

¿Qué es el funcionalismo?

David Villena Saldaña

Resumen

El presente trabajo explica y articula la tesis funcionalista sobre los estados mentales. La clasifica como una respuesta al problema mente-cuerpo en su aspecto metafísico. Señala en qué radica la particularidad de las propiedades funcionales en tanto propiedades de segundo orden y muestra el modo de conjugar las tesis de la superveniencia y la realización múltiple. Aplica finalmente todas estas ideas a la presentación del funcionalismo de máquina y del funcionalismo analítico, que son las dos versiones más influyentes del funcionalismo.

Palabras clave: Funcionalismo, estados mentales, problema mente-cuerpo, filosofía de la mente, filosofía de la psicología.

Abstract

This paper explains the main theses of functionalism about mental states. This view is taken as a response addressed to the metaphysical aspect of the mind-body problem. Thus it is said what distinguishes functional properties as second-order properties, and how to understand supervenience and multiple realizability. The paper applies these ideas to machine functionalism and analytic functionalism, the two main versions of functionalism.

Keywords: Functionalism, mental states, mind-body problem, philosophy of mind, philosophy of psychology.

1. Introducción

El funcionalismo intenta dar cuenta de cuál es la *naturaleza* de los estados mentales. Esto es, aborda específicamente la pregunta “¿qué son los estados mentales?”. En tal sentido, es una respuesta al problema mente-cuerpo en su aspecto metafísico. Se puede afirmar que al día de hoy, habiendo desbancado tanto al conductismo lógico como al fisicalismo, constituye la posición hegemónica en torno a este problema¹. Es, por

¹ Cf. Maloney (1999) y Levin (2013).

decirlo así, la *concepción recibida* (o la *concepción heredada*) en la filosofía de la mente y en la ciencia cognitiva.

El núcleo de esta teoría – aquella idea común a sus diferentes manifestaciones – es la comprensión de los estados mentales a modo de intermediarios con un rol causal específico entre *inputs* sensoriales, *outputs* conductuales y otros estados mentales. Un estado mental particular es, consecuentemente, un estado mental de cierto tipo según el rol causal o la función que desempeñe dentro del sistema del cual forma parte. Esto quiere decir que los estados mentales no son causalmente neutrales y que tienen como naturaleza – diríase incluso que como “esencia” – un rol causal específico, una función específica. No son meros epifenómenos como creían los materialistas del siglo XIX ni simples postulados para hacer referencia a una disposición a la conducta, como aducían los conductistas lógicos de mitad del siglo XX. Así, el estado mental que experimento ahora, aquel que es consecuencia de haberme encontrado en estado de alivio, recibir el *input* sensorial de una piedra cayendo sobre mi pie, y que me lleva a exclamar ‘¡Au!’ como *output* conductual, es un episodio o caso [*token*] del tipo [*type*] mental dolor. Mi estado mental es de dolor, pues desempeña el rol causal atribuido al dolor como tipo y encaja con la descripción funcional del dolor.

2. Propiedades funcionales

Desde un punto de vista lógico, las descripciones (caracterizaciones o definiciones) de propiedades funcionales son de segundo orden. Cuantifican sobre variables predicativas al decir que algún objeto tiene cierta propiedad funcional siempre y cuando tenga una propiedad en particular que satisfaga un rol causal específico. Repárese en el siguiente esquema²:

(F) Algún x tiene F = *def* ser un x tal que $\exists P(C(P) \ \& \ x \ \text{tiene} \ P)$

Esto se lee como: algún x tiene la propiedad funcional F si y sólo si hay un x, hay una propiedad P, P tiene la propiedad de desempeñar el rol causal C y x tiene la propiedad P. La propiedad funcional F está siendo definida en términos de la satisfacción de un rol causal C específico por parte de una propiedad P que no es específica. Pues el término ‘P’ es una variable. De este modo, la propiedad funcional F no está asociada a ninguna propiedad concreta, a esta de aquí o a aquella otra, pues cabe la posibilidad de que más de una propiedad concreta satisfaga el rol causal C – es decir,

² Aquí sigo la línea de explicación en torno a las propiedades funcionales ofrecida por Kim (1996).

cabe la posibilidad de que más de una constante predicativa ocupe el lugar de la variable 'P' y satisfaga el rol causal C. La propiedad funcional F es, por tanto, una propiedad de carácter abstracto y formal. Es una propiedad de segundo orden.

La literatura funcionalista suele poner como ejemplo de propiedad funcional a la propiedad de ser un carburador. Considerando el esquema (F) de definición funcional dado líneas más arriba, un objeto y tiene la propiedad de ser un carburador si cuenta con la propiedad de desempeñar un rol causal específico: si cuenta con una propiedad que le permita mezclar aire con combustible y proveer de dicha mezcla a un motor. Todo aparato que cubra con esta definición funcional será un carburador, no importa cuál sea la propiedad concreta que le permita desempeñar el rol causal especificado. En efecto, resulta indiferente cuál sea su estructura o de qué esté hecho, si de hierro, bronce, oro o plástico; estas propiedades pueden variar (equivalen a eventuales valores de la variable predicativa 'P' del esquema (F)), lo que importa es que permitan al objeto y mezclar aire con combustible y proveer de dicha mezcla a un motor. El carburador es un tipo funcional, no un tipo físico. Su naturaleza está determinada por *lo que hace*, no por aquello de lo cual está hecho.

Lo mismo puede afirmarse de una ratonera. Un objeto z tiene la propiedad de ser una ratonera en caso de que cuente con una propiedad que le permita desempeñar un rol causal específico: *input* de ratón vivo y *output* de ratón muerto. Todo aquello que sea un intermediario causal entre dicho *input* y dicho *output* será una ratonera. Nuevamente, resulta irrelevante cuál sea la propiedad concreta – de qué está hecho o cuál es su estructura – que permita al objeto satisfacer el rol causal mencionado. Si cumple con ese rol, es una ratonera. No hay nada más que decir sobre la naturaleza o “esencia” de esta propiedad.

