida que nos vamos infectando de homúnculos no es un cambio que produzca ninguna diferencia en nuestro procesamiento psicológico (es decir, procesamiento de información) o en nuestro procesamiento neurológico, sino sólo en nuestra microfísica. Ninguna de las técnicas propias de la psicología o de la neurofisiología humanas revelaría diferencia alguna en uno. Sin embargo, las simulaciones que apelan a la cabeza-homuncular, descriptas al comienzo del trabajo, no son cosas a las que se apliquen las teorías neurofisiológicas que son verdaderas de nosotros, y si son interpretadas como simulaciones *Funcionales* (más que como Psicofuncionales) no necesitan ser cosas a las que se apliquen las teorías psicológicas (procesamiento de información) verdaderas de nosotros. Esta diferencia sugiere que nuestras intuiciones están, en parte, controladas por el punto de vista, no del todo razonable, de que nuestros estados mentales dependen de que tengamos la psicología y/o la neurofisiología que tenemos. Así, algo que difiera marcadamente de nosotros en ambos aspectos (recuérdese que se trata de una simulación Funcional más que Psicofuncional) no debe suponerse que tenga mentalidad, sólo sobre la base de que ha sido diseñado para ser Funcionalmente equivalente a nosotros.

#### 1.4 ¿Es la duda prima facie meramente prima facie?

El Argumento de los *Qualia* Ausentes [*Absent Qualia Argument*] descansó en una apelación a la intuición de que las simulaciones que apelan a la cabeza-homuncular carecían de mentalidad, o al menos, de *qualia*. He dicho que esta intuición dio origen a la duda, prima facie, de que el funcionalismo sea verdadero. Pero las intuiciones que no se apoyan en argumentos fundados [*principled*] difícilmente han de ser consideradas sólidas. Ciertamente, las intuiciones incompatibles con una teoría bien fundada, tal como la intuición precopernicana de que la Tierra no se mueve, felizmente desaparecen pronto. Aun en ámbitos como el de la lingüística, cuyos datos consisten principalmente de intuiciones,

a menudo se rechazan intuiciones tales como que las siguientes oraciones son no-gramaticales (sobre bases teóricas):

El caballo corrido pasó el establo cayó.

El muchacho la chica el gato mordió rasguñó murió.

Estas oraciones son, de hecho, gramaticales, aunque difíciles de procesar.<sup>8</sup>

Apelar a las intuiciones cuando se juzga la posesión de mentalidad es, sin embargo, especialmente sospechoso. *Ningún* mecanismo físico parece intuitivamente plausible como asiento de los *qualia*, y mucho menos un *cerebro*. ¿Es una gudeja de tembloroso material [*stuff*] gris más intuitivamente apropiada como asiento de los *qualia* que un grupo de hombrecitos? Si no lo es, quizás haya también una duda prima facie, acerca de los *qualia* de los sistemas con cabeza-cerebral [*brain-headed*].

Sin embargo, existe una diferencia muy importante entre los sistemas con cabeza-cerebral y los sistemas con cabeza-homuncular. Dado que sabemos que *nosotros somos sistemas con cabeza-cerebral* y que tenemos *qualia*, sabemos que los sistemas con cabeza-cerebral pueden tener *qualia*. Así, aunque carecemos de una teoría de los *qualia* que explique cómo es ello *possible*, tenemos una razón contundente para desechar toda duda prima facie que haya acerca de los *qualia* de los sistemas con cabeza-cerebral. Por supuesto que esto hace a mi argumento parcialmente *empírico*: depende del conocimiento que nos marca [*makes us tick*]. Pero dado que este es un conocimiento que de hecho poseemos, depender de tal conocimiento no debería ser considerado un defecto.<sup>9</sup>

Existe otra diferencia entre nuestras cabezas-de-carne-y-hueso y las cabezas-homunculares: éstos son sistemas diseñados para imitarnos,

8. Compárese la primera oración con 'El pescado comido en Boston apesta'. La razón de que sea difícil de procesar es que 'corrido' se lee de manera natural como activo más que como pasivo. Véase Fodor *et al.*, 1974, pág. 360. Para una discusión de por qué la segunda oración es gramatical, véase Fodor y Garrett, 1967; Bever, 1970, y Fodor *et al.*, 1974.

9. A menudo no podemos concebir cómo algo es posible porque carecemos de los conceptos teóricos relevantes. Por ejemplo, antes del descubrimiento de los mecanismos de duplicación genética, Haldane argumentó persuasivamente que ningún mecanismo físico concebible podría hacer ese trabajo. Estuvo en lo correcto. Pero en vez de argumentar que los científicos deberían desarrollar ideas que nos permitiesen concebir un mecanismo físico tal, concluyó que un mecanismo *no-físico* estaba involucrado. (Debo este ejemplo a Richard Boyd.)

pero nosotros no estamos diseñados para imitar nada (aquí me apoyo en otro hecho empírico). Este hecho cancela cualquier intento de argumentar sobre la base de una inferencia a la mejor explicación a favor de los *qualia* de cabezas-homunculares. La mejor explicación de los gritos y muecas de las cabezas-homunculares no son sus dolores, sino que fueron diseñadas para imitar nuestros gritos y muecas.

Algunas personas parecen sentir que la conducta compleja y sutil de las cabezas-homunculares (conducta tan compleja y sutil, aun tan "sensitiva" a los rasgos del entorno, humano y no-humano, como nuestra conducta) es por sí misma una razón suficiente para desuchar la duda, prima facie, de que las cabezas-homunculares tengan *qualia*. Pero esto es crudo conductismo...

Mi argumento contra el Funcionalismo depende del siguiente principio: si una doctrina tiene una conclusión absurda para creer en la cual no hay una razón independiente, y si no hay manera de salvar el absurdo o de mostrar que es engañoso o irrelevante, y si no hay una buena razón para creer en la doctrina que lleva directamente al absurdo, entonces no se acepte la doctrina. Sostengo que no hay una razón independiente para creer en la mentalidad de una cabeza-homuncular, y sé que no hay manera de salvar el absurdo de la conclusión de que tiene mentalidad (aunque por supuesto mi argumento es vulnerable a la introducción de tal explicación). La cuestión, entonces, es si hay alguna buena razón para creer en el Funcionalismo. Un argumento a favor del Funcionalismo es que es la mejor solución disponible para el problema mente-cuerpo. Creo que éste es un mal argumento, pero puesto que también creo que el Psicofuncionalismo es preferible al Funcionalismo (por razones que mencionaré), pospondré la consideración de esta forma de argumentar hasta la discusión del Psicofuncionalismo.