Ahora bien, si, como reza la tesis funcionalista, los estados mentales son estados funcionales, de ello se sigue que los estados mentales son estados abstractos y de segundo orden, tal y como lo son las propiedades de ser un carburador y de ser una ratonera – y también, por cierto, las propiedades de ser un corazón (rol causal: bombear sangre) y de ser un riñón (rol causal: filtrar sangre y conservar ciertos balances químicos). Los funcionalistas consideran que los estados mentales no están atados a su realización (instanciación o ejemplificación) por parte de un estado concreto en específico. Se declaran neutrales en relación con el sustrato. En tanto estados funcionales, los estados mentales pueden ser realizados (instanciados o ejemplificados) por varios estados concretos posibles siempre y cuando estos estados concretos

satisfagan el rol causal especificado por la definición funcional del caso, e.g., la definición de dolor como estado funcional es indiferente a la constitución fisicoquímica de los estados que lo realicen (instancien o ejemplifiquen), lo único que cuenta es que estos estados concretos satisfagan el rol causal específico del dolor.

Esta última implicancia calza con la intuición de que un ente cuyo cerebro sea diferente al nuestro – o que quizá no tenga cerebro – pueda tener dolor. Este ente puede ser un pulpo, un extraterrestre o un robot. El corolario es una negación abierta de la tesis fisicalista de la identidad psicofísica tipo-tipo: el dolor como tipo no es idéntico a la estimulación de las fibras-C. La mente no es el cerebro humano. La mente es un conjunto de propiedades funcionales definidas por un rol causal que eventualmente satisface el cerebro humano y que también eventualmente podría ser satisfecho por otro órgano u objeto. Los estados mentales no son estados físicos; de serlo, constituirían estados de primer orden, estados con una “esencia real” dada por un tipo de sustrato específico. Son estados funcionales, estados de segundo orden y, en consecuencia, estados sin “esencia real.” Su esencia, si alguna, es *nominal*, como sugiere Kim (1996) recordando la distinción lockeana entre esencias reales y esencias nominales. Esta esencia nominal es el rol causal o función.

Resulta preciso indicar que la negación funcionalista del fisicalismo de tipos no implica la negación de todo fisicalismo. De hecho, la tesis funcionalista es compatible con el fisicalismo de casos [*token physicalism*]. Es decir, resulta compatible con afirmar que mi caso o episodio singular de dolor – lo que experimento ahora, en el *hic et nunc* – es idéntico a *este* caso o episodio singular de estimulación de las fibras-C, aunque no haya identidad entre tipos, pues *este* caso o episodio singular de estimulación de las fibras-C desempeña el rol causal específico del dolor en tanto estado funcional. Desde luego, aun cuando sea una cuestión de hecho que hasta aquí todos los casos o episodios singulares registrados de dolor (léase: realizaciones, ejemplificaciones o instanciaciones del estado funcional de dolor como tipo) sean idénticos a casos o episodios singulares de estimulación de las fibras-C, eso no coloca en relación de identidad al dolor como tipo con la estimulación de las fibras-C también como tipo. Si los casos o episodios singulares de estimulación de las fibras-C son idénticos a todos los casos o episodios singulares registrados de dolor, no es en virtud de una necesidad nomológica, sino por el simple hecho de que las estimulaciones de las fibras-C desempeñan el rol causal específico del dolor. El dolor es idéntico a este rol causal, no a la estimulación de las fibras-C, pues en principio casos o episodios singulares de estados físicos de otro tipo

también podrían desempeñar el mismo rol causal y, así, ser idénticos a casos o episodios singulares de dolor. La caracterización funcional del dolor, como ya se ha subrayado, coloca a este estado en un plano diferente al de los estados físicos o de primer orden, aquellos cuyas caracterizaciones, a diferencia de los de segundo orden, no cuantifican sobre propiedades o variables predicativas, sino únicamente sobre variables individuales. Por lo mismo, el dolor no puede estar en relación de identidad con estados físicos de primer orden: el dolor no es idéntico a la estimulación de las fibras-C. Ello sería incurrir en un error categorial.

3. Superveniencia y realización múltiple

Llegados a este punto es relevante introducir la noción de *superveniencia*³. Considérese para ello que si dos objetos m y n no son equivalentes en términos funcionales, tendrán que diferir en lo que respecta a sus propiedades físicas. O sea, es imposible que dos objetos no tengan las mismas propiedades funcionales y, sin embargo, tengan las mismas propiedades físicas. En efecto, bajo el supuesto de que m y n tengan las mismas propiedades físicas, si la propiedad física Φ de m satisface cierto rol causal C y, por tanto, realiza (instancia o ejemplifica) cierta propiedad funcional Ψ , entonces n debe presentar la misma propiedad funcional Ψ , pues cuenta con la misma propiedad física Φ de m y esta propiedad no puede desempeñar el rol causal C en el caso de m y dejar de hacerlo en el de n . Si sucede que m tiene la propiedad funcional Ψ y n no la tiene, es decir, si sucede que m y n no son funcionalmente equivalentes, debe haber una propiedad física Φ que m tenga y con la cual n no cuente, esto es, debe ocurrir que m y n no tengan las mismas propiedades físicas.

Las diferencias funcionales entre dos objetos indican diferencias de carácter físico entre ellos. Esto manifiesta un sentido de dependencia de las propiedades funcionales en relación con las propiedades físicas. Pero esta dependencia no es estricta. Pues, aunque es cierto que a iguales propiedades físicas, iguales propiedades funcionales; no es cierto que a iguales propiedades funcionales, iguales propiedades físicas. Puede haber diferencias en propiedades físicas y, no obstante, haber equivalencia en propiedades funcionales (piénsese si no, en los ejemplos dados sobre el carburador, la ratonera, el corazón y el riñón), pero no puede ocurrir que, habiendo las mismas propiedades físicas en dos o más objetos, no haya equivalencia en propiedades

³ La explicación de Moya (2006) sobre el particular es sumamente útil. Véase también Kim (1993).

funcionales entre estos objetos. Este tipo de dependencia, que más que lógica resulta metafísica, es lo que se conoce como superveniencia. En tal sentido, se dice que las propiedades funcionales guardan una relación de superveniencia con las propiedades físicas. Ciertas propiedades físicas son suficientes para la realización (instanciación o ejemplificación) de cierta propiedad funcional – si se las tiene, es imposible no tener la propiedad funcional del caso; si se tiene propiedades físicas que permitan mezclar aire con combustible y proveer de esta mezcla a un motor, es imposible no tener la propiedad funcional de ser un carburador. Pero ninguna propiedad física es necesaria para la realización (instanciación o ejemplificación) de cierta propiedad funcional – puede haber objetos que cuenten con otras propiedades físicas y que tengan las mismas propiedades funcionales. Esto último es lo que se quiere decir al señalar que no hay relación de necesidad nomológica entre propiedades funcionales y propiedades físicas.