El otro argumento que conozco a favor del Funcionalismo es que puede mostrarse que las identidades Funcionales son verdaderas sobre la base de los análisis de los significados de la terminología mental. De acuerdo con este argumento, las identidades Funcionales tienen que ser justificadas de la misma manera en que uno podría tratar de justificar la afirmación de que el estado de ser soltero es idéntico al estado de ser un hombre no casado. Un argumento similar apela a las trivialidades del sentido común acerca de los estados mentales en lugar de apelar a verdades acerca del significado. Lewis dice que las caracterizaciones funcionales de los estados mentales pertenecen al ámbito de la "psicología de sentido común, a la ciencia *folk*, más que a la ciencia profesional" (Lewis, 1972, pág. 250). (Véase también Shoemaker, 1975 y Armstrong,

1968. Armstrong tergiversa la cuestión de la analiticidad. Véase Armstrong, 1968, págs. 84-5, y pág. 90.) Y luego insiste en que las caracterizaciones Funcionales "deberían incluir sólo trivialidades que entre nosotros constituyen conocimiento común: todos las conocen, todos saben que todos las conocen, y así sucesivamente" (Lewis, 1972, pág. 256). Me referiré fundamentalmente a la versión "trivial" del argumento. La versión de la analiticidad es vulnerable a las mismas consideraciones, así como a dudas quineanas acerca de la analiticidad...

Estoy dispuesto a conceder, a los efectos del argumento, que es posible definir cualquier término de estado mental según las trivialidades concernientes a otros términos de estado mental, a términos de *input* y a términos de *output*. Pero esto no me compromete con el tipo de definición de términos de estado mental en la cual toda la terminología mental ha sido eliminada *via* la Ramsificación o algún otro mecanismo. Es simplemente falaz suponer que si cada término mental es definible en términos de los otros (más *inputs* y *outputs*), entonces cada término mental es definible no-mentalísticamente. Para ver esto, consideremos el ejemplo dado con anterioridad. Simplifiquemos la cuestión, claro, ignorando los *inputs* y los *outputs*. Definamos dolor como la causa de molestia y molestia como el efecto del dolor. Quien estuviera tan equivocado como para aceptar esto, no precisa aceptar una definición de dolor como *la causa de algo*, o una definición de molestia como *el efecto de algo*. Lewis sostiene que es analítico que dolor sea el ocupante de un cierto rol causal. Aun si estuviera en lo correcto acerca de un rol causal, especificado en parte mentalísticamente, uno no puede concluir que es analítico que dolor sea el ocupante de cualquier rol causal, especificado no-mentalísticamente.

No veo ningún argumento razonable a favor del Funcionalismo que se base en trivialidades o en la analiticidad. Además, la concepción que basa el Funcionalismo en trivialidades conduce a dificultades en los casos en que las trivialidades no tienen nada que decir. Recuérdese el ejemplo de los cerebros que son removidos para limpiarlos y rejuvenecerlos, en el que las conexiones entre nuestro cerebro y nuestro cuerpo se mantienen mediante radiotransmisión mientras continuamos con nuestra vida habitual. El proceso lleva unos pocos días y cuando se completa, el cerebro es reinsertado en el cuerpo. Ocasionalmente puede ocurrir que el cuerpo de una persona se destruya a causa de un accidente mientras el cerebro es limpiado y rejuvenecido. Si estuviera conectado a órganos sensoriales de *input* (pero no a órganos de *output*) tal cerebro no exhibiría *ninguna* de las conexiones usuales de carácter tri-

vial entre la conducta y los conjuntos de *inputs* y de estados mentales. Si como parece plausible, tal cerebro pudiera tener casi los mismos estados mentales (en sentido estrecho) que nosotros tenemos (y dado que tal estado de cosas podría volverse típico), el Funcionalismo estaría equivocado.

Resulta instructivo comparar la manera en que el Psicofuncionalismo intenta lidiar con los cerebros en cubetas. De acuerdo con el Psicofuncionalismo, es una cuestión empírica lo que va a valer como *inputs* y *outputs* de un sistema. Considerar a los impulsos neurales como *inputs* y *outputs* evitaría los problemas esquematizados, puesto que los cerebros en cubetas y los paralíticos podrían tener los impulsos neurales correctos aun sin tener movimientos corporales. Objección: podría darse una parálisis que afecte el sistema nervioso, y afecte, de este modo, a los impulsos neurales, así el problema que se le plantea al Funcionalismo se le plantea también al Psicofuncionalismo. Respuesta: las enfermedades del sistema nervioso pueden, en efecto, *cambiar la mentalidad*, por ejemplo pueden hacer que los pacientes sean incapaces de sentir dolor. De este modo, podría ser verdad que una enfermedad del sistema nervioso ampliamente extendida que causa parálisis intermitente, hiciera a la gente incapaz de tener ciertos estados mentales.

De acuerdo con las versiones plausibles del Psicofuncionalismo, la tarea de decidir qué procesos neurales contarían como *inputs* y como *outputs* es, en parte, una cuestión de decidir qué *disfunciones cuentan como cambios en la mentalidad* y qué *disfunciones cuentan como cambios en las conexiones de input y de output periféricas*. El Psicofuncionalismo cuenta con un recurso que el Funcionalismo no tiene, puesto que el Psicofuncionalismo nos permite *corregir la línea que trazamos entre dentro y fuera del organismo, de modo de evitar problemas del tipo que hemos discutido*. Todas las versiones del Funcionalismo yerran al intentar trazar esta línea sólo sobre la base del conocimiento del sentido común; las versiones "analíticas" del Funcionalismo yerran especialmente al intentar trazar la línea a priori.

## 2. Psicofuncionalismo

Al criticar el Funcionalismo apelé al siguiente principio: si una doctrina tiene una conclusión absurda para creer en la cual no hay una razón independiente, y si no hay manera de salvar el absurdo o de mostrar que es engañoso o irrelevante, y si no hay una buena razón para

creer en la doctrina que lleva directamente al absurdo, entonces no se acepte la doctrina. Dije que no había ninguna razón independiente para creer que la simulación funcional que apela a la cabeza-homuncular, tiene estados mentales. Sin embargo, *hay* una razón independiente para creer que la simulación *Psicofuncional* que apela a la cabeza-homuncular tiene estados mentales, es decir, que una simulación Psicofuncional de uno sería Psicofuncionalmente equivalente a uno, de modo que toda teoría psicológica verdadera de uno sería también verdadera de la simulación. ¿Qué mejor razón podría haber para atribuirle estados mentales, cualesquiera que sean ellos, que estén dentro del dominio de la psicología?

Este punto muestra que cualquier simulación Psicofuncional de uno comparte nuestros estados mentales *no-cualitativos*. Sin embargo, en la próxima sección argumentaré que hay, no obstante, algunas dudas de que comparta nuestros estados mentales *cualitativos*.

### 2.1 ¿Son los *qualia* estados Psicofuncionales?

Comencé este artículo describiendo un dispositivo de cabeza-homuncular y sosteniendo que hay dudas, prima facie, acerca de que tenga estados mentales, especialmente de que tenga estados mentales *cualitativos*, como dolores, picazones y sensaciones de rojo. La duda especial acerca de los *qualia* puede ser explicada, quizás, pensando en los *qualia invertidos* más que en los *qualia ausentes*. Tiene sentido, o parece tenerlo, suponer que los objetos que dos personas llaman verdes, lucen a una de ellas de la manera en que lucen los objetos que ambas llaman rojos. Parece que podríamos ser funcionalmente equivalentes aun cuando la sensación que las fresas evocan en uno sea cualitativamente la misma que la sensación que el césped evoca en la otra. Imagínese una lente invertida que cuando se coloca en el ojo de un sujeto produce exclamaciones como "Las cosas rojas lucen ahora de la manera en que las cosas verdes acostumbraban lucir, y viceversa". Imagínese además, a un par de gemelos idénticos, a uno de los cuales se le han insertado las lentes al nacer. Los gemelos crecen normalmente, y a la edad de 21 años son funcionalmente equivalentes. Esta situación ofrece, al menos, alguna evidencia de que el espectro de cada uno está invertido con relación al del otro (véase Shoemaker, 1975, nota 17, para una descripción convincente de la inversión intrapersonal del especíro). Sin embargo, resulta difícil ver cómo dar sentido al análogo de la inversión del espectro con respecto a estados no-cualitativos. Imagínese un par de

personas, una de las cuales cree que *p* es verdadera y que *q* es falsa, mientras que la otra cree que *q* es verdadera y que *p* es falsa. ¿Podrían esas personas ser funcionalmente equivalentes? Resulta difícil ver cómo podrían serlo.<sup>10</sup> Ciertamente, resulta difícil ver cómo dos personas podrían tener sólo esa diferencia en las creencias y sin embargo que no existiese ninguna circunstancia posible en la cual esa diferencia en la creencia se revelara por sí misma en conductas diferentes. Los *qualia* parecen ser supervenientes [*supervenient*] a la organización funcional, de una manera como las creencias no lo son...

Existe otra razón para distinguir firmemente entre estados mentales cualitativos y no-cualitativos cuando hablamos de teorías funcionalistas:

10. Supongamos que un hombre que tiene una buena visión de los colores utiliza erróneamente 'rojo' para denotar verde y 'verde' para denotar rojo. Es decir que confunde las dos palabras. Dado que esta confusión es puramente lingüística, aunque diga de una cosa verde que es roja, no *cree* que sea roja, así como un extranjero que ha confundido 'estuche' con 'sandwich' no cree que la gente come estuches en el almuerzo. Digamos que la persona que ha confundido de esta manera 'rojo' y 'verde', es una víctima de Cambio de Palabras [*Switching Word*].

Considérese ahora una enfermedad diferente: tener lentes que invierten rojo/verde ubicadas en los ojos, sin saberlo. Digamos que una víctima de esta enfermedad es una víctima de Cambio de Estímulo [*Stimulus Switching*]. Como la víctima de Cambio de Palabra, la víctima de Cambio de Estímulo aplica 'rojo' a las cosas verdes y viceversa. Pero la víctima de Cambio de Estímulo *tiene* creencias falsas acerca del color. Si se le muestra una mancha verde dice y *cree* que es roja.

Supongamos ahora que una víctima de Cambio de Estímulo de pronto se vuelve también una víctima de Cambio de Palabra (supongamos además que es un residente nativo de una villa remota del Ártico y que no posee creencias respecto de que el pasto sea verde, las fresas sean rojas, etcétera). Habla normalmente, aplicando 'verde' a las manchas verdes y 'rojo' a las manchas rojas. Por cierto, es funcionalmente normal. Pero sus *creencias* son tan anormales como eran antes de que se tornara una víctima de Cambio de Palabra. Antes de confundir las palabras 'rojo' y 'verde', aplicaba 'rojo' a una mancha verde, y erróneamente creía que la mancha era roja. Ahora (correctamente) dice 'rojo', pero su creencia sigue siendo errónea.

Así, dos personas pueden ser funcionalmente las mismas, aunque tengan creencias incompatibles. En consecuencia, el problema de los *qualia* invertidos infecta tanto a las creencias como a los *qualia* (aunque, presumiblemente, sólo a las creencias cualitativas). Este hecho debe interesar no sólo a quienes sostienen teorías de identidad de estados funcionales referidas a creencias, sino también a quienes se sienten atraídos por las explicaciones al estilo de Harman acerca del significado como rol funcional. Nuestra doble víctima —de Cambio de Palabra y de Cambio de Estímulo— es un contraejemplo para tales explicaciones. Porque su palabra 'verde' juega el rol normal en su razonamiento e inferencia, pero dado que al decir de algo que "es verde" expresa su creencia de que es *rojo*, usa 'verde' con un significado anormal. Estoy en deuda con Sylvain Bromberger por la discusión de esta cuestión.

el Psicofuncionalismo evita los problemas que el Funcionalismo tiene con los estados no-cualitativos, por ejemplo, las actitudes proposicionales como creencias y deseos. Pero el Psicofuncionalismo puede ser tan poco capaz de lidiar con los estados cualitativos, como lo es el Funcionalismo. La razón es que los *qualia* pueden muy bien no caer en el dominio de la psicología.

Para ver esto, permítasenos tratar de imaginar cómo sería una realización de la psicología humana que apelara a la cabeza-homuncular. La teorización psicológica corriente parece estar dirigida a la descripción de las relaciones del flujo-de-información [*information-flow*] entre mecanismos psicológicos. El objetivo principal parece consistir en descomponer tales mecanismos en mecanismos psicológicos primitivos, "cajas negras", cuya estructura interna cae en el dominio de la fisiología más que en el dominio de la psicología. (Véanse Fodor, 1968; Dennett, 1975 y Cummins, 1975; se plantean objeciones interesantes en Nagel, 1969.) Por ejemplo, un mecanismo quasi-primitivo podría aparear dos ítems en un sistema representacional y determinar si son casos del mismo tipo. O los mecanismos primitivos podrían ser como los de un computador digital, por ejemplo podrían ser (a) *agregue 1 a un registro dado*, y (b) *substraiga 1 de un registro dado, o si el registro contiene 0, pase a la instrucción n (indicada)*. (Estas operaciones pueden combinarse para realizar cualquier operación de un computador digital; véase Minsky, 1967, pág. 206). Considérese un computador cuyo código de lenguaje-de-máquina contiene sólo dos instrucciones que corresponden a (a) y a (b). Si se pregunta cómo multiplica o resuelve ecuaciones diferenciales o compone nóminas, puede que se le conteste mostrándole un programa expresado en términos de las dos instrucciones del lenguaje-de-máquina. Pero si se pregunta cómo agrega 1 a un registro dado, la respuesta apropiada se da mediante un diagrama de los circuitos [*wiring diagram*], no mediante un programa. La máquina está construida [*hardwired*] para agregar 1. Cuando la instrucción que corresponde a (a) aparece en un cierto registro, los contenidos del otro registro cambian "automáticamente" de una cierta manera. La estructura computacional de un computador está determinada por un conjunto de operaciones primitivas y por las maneras en que las operaciones no-primitivas se arman a partir de aquéllas. De este modo, no importa a la estructura computacional del computador si los mecanismos primitivos son realizados mediante circuitos de tubos, circuitos de transistores o de relés. Del mismo modo no importa a la psicología de un sistema mental si sus mecanismos primitivos se realizan en uno u otro mecanismo neu-