Resumiendo la cuestión de la superveniencia, las propiedades funcionales guardan una relación de superveniencia con las propiedades físicas porque no puede haber dos objetos con las mismas propiedades físicas y con propiedades funcionales diferentes, pero sí puede ocurrir que dos objetos con las mismas propiedades funcionales tengan diferentes propiedades físicas. No puede ocurrir, asimismo, que un mismo objeto en dos tiempos diferentes, contando con las mismas propiedades físicas, presente diferentes propiedades funcionales; ni puede pasar que un mismo objeto cambie de propiedades funcionales permaneciendo, no obstante, con las mismas propiedades físicas. Es decir:

- (S1) Es imposible que dos objetos m y n tengan las mismas propiedades físicas Φ y no tengan las mismas propiedades funcionales Ψ ;
- (S2) Es posible que dos objetos m y n tengan las mismas propiedades funcionales Ψ y no tengan las mismas propiedades físicas Φ ;
- (S3) Es imposible que un objeto m tenga las mismas propiedades físicas Φ en dos tiempos diferentes t_1 y t_2 y no cuente con las mismas propiedades funcionales Ψ en t_1 y t_2 ; y
- (S4) Es imposible que un objeto m cambie de propiedades funcionales Ψ y permanezca con las mismas propiedades físicas Φ .

Lo argumentado muestra al funcionalismo como una teoría de carácter abstracto sobre lo mental – la teoría óptima, dirían algunos, para una visión computacional de la mente. Téngase en cuenta que la tesis de la realización (instanciación o ejemplificación)

múltiple subyace a estos razonamientos y al esquema de definición funcional (F)⁴. La vemos explícitamente formulada en (S2). Una propiedad funcional puede ser realizada (instanciada o ejemplificada) por múltiples propiedades concretas, no obstante ser éstas de diverso tipo o naturaleza. Ni el material del sustrato ni su estructura hacen de algo un carburador o una ratonera – tampoco al corazón o al riñón, de allí que podamos hablar de corazones artificiales y de riñones artificiales sin incurrir en errores conceptuales o empíricos; piénsese, además, en la multiplicidad de corazones y riñones que apreciamos como efecto de procesos evolutivos separados en el mundo animal. Los carburadores y las ratoneras son múltiplemente realizables (instanciables o ejemplificables). El ser un carburador o una ratonera es algo que se abstrae del sustrato. Lo mismo pasa con los estados mentales. Referirse a ellos supone una abstracción, una lógica de orden superior. Por eso, “si el funcionalismo es verdadero, las características más generales de la cognición [y el pensamiento] son independientes de la neurología”⁵ (Maloney, 1999, p. 333). Esta es precisamente la razón por la cual se dice que la teoría funcionalista de la mente se configura como la teoría óptima para una visión computacional de lo mental.

Los procesos computacionales pueden implementarse sobre múltiples sustratos físicamente disímiles y el funcionalismo no exige que los estados mentales estén atados a estados concretos específicos. Son múltiplemente realizables (instanciables o ejemplificables). Los estados mentales pueden, en consecuencia, ser vistos como estados computacionales, y así del mismo modo que *hardwares* de características completamente diferentes pueden implementar el mismo *software*, organismos biológica y físicamente diferentes pueden exhibir los mismos estados mentales, es decir, tener una misma psicología. Esto sucede ya que en uno y otro caso los estados son definidos (singularizados o individuados) en términos de un rol causal específico y no de un sustrato específico. No hay restricciones biológicas o físicas de tipo alguno.

Veamos ahora dos versiones del funcionalismo: el funcionalismo de máquina y el funcionalismo analítico.

4. Funcionalismo de máquina

⁴ Rabossi (2004) critica esta tesis aplicada al campo de la psicología.

⁵ Mi traducción.

Putnam (1960) (1967) populariza al funcionalismo apelando al concepto de máquina de Turing⁶. Su defensa de él es indirecta, pero expresada de manera categórica y con un efecto sumamente persuasivo. Marca un hito en la filosofía de la mente. Los argumentos que ofrece en su favor son, sobre todo, argumentos en contra del fisicalismo de tipos (teoría de la identidad psicofísica) y del conductismo lógico. Su crítica es frontal en relación con el primero. Señala sin remilgos que los estados mentales *no* son estados cerebrales. La razón sobre la que se apoya para negar tal identidad no es de índole *a priori* – tampoco nadie puede negar *a priori* al funcionalismo. Es decir, no critica al fisicalismo de tipos por incurrir en errores de carácter conceptual. No es necesariamente falso ni una contradicción. Tampoco es un sinsentido. La razón en contra de este fisicalismo es empírica: pasa sencillamente que hay una hipótesis más aceptable, a saber, que los estados mentales son estados funcionales. Tal es la tesis del funcionalismo. La idea de realización múltiple formulada en (S2) subyace a esta tesis y constituye por sí sola un argumento en contra del fisicalismo de tipos. Los estados mentales *no* son estados cerebrales, pues hay estados mentales como el hambre en los que, por ejemplo, pueden encontrarse tanto un humano como un pulpo, y la evidencia muestra que estos organismos, encontrándose en un mismo tipo de estado mental, no se encuentran en un mismo tipo de estado físico. Y si hay un mismo tipo de estado mental que tiene como correlato dos tipos diferentes de estados físicos, eso quiere decir que es falso que los estados mentales como tipos sean idénticos a los estados cerebrales como tipos. Un solo ejemplo basta para deponer al fisicalismo de tipos y posicionar al funcionalismo.

El rechazo funcionalista del conductismo lógico, mientras tanto, no es en el mismo grado frontal. Recuérdese que el funcionalismo en sus diferentes versiones puede ser visto como una suerte de conductismo mejorado. Entre otras razones porque, apelando a un lenguaje lógico-matemático y a descripciones físicas de *inputs* sensoriales y de *outputs* conductuales, satisface el deseo conductista de caracterizar lo mental en términos no mentales (véase la figura 3 y su correspondiente explicación). Además, es cierto que, siguiendo el estilo conductista, el funcionalismo entiende a los estados mentales en términos de sus relaciones con *inputs* sensoriales y *outputs* conductuales, pero, a diferencia del conductismo, suma a esta caracterización las relaciones que los estados mentales tienen con otros estados mentales. Los estados mentales son estados

⁶ Shagrir (2005) ofrece un interesante e informado recorrido de la asunción y el posterior abandono del funcionalismo de máquina por parte de Putnam.

reales y esta realidad, que es negada por el eliminativismo conductista, les permite ser causa efectiva de conducta y de otros estados mentales.

Debe señalarse que el objetivo de Putnam no es desarrollar ni verificar la tesis de que los estados mentales son estados funcionales. Ello sería una empresa abrumadora y de difícil consecución para un investigador en aislado, pues tendría que ofrecer una lista con las descripciones funcionales de cada uno de los estados mentales. Esa no es su tarea y se la delega a los psicólogos empíricos, quienes deben de recibirla agradecidos, pues con ella se les suministra un potente modelo de explicación de los estados mentales, uno que les puede dar trabajo por décadas y que puede tender puentes serios entre sus investigaciones y las que se llevan a cabo en las ciencias de la computación y los estudios sobre inteligencia artificial. Lo único que busca Putnam es mostrar la viabilidad o plausibilidad de la tesis funcionalista. La clave de esta formulación inicial del funcionalismo – la clave, en otras palabras, del funcionalismo de máquina – reside, como se ha observado, en el concepto matemático de máquina de Turing. Se usa este concepto para modelar a un organismo del que sea posible decir que tiene estados mentales. Nosotros, por ejemplo.

La máquina de Turing es descrita a través de una serie finita de estados discretos y una tabla de instrucciones (tabla de máquina) que, considerando *inputs* determinados, estipula el tránsito de un estado a otro con carácter de necesidad, así como la producción de *outputs* determinados. No hay nada más acerca de la máquina de Turing que las series finitas de estados, *inputs*, *outputs* y la tabla de máquina que relaciona causalmente los elementos anteriores. La tabla de máquina da cuenta de la organización funcional del sistema en cuestión. Es su programa. Esta descripción de estados mediante la tabla de máquina constituye una especificación implícita de cada estado. No nos dice cuál es la realización (instanciación o ejemplificación) física de ningún estado. Ello no constituye un defecto. Se trata más bien de una virtud. Pues da cuenta de una característica que responde a la naturaleza de los estados con los que estamos trabajando, que son estados funcionales, o sea, estados cuyo ser es un rol causal definido y no una composición o estructura determinadas. La descripción de los estados de máquina es completa a través de la tabla y excluye abiertamente una descripción de sus eventuales realizaciones (instanciaciones o ejemplificaciones) físicas.

Ilustremos estas ideas mostrando la tabla de máquina de un autómata detector de paridad dada por Block (1996). Este autómata cuenta con dos estados, E_1 y E_2 , un *input*, '1,' y dos *outputs*, emitir la expresión 'par' y emitir la expresión 'impar.' Esto quiere

decir que el autómata puede indicarnos si percibe un número par o impar de cifras 1. La tabla de máquina que expresa la organización funcional del autómata es la siguiente:

Estado		
<i>Input</i>	E ₁	E ₂
1	“Impar,” E ₂	“Par,” E ₁

Figura 1. Autómata detector de paridad.

Se aprecia que cuando el autómata se encuentra en el estado E₁ y recibe el *input* ‘1,’ emite la expresión ‘impar’ como *output* y pasa al estado E₂. Si estando en E₂ recibe el *input* ‘1,’ emite la expresión ‘par’ como *output* y pasa al estado E₁. Así, dado el caso de que reciba dos veces el *input* ‘1,’ dirá “par,” lo cual significa que ha percibido un número par de cifras 1. Esta tabla de máquina constituye el programa o modelo adecuado para un autómata detector de paridad. Los estados E₁ y E₂ son estados funcionales. La naturaleza de E₁ es ser mediador causal entre E₂ y los *outputs* posibles considerando el único *input* permitido. Asimismo, la naturaleza de E₂ es ser mediador causal entre E₁ y los *outputs* posibles considerando el único *input* permitido. Nada más hay que decir acerca de estos dos estados. Todo lo que son está capturado por la tabla de máquina recién ofrecida. Sus descripciones están completas. La lógica es de orden superior, como corresponde a toda descripción funcional. Para ser explícitos:

(E₁) Algún x está en E₁ = *def* Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe un *input* de ‘1,’ pasa a Q y emite la expresión “impar;” si x está en Q y recibe un *input* ‘1,’ entra en P y emite la expresión “par” & x está en P]

(E₂) Algún x está en E₂ = *def* Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe un *input* de ‘1,’ pasa a Q y emite la expresión “impar;” si x está en Q y recibe un *input* ‘1,’ entra en P y emite la expresión “par” & x está en Q]

Pensemos ahora en una tabla de máquina con dos *inputs* posibles. García-Carpintero (1995) brinda la organización funcional de un autómata expendedor de billetes de autobús. Adaptándola a nuestro caso, pensemos que la máquina vende billetes de autobús por el precio de S/. 1.00 cada uno. Puede encontrarse en dos estados, B₁ y B₂, recibir dos *inputs*, monedas de S/. 1.00 y monedas de S/. 0.5, y emitir dos *outputs*, un billete de autobús y monedas de S/. 0.5. La tabla de máquina es la siguiente:

Estado <i>Input</i>	B ₁	B ₂
S/. 1.00	Billete de autobús, B ₁	Billete de autobus y S/. 0.5, B ₁
S/. 0.5	No <i>output</i> , B ₂	Billete de autobús, B ₁

Figura 2. Autómata expendedor de billetes de autobús.