rológico. Llámese a un sistema una “realización de la psicología humana” si toda teoría psicológica verdadera de nosotros es verdadera de él. Considérese una realización de la psicología humana cuyas operaciones psicológicas primitivas son efectuadas por hombrecitos, de la manera como lo fueron las simulaciones que apelan a la cabeza-homuncular ya discutidas. Así, quizás un hombrecito produzca ítems de una lista uno a uno, otro compare estos ítems con otras representaciones para determinar si se aparean, etcétera.

Ahora bien, existen buenas razones para suponer que este sistema tiene algunos estados mentales. Las actitudes proposicionales son un ejemplo. Quizá, la teoría psicológica identificará recordar que P con haber “almacenado” [“stored”] un objeto de carácter oracional [*sentence like*] que exprese la proposición que P (Fodor, 1975). Entonces, si uno de los hombrecitos puso un cierto objeto de carácter oracional en “almacenamiento”, podemos tener razón para considerar al sistema como recordando que P. Pero a menos que tener *qualia* sea tener cierto procesamiento de información (en el mejor de los casos, una propuesta discutible) no existe una razón teórica tal para considerar al sistema como teniendo *qualia*. En resumen, hay quizá tantas dudas acerca de los *qualia* de este sistema con cabeza-homuncular como las que hay acerca de los *qualia* de la simulación Funcional que apela a cabeza-homuncular, discutida previamente en este artículo.

Pero, *ex hypothesi*, cualquier teoría psicológica es verdadera del sistema que estamos discutiendo. Así, cualquier duda acerca de que tenga *qualia* es una duda acerca de que los *qualia* caigan en el dominio de la psicología.

Podría objetarse: “¡La clase de psicología que se tiene en mente es la psicología *cognitiva*, es decir, la psicología de los procesos de pensamiento, y no es de extrañar que los *qualia* no caigan en el dominio de la psicología *cognitiva*!”. Pero yo no tengo en mente a la psicología cognitiva, y si suena de esa manera, es fácilmente explicable: nada de lo que sabemos acerca de los procesos psicológicos que subyacen a nuestra vida mental consciente tiene que ver con los *qualia*. Lo que suele pasar por “psicología” de la sensación o del dolor es, por ejemplo, (a) fisiología; (b) psicofísica (es decir, el estudio de las funciones matemáticas que relacionan las variables de estímulo con variables de sensación; por ejemplo, la intensidad del sonido como una función de la amplitud de las ondas sonoras), o (c) un conjunto heterogéneo de estudios descriptivos (véase Melzack, 1973, cap. 2). De ellos, sólo la psicofísica podría ser interpretada como ocupándose de los *qualia* per se. Y es obvio que

la psicofísica sólo toca el aspecto *funcional* de la sensación, no su carácter cualitativo. Los experimentos psicofísicos hechos con uno tendrían los mismos resultados que si se hicieran con cualquier sistema Psicofuncionalmente equivalente a uno, aun si tuviera *qualia* invertidos o ausentes. Si los resultados experimentales no cambian, sea que los sujetos experimentales tengan o no tengan *qualia* invertidos o ausentes, difícilmente pueda esperarse que echen luz sobre la naturaleza de los *qualia*.

Por cierto que basándonos en la clase de aparato conceptual del que ahora disponemos en psicología, no veo cómo la psicología en alguna de sus encarnaciones actuales *podría* explicar los *qualia*. No podemos concebir ahora cómo la psicología *podría* explicar los *qualia*, aunque *podemos* concebir cómo la psicología *podría* explicar creer, desear, esperar, etcétera (véase Fodor, 1975). Que algo sea considerado inconcebible no es una buena razón para pensar que sea imposible. Mañana podrían desarrollarse conceptos que hicieran conceivable lo que ahora es inconcebible. Pero todo lo que tenemos para seguir adelante es lo que sabemos y, si nos basamos en lo que tenemos para seguir adelante, pareciera que los *qualia* no caen en el dominio de la psicología...

No es una objeción a la sugerencia de que los *qualia* no son entidades psicológicas, afirmar que los *qualia* sean el paradigma mismo de algo que cae en el dominio de la psicología. Como se ha señalado a menudo, qué cae en el dominio de una rama particular de la ciencia es, en parte, una cuestión empírica. La liquidez del agua no resulta ser explicable por la química, sino más bien por la física subatómica. Las ramas de la ciencia abarcan en todo momento un conjunto de fenómenos que pretenden explicar. Pero puede descubrirse que algún fenómeno que parecía central a una rama de la ciencia pertenece, realmente, al ámbito de una rama diferente...

El Argumento de los *Qualia Ausentes* explota la posibilidad de que el estado Funcional o Psicofuncional que los Funcionalistas o Psicofuncionalistas querrían identificar con el dolor, pueda ocurrir sin que ningún *quale* ocurra. También parece ser conceivable que ocurra un *quale* sin que ocurra dolor. Ciertamente, hay hechos que prestan apoyo a este punto de vista. Luego de las lobotomías frontales, los pacientes informan, típicamente, que todavía tienen dolores, aunque los dolores no los molestan ya (Melzack, 1973, pág. 95). Esos pacientes exhiben todos los signos “sensoriales” de dolor (tal como reconocer la agudeza de un pinchazo), pero a menudo no tienen deseo, o tienen pocos deseos, de evitar los estímulos “dolorosos”.

Un punto de vista sugerido por estas observaciones es que cada

dolor es, en realidad, un estado *compuesto* cuyos componentes son un  *quale* y un estado Funcional o Psicofuncional.<sup>11</sup> O lo que equivale a la misma idea, cada dolor es un  *quale* que desempeña un cierto rol Funcional o Psicofuncional. Si este punto de vista es correcto, ayuda a explicar cómo la gente puede haber creído en teorías tan diferentes acerca de la naturaleza del dolor y de otras sensaciones; han enfatizado un componente a expensas del otro. Quienes proponen el conductismo y el funcionalismo tuvieron en mente un componente; quienes proponen la definición ostensiva privada han tenido en mente al otro. Ambas aproximaciones yerran en tratar de dar una explicación de algo que tiene dos componentes de naturalezas completamente diferentes.