Se aprecia que cuando el autómata se encuentra en el estado B₁ y recibe el *input* de S/. 1.00, emite un billete de autobús como *output* y permanece en el estado B₁. Si estando en B₁ recibe el *input* S/. 0.5, no emite *output* alguno, pues no se ha completado

el precio del billete, y pasa al estado B_2 . Si se encuentra en B_2 y recibe el *input* de S/. 1.00, emite dos *outputs*, a saber, un billete de autobús y S/. 0.5 a manera de cambio o vuelto, y pasa al estado B_1 . Finalmente, si estando en B_2 recibe el *input* de S/. 0.5, emite como *output* un billete de autobús, pues se ha completado el precio del mismo, y pasa al estado B_1 . No hay nada más que decir acerca de la organización funcional de este autómata expendedor de billetes de autobús. La descripción de sus estados está completa. Nuevamente, la lógica de estas descripciones es de orden superior:

(B₁) Algún x está en $B_1 =_{def}$ Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe un *input* de S/. 1.00, permanece en P y emite un billete de autobús como *output*; si x está en P y recibe un *input* de S/. 0.5, pasa a Q y no emite *output*; si x está en Q y recibe un *input* de S/. 1.00, pasa P y emite un billete de autobús y S/. 0.5 como *output*; si x está en Q y recibe un *input* de S/. 0.5, pasa a P y emite un billete de autobús como *output* & x está en P]

(B₂) Algún x está en $B_2 =_{def}$ Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe un *input* de S/. 1.00, permanece en P y emite un billete de autobús como *output*; si x está en P y recibe un *input* de S/. 0.5, pasa a Q y no emite *output*; si x está en Q y recibe un *input* de S/. 1.00, pasa P y emite un billete de autobús y S/. 0.5 como *output*; si x está en Q y recibe un *input* de S/. 0.5, pasa a P y emite un billete de autobús como *output* & x está en Q]

Dos objetos de composición material diferente pueden cubrir con esta descripción y ser, así, realizaciones (instanciaciones o ejemplificaciones) de la misma máquina expendedora de billetes de autobús por tener la misma organización funcional. La originalidad del funcionalismo de máquina consiste en postular que los estados mentales pueden entenderse en términos de tablas de máquina como las mostradas en las figuras 1 y 2.

La idea, desde luego, no es que el uso de los términos mentales en la cotidianidad sea efectivamente así. El funcionalismo de máquina no considera que su propuesta sea formalmente correcta y empíricamente viable porque la gente en el día a día hable del dolor o las creencias pensando en tablas de máquina o descripciones funcionales. La intención no es capturar el uso cotidiano de los términos mentales. Esta teoría no es tributaria de la filosofía lingüística. Su objeto es, más bien, persuadirnos de la utilidad de entender los términos mentales a través de la tabla de máquina que describa las relaciones causales entre todos los estados mentales posibles – o sea, todos los estados que los términos mentales designan – junto con sus vínculos con *inputs* y

outputs específicos. Ello resultaría en un modelo mecánico materialmente adecuado de la mente, que sin recoger el significado cotidiano de los términos mentales calce con él. La particularidad de este modelo es que equivaldría a un programa o *software* de lo mental. Especificarlo, como ya se ha dicho, no fue nunca el objetivo de Putnam. Se presume que el desarrollo de esa tarea motive a los psicólogos y a quienes estén detrás de un marco teórico fructífero para los estudios e implementación de la inteligencia artificial⁷.

Demos una tabla de máquina mínima para lo mental. Está basada en Byrne (2011). Supongamos que podemos encontrarnos sólo en dos estados, M_1 y M_2 , recibir dos *inputs*, una piedra cayendo sobre el pie y un trozo de hielo colocado sobre el pie, y emitir dos *outputs*, las exclamaciones de ‘¡Au!’ y ‘¡Uh!’. Dadas estas condiciones, nuestra tabla de máquina es la siguiente:

Estado <i>Input</i>	M_1	M_2
Piedra cayendo sobre el pie	¡Au!, M_1	¡Au!, M_1

⁷ A propósito de esto, Heil (2000) hace una analogía entre los trabajos propios de programadores e ingenieros, por un lado, y los trabajos propios de psicólogos y neurocientíficos, por el otro. Así como el programador explica el *software* (aspecto lógico) y el ingeniero explica el *hardware* (aspecto físico), y sus explicaciones no compiten entre sí ya que responden a dos niveles diferentes de organización de una misma máquina, ocurre que el psicólogo explica la mente (aspecto lógico) y el neurocientífico explica el sistema nervioso (aspecto físico), y sus explicaciones no deben ser vistas como contrapuestas. Hablan, como en el caso anterior, de dos niveles diferentes de organización de una misma máquina.

Hielo colocado sobre el pie	¡Uh!, M ₂	No <i>output</i> , M ₂
--------------------------------------	-----------------------------	--

Figura 3. Teoría mínima de la mente.

Se aprecia que cuando nos encontramos en el estado M_1 y recibimos el *input* sensorial de una piedra cayendo sobre nuestro pie, exclamamos “¡Au!” como *output* conductual y permanecemos en el mismo estado M_1 . Si estando en M_1 recibimos el *input* sensorial de un trozo de hielo colocado sobre nuestro pie, exclamamos “¡Uh!” como *output* conductual y pasamos al estado M_2 . Si nos encontramos en M_2 y recibimos el *input* sensorial de una piedra cayendo sobre nuestro pie, exclamamos “¡Au!” como *output* conductual y pasamos al estado M_1 . Finalmente, si estando en M_2 recibimos el *input* sensorial de un trozo de hielo colocado sobre nuestro pie, no emitimos *output* conductual alguno y pasamos al estado M_1 . No hay nada más que decir acerca de los estados M_1 y M_2 . M_1 y M_2 son los estados mentales de dolor y alivio, respectivamente. Tales estados han quedado definidos por la tabla de máquina. Nada más queda por decir acerca de ellos desde un punto de vista funcionalista. Cualquier cosa que presente esta organización funcional contará con los estados mentales de dolor y alivio.