### 3. Chauvinismo vs. liberalismo

Resulta natural entender las teorías psicológicas a las que el Psicofuncionalismo refiere, como teorías de la psicología *humana*. Entendido de este modo, es imposible para el Psicofuncionalismo que un sistema tenga creencias, deseos, etcétera, excepto que las teorías psicológicas que son verdaderas de nosotros sean verdaderas de él. El Psicofuncionalismo (entendido de ese modo) estipula que la equivalencia Psicofuncional con nosotros es necesaria para lo mental [*mentality*].

Pero aun cuando la equivalencia Psicofuncional con nosotros sea una condición de nuestro *reconocimiento de lo mental*, ¿qué razón hay para pensar que sea una condición de lo mental en sí mismo? ¿No podría existir una amplia variedad de procesos psicológicos posibles que subyazgan a lo mental, de los cuales instanciamos sólo un tipo? Supongamos que nos encontramos con marcianos y descubrimos que ellos son, de manera aproximada, Funcionalmente (pero no Psicofuncionalmente) equivalentes a nosotros. Cuando llegamos a conocerlos descubrimos que son tan diferentes de nosotros como los humanos que conocemos. Desarrollamos vastas relaciones culturales y comerciales con ellos. Cada cual estudia la ciencia y los periódicos filosóficos del otro, asiste a las películas del otro, cada cual lee las novelas del otro, etcétera. Entonces los psicólogos marcianos y los terrestres comparan sus anotaciones, sólo para descubrir que en la psicología subyacente, marcianos y terrestres

11. El  *quale* podría ser identificado con un estado físico-químico. Este punto de vista concordaría con una sugerencia hecha por Hilary Putnam a fines de los 60 en su seminario de filosofía de la mente. Véase también cap. 5 de Gunderson, 1971.

son muy diferentes. Pronto acuerdan que la diferencia puede describirse como sigue. Pensemos en los humanos y en los marcianos como si fueren productos de un diseño consciente. En tal proyecto de diseño habrá varias opciones. Algunas capacidades pueden ser asignadas por el diseño [*built in*] (innatas), otras pueden ser aprendidas. El cerebro puede ser diseñado para llevar a cabo tareas usando tanta capacidad de memoria como sea necesaria para minimizar el uso de la capacidad computacional, o, por otra parte, el diseñador podría preferir conservar espacio de memoria y contar principalmente con capacidad computacional. Las inferencias pueden ser llevadas a cabo por sistemas que utilicen pocos axiomas y muchas reglas de inferencia o, en cambio, pocas reglas y muchos axiomas. Imagíñese, ahora, que lo que los psicólogos marcianos y terrestres descubren cuando comparan sus anotaciones es que los marcianos y los terrestres difieren como si fueran los productos finales de elecciones de diseño maximalmente diferentes (compatibles con la equivalencia Funcional aproximada en los adultos). ¿Deberíamos rechazar nuestro supuesto de que los marcianos pueden disfrutar de nuestras películas, creer en los resultados científicos aparentes, etcétera? ¿Deberían "rechazar" su "supuesto" de que nosotros "disfrutamos" sus novelas, "aprendemos" de sus libros de texto, etcétera? Quizá no he proporcionado información suficiente para responder a esta pregunta. Después de todo, puede haber muchas maneras de completar la descripción de las diferencias humano-marcianas respecto de las cuales sería razonable suponer que no hay, simplemente, hechos decisivos, o suponer aun que los marcianos no merecen adscripciones mentales. Pero seguramente hay muchas maneras de completar la descripción de la diferencia marciano-terráquea que he esquematizado, según la cual sería perfectamente evidente que aun cuando los marcianos se comportaran de manera diferente de nosotros de acuerdo con experimentos psicológicos sutiles, no obstante, piensan, desean, disfrutan, etcétera. Suponer lo contrario sería puro chauvinismo humano. (Recuérdese que las teorías son chauvinistas en tanto que *niegan* falsamente que los sistemas tengan propiedades mentales, y liberales, en tanto *adscriben* falsamente propiedades mentales.)

Una sugerencia obvia para salir de esta dificultad consiste en identificar a los estados mentales con estados Psicofuncionales, considerando al dominio de la psicología de modo que incluya a *todas las criaturas con mentalidad*, incluidos los marcianos. La sugerencia es que definamos "Psicofuncionalismo" en términos de una psicología "universal" o "intersistémica" [*"cross-system"*], en lugar de la psicología humana, tal

como supuse antes. La psicología universal, sin embargo, es una empresa sospechosa. Porque, ¿cómo hemos de decir nosotros qué sistemas deben ser incluidos en el *dominio* de la psicología universal? Una manera posible de decidir qué sistemas tienen mentalidad y así cuáles caen en el dominio de la psicología universal, sería usar alguna *otra teoría* desarrollada de lo mental tal como el conductismo o el Funcionalismo. Pero tal procedimiento sería al menos tan carente de justificación como la otra teoría usada. Además, si el Psicofuncionalismo tiene que presuponer alguna otra teoría de la mente, podríamos muy bien aceptar en su lugar a la otra teoría de la mente.

Quizá la psicología universal evitará este problema del "dominio", del mismo modo que otras ramas de la ciencia lo evitan o buscan evitarlo. Otras ramas de la ciencia comienzan con dominios tentativos, basadas en versiones intuitivas y precientíficas de los conceptos que ellas, se supone, explican. Luego, intentan desarrollar clases naturales [*natural kinds*] de una manera que permite la formulación de generalizaciones legaliformes que se aplican a todas o a la mayoría de las entidades de los dominios precientíficos. En el caso de la mayoría de las ramas de la ciencia —incluyendo las ciencias biológicas y sociales tales como la genética y la lingüística—, el dominio precientífico resultó ser adecuado para la articulación de generalizaciones legaliformes.

Ahora bien, podría ser que fuéramos capaces de desarrollar una psicología universal de la misma manera en que desarrollamos la psicología terrestre. Decidimos, apoyados en una base intuitiva y precientífica, qué criaturas incluir en su dominio, y trabajamos para desarrollar clases naturales de la teoría psicológica, que aplicamos a todas o al menos a la mayoría de ellas. Quizás el estudio de una clase amplia de organismos que se encuentren en mundos diferentes conducirá un día a desarrollar teorías que determinen condiciones de verdad para la adscripción de estados mentales como creencia, deseo, etcétera, aplicables a sistemas que son preteóricamente diferentes de nosotros. Por cierto que tal psicología inter-mundos requerirá, sin duda, una clase completamente diferente de conceptos mentales. Quizás, habrá familias de conceptos que se correspondan con creencia, deseo, etcétera, es decir, una familia de conceptos similares a creencia [*belief-like concepts*], conceptos similares a deseo [*desire-like concepts*], etcétera. Si tal fuera el caso, la psicología universal que desarrollemos, sin duda dependerá, de alguna manera, de qué nuevos organismos descubramos primero. Aun cuando la psicología universal fuera de hecho posible, sin embargo, ciertamente habrá muchos organismos posibles cuyo *status* mental sea indeterminado.

Por otra parte, puede ser que la psicología universal *no* sea posible. Quizá la vida en el universo sea tal que, sencillamente, no tengamos bases para decisiones razonables acerca de cuáles son los sistemas que entran en el dominio de la psicología y cuáles no.

Si la psicología universal *fuerza* posible, el problema que he estado planteando se desvanece. El Psicofuncionalismo-universal evita el liberalismo del funcionalismo y el chauvinismo del Psicofuncionalismo-humano. Pero la pregunta de si es posible la psicología universal es un interrogante que ahora no tenemos manera de responder.

He aquí una síntesis de lo argumentado:

1. El Funcionalismo tiene la consecuencia extraña de que la simulación de uno, apelando a la cabeza-homuncular, tiene *qualia*. Esto pone la carga de la prueba en el Funcionalista, a fin de que nos dé alguna razón para creer en su doctrina. Sin embargo, el único argumento que hay en la literatura a favor del Funcionalismo no es bueno, y así el Funcionalismo no da señales de satisfacer la carga de la prueba.
2. Las simulaciones Psicofuncionales que se hacen de nosotros comparten los estados mentales que caen en el dominio de la psicología, de modo que la cabeza-homuncular Psicofuncional no arroja duda sobre las teorías Psicofuncionales de los estados cognitivos, sino sólo sobre las teorías Psicofuncionales de los *qualia*, quedando la duda de si los *qualia* caen en el dominio de la psicología.
3. Las teorías Psicofuncionalistas de los estados mentales que caen en el dominio de la psicología son, sin embargo, irremediablemente chauvinistas.

Así, una versión del funcionalismo tiene problemas con el liberalismo y la otra con el chauvinismo. Porque en lo que respecta a los *qualia*, si caen en el dominio de la psicología, entonces el Psicofuncionalismo es a los *qualia* tan chauvinista como el Psicofuncionalismo lo es a la creencia. Por otra parte, si los *qualia* no caen en el dominio de la psicología, la cabeza-homuncular Psicofuncionalista puede ser usada contra el Psicofuncionalismo respecto de los *qualia*. Porque lo único que protege al Psicofuncionalismo del argumento de la cabeza-homuncular con respecto al estado mental S, es que si uno tiene S entonces cualquier simulación Psicofuncional de uno tiene que tener S; porque la teoría correcta de S se aplica tanto a ella como a uno.

### 3.1 El problema de los inputs y de los outputs

He estado suponiendo (como a menudo lo hacen los Psicofuncionalistas, véase Putnam, 1967) que los *inputs* y los *outputs* pueden especificarse mediante descripciones de impulsos neurales. Pero esta es una afirmación chauvinista, puesto que impide que organismos sin neuronas (por ejemplo, máquinas) tengan descripciones funcionales. ¿Cómo puede uno evitar el chauvinismo con respecto a la especificación de *inputs* y de *outputs*? Una manera sería caracterizar los *inputs* y los *outputs* sólo como *inputs* y *outputs*. Así, la descripción funcional de una persona podría listar *outputs* numerándolos:  $output_1, output_2, \dots$ . Entonces un sistema podría ser funcionalmente equivalente a uno si tuviera un conjunto de estados, *inputs* y *outputs* causalmente relacionados entre sí de la manera en que lo están los nuestros, cualesquiera sean los estados, *inputs* y *outputs*. Por cierto que aunque este enfoque viola la exigencia de algunos funcionalistas de que los *inputs* y los *outputs* sean especificados físicamente, otros funcionalistas —aquellos que insisten en que sólo las descripciones de *input* y de *output* sean *no-mentales*— pueden haber tenido en mente algo como eso. Esta versión del funcionalismo no “liga” [*tack down*] las descripciones relativamente específicas de los *inputs* y de los *outputs* a las descripciones funcionales en la periferia; más bien, esta versión del funcionalismo trata a los *inputs* y a los *outputs* como todas las versiones del funcionalismo tratan a los estados internos. Es decir, esta versión especifica estados, *inputs* y *outputs* requiriendo sólo que sean estados, *inputs* y *outputs*.

El problema con esta versión del funcionalismo es que es extremadamente liberal. Los sistemas económicos tienen *inputs* y *outputs*, tal como la entrada y salida de créditos y débitos. Y los sistemas económicos tienen también una rica variedad de estados internos, tal como tener una tasa de incremento del PBN igual al doble de la tasa mínima de interés. No parece imposible que un jeque acaudalado pudiera ganar el control de la economía de un país pequeño, por ejemplo Bolivia, y manipular su sistema financiero para hacerlo funcionalmente equivalente a una persona, por ejemplo a él mismo. Si esto parece implausible, recuérdese que los estados, *inputs* y *outputs* económicos que el jeque hace corresponder a sus estados, *inputs* y *outputs* mentales, no precisan ser magnitudes económicas “naturales”. Nuestro jeque hipotético podría tomar cualesquiera de las magnitudes económicas, por ejemplo la quinta derivada del balance de pago. Su única limitación es que las magnitudes que elija sean económicas, que las magnitudes tengan tales

y cuales valores, sean *inputs*, *outputs* y estados, y que él sea capaz de montar una estructura financiera que se acomode al modelo formal propuesto. El mapeado [*mapping*] de las magnitudes psicológicas en las magnitudes económicas podría ser tan estrañamente como el jeque quisiera.

Esta versión del funcionalismo es demasiado liberal y en consecuencia tiene que ser rechazada. Si hay puntos acordados cuando se discute el problema mente-cuerpo, uno de ellos es que la economía de Bolivia no podría tener estados mentales, no importa cuánto sea distorsionada por aficionados poderosos. Obviamente, tenemos que ser más específicos en nuestras descripciones de *inputs* y de *outputs*. La pregunta es: ¿existe una descripción de *inputs* y de *outputs* suficientemente específica como para evitar el liberalismo y, sin embargo, lo suficientemente general como para evitar el chauvinismo? Dudo que la haya.