(M_1) Algún x se encuentra en estado de dolor (M_1) = *def* Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe el *input* sensorial de una piedra cayendo sobre su pie, permanece en P y exclama “¡Au!” como *output* conductual; si x está en P y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, pasa a Q y exclama “¡Uh!” como *output* conductual; si x está en Q y recibe el *input* sensorial de una piedra cayendo sobre su pie, pasa a P y exclama “¡Au!” como *output* conductual; si x está en Q y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, permanece en Q y no emite *output* conductual alguno & x está en P]

(M₂) Algún x se encuentra en estado de alivio (M₂) = *def* Ser un x tal que $\exists P \exists Q$ [si x está en P y recibe el *input* sensorial de una piedra cayendo sobre su pie, permanece en P y exclama “¡Au!” como *output* conductual; si x está en P y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, pasa a Q y exclama “¡Uh!” como *output* conductual; si x está en Q y recibe el *input* sensorial de una piedra cayendo sobre su pie, pasa a P y exclama “¡Au!” como *output* conductual; si x está en Q y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, permanece en Q y no emite *output* conductual alguno & x está en Q]

Lo dado en la figura 3 y en este último par de definiciones funcionales de términos mentales es un ejemplo de la caracterización funcional de los estados mentales. Repárese, no obstante, en el hecho de que los estados especificados por las tablas de máquina son estados totales. Eso significa que las máquinas modeladas sólo se encuentran en un estado en un tiempo dado, nunca en dos, tres o más. Dicha característica constituye una primera objeción a la pretensión de modelar de manera mecánica a la mente a partir del concepto de máquina de Turing, pues nuestra experiencia dice que un organismo con psicología puede encontrarse en más de un estado mental a la vez y que estos, por tanto, no son totales. Puedo, así, al mismo tiempo (1) ver una manzana, (2) sentir calor, (3) escuchar el canto de los pájaros, (4) creer que se ha hecho tarde para llegar al trabajo y (5) sentirme preocupado.

Otro punto de interés es que el organismo modelado por una máquina de Turing presenta determinismo en sus cambios de estados. Para evitar esta odiosa posición en el caso de que el modelado sea un organismo humano con presunta libertad, Putnam modifica el concepto de máquina de Turing de tal manera que el tránsito de un estado a otro no se estipule con carácter de necesidad, sino tan sólo en términos de probabilidad. Propone denominar a este nuevo concepto de máquina de Turing como autómata probabilista.

5. Funcionalismo analítico

David K. Lewis (1972) presenta una versión analítica del funcionalismo. Esta denominación responde al hecho de que en sus textos sugiera recurrir a una teoría cuyas definiciones de términos mentales – como ‘dolor’ o ‘alivio’ – son presuntas verdades analíticas. La teoría aludida es la psicología popular, una conjunción de proposiciones generales que, no obstante el paso de miles de años, permanece relativamente estable –

y es de esperar que continúe en dicho estado a futuro, a pesar de las denuncias en su contra formuladas por los esposos Paul Churchland (1981) y Patricia Churchland (1986). La analiticidad de estas proposiciones generales tiene como indicador el que se les considere perogrulladas u obviedades, tales como que si alguien *teme* que ocurra el evento *e*, entonces *desea* que no ocurra el evento *e*. ¿Quién puede estar en contra de afirmaciones como éstas? Ellas dan el significado de los términos mentales: temer algo significa desear que no ocurra ese algo. Su manejo nos muestra como hablantes competentes y, por lo tanto, son condición necesaria para la interpretación de los otros e incluso para el conocimiento de nosotros mismos.

El funcionalismo analítico de D. K. Lewis tiene por objeto proveer al fisicalismo de las herramientas técnicas que faciliten una próxima identificación teórica entre estados mentales y estados físicos. Es un medio, no un fin. La eventual identificación entre ‘dolor’ y ‘estimulación de las fibras-C’ no será consecuencia de postular leyes puente para simplificar la ciencia, identificando las entidades de las que se habla en una teoría con las entidades de las que se habla en otra. Este tipo de identificaciones serían postuladas, no descubiertas. Eso no resulta muy atractivo. Por lo contrario, D. K. Lewis augura que las identificaciones psicofísicas caerán lógicamente de dos premisas: las definiciones funcionales de términos mentales y los descubrimientos de la fisiología.

(1) Estado mental M = el ocupante del rol causal R

(2) Estado neuronal N = el ocupante del rol causal R

∴ Estado mental M = estado neuronal N

La premisa (1) es una definición funcional que da el significado del término mental ‘M,’ el cual designa al estado mental M, e.g, el término mental ‘dolor’ designa al estado mental dolor. La definición funcional del término mental ‘M’ no incluye término mental alguno, esto equivale a decir que el extremo derecho de definiciones como (1) no cuenta con ningún término mental. La premisa (2), por su lado, es un descubrimiento de la fisiología. La conclusión, que es una identificación psicofísica, cae deductivamente de las premisas (1) y (2) por el carácter transitivo de la identidad (‘=’). La contribución de D. K. Lewis a esta empresa de identificación viene dada por el suministro de la técnica que permita conseguir definiciones como las representadas por (1). Es preciso señalar que él mismo no da ninguna definición en esta línea. Sólo nos dice cómo hacerlas. Con su técnica, dicho sea de paso, ayuda en cierto modo al conductismo, aunque ciertamente no tanto como a la concepción fisicalista. Una manera de expresar esto es sostener que, mientras su agenda es fisicalista, su inspiración es

conductista⁸. Ayuda al conductismo porque, al ofrecer un *definiendum* carente de términos mentales, evita las circularidades a las cuales daban lugar las definiciones conductistas.