Toda propuesta para la descripción de *inputs* y de *outputs* que he visto o pensado peca o bien de liberalismo o de chauvinismo. Aunque este trabajo se ha concentrado en el liberalismo, el chauvinismo es el problema más extendido. Considérense las descripciones Funcionales y Psicofuncionales estándar. Los Funcionalistas tienden a especificar los *inputs* y los *outputs* a la manera de los conductistas: los *outputs* en términos de movimientos de brazos y piernas, sonidos emitidos y cosas similares; los *inputs* en términos de luz y sonido penetrando en ojos y oídos... Tales descripciones son descaradamente *específicas-de-la-especie* [*species-specific*]. Los humanos tienen brazos y piernas, pero las víboras no, y sea que las víboras tengan o no tengan mentalidad, se puede imaginar fácilmente criaturas similares a las víboras que las tengan. Ciertamente, se puede imaginar criaturas con todo tipo de dispositivos *input-output*, por ejemplo, criaturas que se comunican y manipulan [cosas] mediante la emisión de fuertes campos magnéticos. Por supuesto, se podrían formular descripciones Funcionales para cada una de tales especies y en algún lugar del paraíso disyuntivo existe una descripción disyuntiva que abarcaría a todas las especies que existen efectivamente en el universo (la descripción puede ser infinitamente extensa). Pero ni siquiera la apelación a entidades sospechosas tales como las disyunciones infinitas liberará al Funcionalismo, pues ni aun el punto de vista corregido nos dirá qué hay en común en los organismos que sienten dolor. Y no permitirá la adscripción de dolor a algunas criaturas hipotéticas (pero inexistentes) que sientan dolor. Más aún, éstas constituyen las bases sobre las cuales los funcionalistas rechazan, acerbamente, a las teorías disyuntivas propuestas a veces con desesperación por los fisicalistas. Si de pronto los funcionalistas vieran con agrado a los estados

como disyuntivas extravagantes para salvarse a sí mismos del chauvinismo, no tendrían manera de defenderse del fisicalismo.

Las descripciones Psicofuncionales estándar de *inputs* y de *outputs* son también específicas-de-la-especie (por ejemplo, en términos de actividad neural) y en consecuencia son también chauvinistas.

No resulta difícil explicar el chauvinismo de las descripciones *input-output* estándar. La variedad de vida inteligente posible es enorme. Dadas descripciones adecuadamente específicas de *inputs* y *outputs*, cualquier aprendiz de ciencia ficción en la edad secundaria será capaz de describir un ser sintiente y sapiente cuyos *inputs* y *outputs* no satisfagan esa descripción.

Argumentaré que *cualquier descripción física* de *inputs* y de *outputs* (recuérdese que muchos funcionalistas han insistido en las descripciones físicas) produce una versión del funcionalismo que inevitablemente es chauvinista o liberal. Imaginémonos con quemaduras tan graves que la manera óptima de comunicarnos con el mundo externo sea vía modulaciones de nuestro patrón personal EEG en Código Morse. Descubrimos que pensar un pensamiento excitante produce un patrón que la audiencia acuerda en interpretar como un punto y un pensamiento triste produce una "raya". Por cierto que esta fantasía no está lejos de la realidad. Según un artículo periodístico aparecido en el *Boston Globe* el 21 de marzo de 1976, "En UCLA, los científicos están trabajando en el uso del EEG para controlar máquinas... Un sujeto pone electrodos en su cuero cabelludo, y piensa un objeto a través de un laberinto". Presumiblemente, el proceso "inverso" es también posible: otros se comunican con uno en Código Morse mediante la producción de una descarga de actividad eléctrica que afecta nuestro cerebro (causando, por ejemplo, una posimagen [*afterimage*] extensa o breve). Recíprocamente, si los cerebroscopios que los filósofos a menudo imaginan se tornaran una realidad, nuestros pensamientos se leerían directamente de su cerebro. Nuevamente, el proceso inverso también parece posible. En estos casos, *el cerebro mismo se vuelve una parte esencial de los dispositivos input y output de uno*. Esta posibilidad tiene consecuencias embarazosas para los funcionalistas. Se recordará que los funcionalistas sostienen que el fisicalismo es falso porque un estado mental único puede ser realizado por una variedad indefinidamente grande de estados físicos que no tienen caracterización física necesaria ni suficiente. Pero si este argumento funcionalista contra el fisicalismo es correcto, *el mismo argumento vale para los inputs y los outputs*, puesto que la realización física de los estados mentales puede servir como una parte esencial de los dispositivos

de *inputs* y de *outputs*. Es decir, en cualquier sentido de 'físico' en el que la crítica funcionalista al fisicalismo es correcta, *no existirá caracterización física que valga para todos los sistemas mentales de inputs y de outputs, y sólo para ellos*. En consecuencia, todo intento de formular una descripción funcional con caracterizaciones físicas de *inputs* y *outputs*, excluirá inevitablemente algunos sistemas con mentalidad o incluirá algunos sistemas con mentalidad... *los funcionalistas no pueden evitar ni el chauvinismo ni el liberalismo*.

Así, las especificaciones físicas de *inputs* y de *outputs* no servirán. Más aún, la terminología mental o de "acción" (tal como "golpear a la víctima") tampoco puede usarse, puesto que usar tales especificaciones de *inputs* o *outputs* sería dejar a un lado al programa funcionalista que caracteriza a la mentalidad en términos no-mentales. Por otra parte, como se recordará, el caracterizar a *inputs* y *outputs* simplemente como *inputs* y *outputs*, es inevitablemente liberal. No veo cómo pueda haber un vocabulario para describir *inputs* y *outputs* que evite al liberalismo y al chauvinismo. No pretendo que este argumento sea concluyente contra el funcionalismo. Más bien, como el argumento funcionalista contra el fisicalismo, es mejor interpretarlo como un argumento acerca de la carga de la prueba. El funcionalista dice al fisicalista: "Es muy difícil ver cómo podría existir una única caracterización física de los estados internos de todas las criaturas con mentalidad, y sólo de ellas". Yo le digo al funcionalista: "Es muy difícil ver cómo podría existir una única caracterización física de los *inputs* y *outputs* de todas las criaturas con mentalidad, y sólo de ellas". En ambos casos, se ha dicho suficiente como para que el esbozo de cómo podrían ser posibles tales caracterizaciones sea responsabilidad de quienes piensan que podría haberlas.<sup>12</sup>

TRADUCTORA: Eleonora Baringoltz.

REVISIÓN TÉCNICA: Eduardo Rabossi.

12. Estoy en deuda con Sylvain Bromberger, Harry Field, Jerry Fodor, David Hills, Paul Horwich, Bill Lycan, Georges Rey y David Rosenthal, por sus comentarios detallados acerca de alguna de las primeras versiones de este trabajo. Partes de las primeras versiones fueron leídas a comienzos del otoño de 1975, en Tufts University, Princeton University, University of North Carolina en Greensboro, y State University of New York en Binghamton.