La técnica empleada ha venido a llamarse “técnica Lewis-Ramsey”⁹. Tiene como base la pauta ofrecida por Ramsey (1929) para describir una teoría con un lenguaje que denote aquellas entidades cuyas relaciones entre sí esa misma teoría dice explicar. Esta técnica ayuda sobremanera cuando la teoría en cuestión introduce términos que nombran entidades inobservables, tales como ‘electrón’ o ‘dolor.’ La pretensión que hay detrás es una pretensión del positivismo lógico, como bien anota Weatherson (2014). El objeto es reducir el discurso sobre inobservables a un discurso sobre observables. Para ello, los términos introducidos por cierta teoría – en nuestro caso, la psicología popular – son definidos echando mano de un lenguaje ya comprendido – en nuestro caso, uno que involucre términos sobre *inputs* sensoriales y *outputs* conductuales, esto es, un lenguaje físico-conductual. La técnica Lewis-Ramsey permite extraer la denotación de los términos mentales introducidos por la psicología popular. Podemos hablar de estas denotaciones sin apelar a término mental alguno (lo que tiene el doble mérito de no apelar a inobservables y de no incurrir en definiciones circulares) y empleando únicamente un lenguaje físico-conductual.

Demos un paradigma para la ramseyficación de teorías.

Considérese que se propone una teoría T y que ésta introduce los términos ‘ $t_1 \dots t_n$.’ Tales términos son nombres, esto es, términos que denotan entidades. En el vocabulario de T también hay términos que no son introducidos por T. Estos otros términos son los términos ‘ $o_1 \dots o_n$.’ La teoría T se formula a través de una oración conjuntiva:

$$(T) T(t_1 \dots t_n, o_1 \dots o_n)$$

(T) dice que las entidades denotadas por los términos ‘ $t_1 \dots t_n$ ’ ocupan ciertos roles causales. También dice de estas entidades que se encuentran causalmente vinculadas con aquellas otras entidades denotadas por los términos del vocabulario que no han sido introducidos por T, esto es, por los términos ‘ $o_1 \dots o_n$.’ (T) dice por último que las entidades denotadas por los términos ‘ $t_1 \dots t_n$ ’ guardan un vínculo causal entre sí. Ahora bien, la matriz de T se obtiene reemplazando los términos ‘ $t_1 \dots t_n$ ’ que aparecen en (T) por variables libres:

⁸ Cf. Schwarz (2015).

⁹ Cf. Shoemaker (1981).

$$\text{(Matriz T)} T(x_1 \dots x_n, o_1 \dots o_n)$$

D. K. Lewis señala con pertinencia que cualquier n-tupla que satisfaga esta fórmula (es decir, cualquier conjunto de entidades cuyos nombres reemplacen las variables libres y den como resultado únicamente oraciones verdaderas) debe ser considerada una realización (instanciación o ejemplificación) de la teoría T. De este modo, si ‘T’ representa a la psicología popular, es posible que una n-tupla constituida por parte de nuestros estados neuronales satisfaga la matriz de T. Si esto es así, dichos estados neuronales constituyen una realización de la psicología popular.

Ahora procedamos a ramseyficar T, de modo que se pueda extraer las denotaciones de los términos ‘ $t_1 \dots t_n$ ’ y hablar de ellas empleando únicamente los términos (el lenguaje) que no ha introducido T, esto es ‘ $o_1 \dots o_n$ ’.

$$\text{(Ramsey T)} \exists x_1 \dots \exists x_n T(x_1 \dots x_n, o_1 \dots o_n)$$

La oración Ramsey de T se obtiene tras una generalización existencial de las variables libres que contiene la matriz de T. Significa literalmente que la teoría T tiene por lo menos una realización (instanciación o ejemplificación). La oración Carnap de T – o el condicional Carnap de T – se obtiene, a su vez, estableciendo una relación condicional entre la formulación de T y la oración Ramsey de T. Este condicional tiene a la oración Ramsey de T como antecedente y a la formulación de T como consecuente.

$$\text{(Carnap T)} \exists x_1 \dots \exists x_n T(x_1 \dots x_n, o_1 \dots o_n) \rightarrow T(t_1 \dots t_n, o_1 \dots o_n)$$

La oración Carnap de T señala que si la teoría T tiene por lo menos una realización (instanciación o ejemplificación), entonces la teoría T es verdadera. Esto es, si la teoría T tiene por lo menos una realización (instanciación o ejemplificación), entonces los términos ‘ $t_1 \dots t_n$ ’ denotan una serie de entidades que se encuentran en las relaciones causales especificadas por (T).

De lo expuesto, inferimos la virtud de la ramseyficación de T. Gracias a este procedimiento los términos ‘ $t_1 \dots t_n$ ’ introducidos por T son eliminados y, no obstante esto, podemos hablar de las entidades denotadas por esos mismos términos. Ello quiere decir que la oración Ramsey de T captura el significado de los términos ‘ $t_1 \dots t_n$ ’ introducidos por T. Retiene el significado que T atribuye a sus términos, aunque, al mismo tiempo, prescindamos de ellos. Este significado no consiste en otra cosa que en las relaciones causales que T establece entre las entidades denotadas por los términos ‘ $t_1 \dots t_n$ ’ y las entidades denotadas por los términos ‘ $o_1 \dots o_n$ ’. Es decir, por lo especificado en (T). La especificación de estas relaciones causales se retiene en la oración Ramsey de T. De allí que se afirme que el significado de los términos ‘ $t_1 \dots t_n$ ’ se retenga y que las

entidades que estos términos denotan se extraigan. Nos quedamos con ellas, a pesar de haber eliminado los términos que T introdujo para denotarlas. Lo que queríamos era precisamente prescindir de tales términos y quedarnos con el significado que les asignaba T. Este deseo se ha visto satisfecho. Gracias a la oración Ramsey de T podemos hablar de las entidades cuyas relaciones explica T olvidándonos de los términos introducidos por T.