## REFERENCIAS BIBLIOGRÁFICAS

- Armstrong, D.: (1968) *A Materialist Theory of Mind*. Londres: Routledge & Kegan Paul.
- Bever, T.: (1970) "The cognitive basis for linguistic structures", en J.R. Hayes (comp.), *Cognition and the Development of Language*. Nueva York, Wiley.
- Block, N.: (1980) "Are absent qualia impossible?", *Philosophical Review* 89 (2).
- Block, N. y Fodor J.: (1972) "What psychological states are not", *Philosophical Review* 81, 159-81.
- Chisholm, Roderick: (1957) *Perceiving*. Ithaca, Cornell University Press.
- Cummins, R.: (1975) "Functional analysis", *Journal of Philosophy* 72, 741-64.
- Davidson, D.: (1970) "Mental events", en L. Swanson y J.W. Foster (comps.), *Experience and Theory*. Amherst, University of Massachusetts Press.
- Dennett, D.: (1969) *Content and Consciousness*. Londres, Routledge & Kegan Paul.
- Dennett, D.: (1975) "Why the law of effect won't go away", *Journal for the Theory of Social Behavior* 5, 169-87.
- Dennett, D.: (1978a) "Why a computer can't feel pain", *Synthèse* 38, 3.
- Dennett, D.: (1978b) *Brainstorms*, Montgomery, Vt., Bradford.
- Feldman, F.: (1973) "Kripke's argument against materialism", *Philosophical Studies*, 416-19.
- Fodor, J.: (1965) "Explanations in psychology", en M. Black (comp.), *Philosophy in America*, Londres, Routledge & Kegan Paul.
- Fodor, J.: (1968) "The appeal to tacit knowledge in psychological explanation", *Journal of Philosophy* 65, 627-40.
- Fodor, J.: (1974) "Special sciences", *Synthèse* 28, 97-115.
- Fodor, J. y Garrett, M.: (1967) "Some syntactic determinants of sentential complexity", *Perception and Psychophysics* 2, 289-96.
- Geach, P.: (1957) *Mental Acts*, Londres, Routledge & Kegan Paul.
- Gendron, B.: (1971) "On the relation of neurological and psychological theories: A critique of the hardware thesis", en R.C. Buck y R.S. Cohen (comps.), *Boston Studies in the Philosophy of Science VIII*. Dordrecht, Reidel.
- Grice, H.P.: (1975) "Method in philosophical psychology (from the banal to the bizarre)", *Proceedings and Addresses of the American Philosophical Association*.

- Gunderson, K.: (1971) *M mentality and Machines*, Garden City, Doubleday Anchor.
- Harman, G.: (1973) *Thought*, Princeton, Princeton University Press.
- Hempel, C.: (1970) "Reduction: Ontological and linguistic facets", en S. Morgenbesser, P. Suppes y White (comps.), *Essays in Honor of Ernst Nagel*. Nueva York, St. Martin's Press.
- Kalke, W.: (1969) "What is wrong with Fodor and Putnam's functionalism?", *Noûs* 3, 83-93.
- Kim, J.: (1972) "Phenomenal properties, psychophysical laws, and the identity theory", *The Monist* 56 (2), 177-92.
- Lewis, D.: (1972) "Psychophysical and theoretical identifications", *Australasian Journal of Philosophy* 50 (3), 249-58.
- Locke, D.: (1968) *Myself and Others*, Oxford, Oxford University Press.
- Melzack, R.: (1973) *The Puzzle of Pain*, Nueva York, Basic Books.
- Minsky, M.: (1967) *Computation*. Englewood Cliffs, NJ, Prentice-Hall.
- Mucciolo, L.F.: (1974) "The identity thesis and neuropsychology", *Noûs* 8, 327-42.
- Nagel, T.: (1969) "The boundaries of inner space", *Journal of Philosophy* 66, 452-8.
- Nagel, T.: (1970) "Armstrong on the mind", *Philosophical Review* 79, 394-403.
- Nagel, T.: (1972) "Review of Dennett's *Content and Consciousness*", *Journal of Philosophy* 50, 220-34.
- Nagel, T.: (1974) "What is it like to be a bat?", *Philosophical Review* 83, 435-50.
- Nelson, R.J.: (1969) "Behaviorism is false", *Journal of Philosophy* 66, 417-52.
- Nelson, R.J.: (1975) "Behaviorism, finite automata and stimulus response theory", *Theory and Decision*, 6, 249-67.
- Nelson, R.J.: (1976) "Mechanism, functionalism, and the identity theory", *Journal of Philosophy* 73, 365-86.
- Oppenheim, P. y Putnam, H.: (1958) "Unity of science as a working hypothesis", en H. Feigl, M. Scriven y G. Maxwell (comps.), *Minnesota Studies in the Philosophy of Science II*, Minneapolis, University of Minnesota Press.
- Pitcher, G.: (1971) *A Theory of Perception*, Princeton, Princeton University Press.
- Putnam, H.: (1963) "Brains and behavior"; reimpreso, como todos los artículos de Putnam citados en este trabajo (excepto "On proper-

- ties"), en *Mind, Language and Reality, Philosophical Papers*, vol. 2, Londres, Cambridge University Press, 1975.
- Putnam, H.: (1966) "The mental life of some machines".
- Putnam, H.: (1967) "The nature of mental states" (publicado originalmente con el título de "Psychological Predicates").
- Putnam, H.: (1970) "On properties", en *Mathematics, Matter and Method: Philosophical Papers*, vol. 1, Londres, Cambridge University Press.
- Putnam, H.: (1975a) "Philosophy and our mental life".
- Putnam, H.: (1975b) "The meaning of 'meaning'".
- Rorty, R.: (1972) "Functionalism, machines and incorrigibility", *Journal of Philosophy* 69, 203-20.
- Scriven, M.: (1966) *Primary Philosophy*, Nueva York, Mc Graw-Hill.
- Sellars, W.: (1956) "Empiricism and the philosophy of mind", en H. Feigl y M. Scriven (comps.), *Minnesota Studies in Philosophy of Science I*, Minneapolis, University of Minnesota Press.
- Sellars, W.: (1968) *Science and Metaphysics* (cap. 6), Londres, Routledge & Kegan Paul.
- Shoemaker, S.: (1975) "Functionalism and *qualia*", *Philosophical Studies* 27, 271-315.
- Shoemaker, S.: (1976) "Embodiment and behavior", en A. Rorty (comp.), *The Identities of Persons*, Berkeley, University of California Press.
- Shallice, T.: (1972) "Dual functions of consciousness", *Psychological Review* 79, 383-93.
- Smart, J.J.C.: (1971) "Reports of immediate experience", *Synthèse* 22, 346-59.
- Wiggins, D.: (1975) "Identity, designation, essentialism, and physicalism", *Philosophia* 5, 1-30.