Apliquemos lo dicho a términos mentales como ‘dolor’ o ‘alivio.’ La teoría que los introduce es la psicología popular¹⁰. Esta psicología también dispone de términos que no introduce. Nos referimos específicamente a términos que denotan *inputs* sensoriales y *outputs* conductuales. Digamos que son términos físico-conductuales como ‘descarga eléctrica de 220V’ o ‘grito.’ Estableciendo el paralelo con la exposición paradigmática de la ramseyficación, los términos mentales son los términos ‘ $t_1 \dots t_n$ ’ y los términos físico-conductuales son los términos ‘ $o_1 \dots o_n$.’ El significado que la psicología popular confiere a los términos mentales se reduce a roles causales. Entendemos que la formulación de la psicología popular es una gran conjunción de oraciones en donde se especifica las relaciones causales que tienen entre sí las entidades denotadas por los términos mentales, así como las relaciones causales que tienen estas mismas entidades con aquellas otras entidades denotadas por los términos físico-conductuales. A juicio de D. K. Lewis, esta gran conjunción de oraciones está constituida únicamente por verdades que todos conocen, verdades que todos conocen que todos conocen, verdades que todos conocen que todos conocen que todos conocen, etc. Es decir, por verdades que ningún hablante competente está dispuesto a negar. De allí la tentación de decir que son verdades analíticas.

Siguiendo con el procedimiento de ramseyficación, reemplazaremos los términos mentales por variables libres, que luego ligaremos mediante cuantificadores existenciales. De este modo, llegaremos a hablar de las entidades denotadas por los términos mentales ‘ $t_1 \dots t_n$ ’ empleando únicamente los términos físico-conductuales ‘ $o_1 \dots o_n$.’ Los significados que la psicología popular daba a sus términos mentales se retienen. No hay pérdida semántica. Pues estos significados no consistían sino en roles causales específicos y la referencia a dichos roles continúa presente en la oración Ramsey de la psicología popular. Aunque hayamos sido desprovistos de todos los términos mentales, la red constituida por los significados de todos los términos mentales

¹⁰ Si la teoría a ser ramseyficada no es la psicología popular, sino la psicología científica, entonces lo expuesto se llamaría psicofuncionalismo y no funcionalismo analítico.

no se ha perdido. La oración Ramsey termina de enseñarnos que las entidades denotadas por los términos mentales son estados funcionales. El estado mental es uno y el mismo con el estado funcional. Y ésta es la tesis del funcionalismo.

Para hacer más didáctico el asunto, empleemos el estilo de ramseyficación de la psicología popular propuesto por Block (1978), y que es recogido por Shoemaker (1981) y Kim (1996), entre otros.

$$(P) P(E_1 \dots E_n, I_1 \dots I_k, O_1 \dots O_m)$$

(P) es la formulación de la psicología popular. Consta, como se ha dicho, de una gran conjunción de oraciones que especifican las relaciones causales que tienen entre sí los estados mentales, así como las relaciones causales que tienen los estados mentales con los *inputs* sensoriales y los *outputs* conductuales. Para estos efectos, los términos ‘ $E_1 \dots E_n$ ’ son términos que denotan estados mentales, los términos ‘ $I_1 \dots I_k$ ’ son términos que denotan *inputs* sensoriales y los términos ‘ $O_1 \dots O_m$ ’ son términos que denotan *outputs* conductuales. Todos estos términos son constantes. La matriz de la psicología popular se obtiene reemplazando los términos ‘ $E_1 \dots E_n$ ’ por las variables libres ‘ $F_1 \dots F_n$:’

$$(\text{Matriz } P) P(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m)$$

Recordando el carácter de una de las observaciones de D. K. Lewis referidas líneas más arriba, cualquier n-tupla que satisfaga esta matriz será una realización de la psicología popular. Las variables ‘ $F_1 \dots F_n$ ’ pueden reemplazarse con éxito en nuestro caso humano por términos que hagan referencia a estados neuronales, digamos, por los términos ‘ $N_1 \dots N_n$.’ Estos estados son una realización de la psicología popular. Pero las mismas variables ‘ $F_1 \dots F_n$ ’ podrían ser reemplazadas con igual éxito por términos que hagan referencia no a estados de nuestro cerebro, sino a estados del *hardware* de cierta computadora o a estados del cerebro de un extraterrestre cuya química no sea la del carbono. Si esto es así, esos sistemas también serían realizaciones de la psicología popular. Desde luego, para un fisicalista de tipos (no de casos) nosotros somos la única realización de la psicología popular.

Ahora bien, la oración Ramsey de la psicología popular es consecuencia de realizar una de serie de generalizaciones existenciales en relación con cada una de las variables libres de la matriz de dicha teoría:

$$(\text{Ramsey } P) \exists F_1 \dots \exists F_n P(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m)$$

Nótese que no se cuantifica sobre los términos físico-conductuales, pues estos son constantes y no variables. Esta oración afirma que hay por lo menos una realización

(instanciación o ejemplificación) de la psicología popular y nos permite hablar de las entidades que esta teoría dice explicar sin emplear los términos que ella misma introduce para tales efectos. O sea, gracias a la oración Ramsey de la psicología popular podemos hablar sobre las entidades denotadas por los términos mentales sin emplear términos mentales. Podemos hablar del dolor sin emplear el término ‘dolor.’ Lo hacemos en un lenguaje puramente físico-conductual refiriéndonos a cierto estado funcional singularizado por tal y cual rol causal específico en relación con tales y cuales *inputs* sensoriales específicos, tales y cuales *outputs* conductuales específicos y tales y cuales estados funcionales específicos adicionales. Este rol causal específico es lo que singulariza al estado funcional del caso, lo que le da una identidad definida y le confiere, en consecuencia, la condición de entidad¹¹. Esta entidad es lo que nuestra psicología popular por razones puramente contingentes designa con el nombre de ‘dolor.’ Está claro que se trata de una entidad de segundo orden, pues su identidad le ha sido conferida por un rol causal específico y no por una constitución (o estructura) física determinadas. No hay nada más que decir acerca del dolor, nada más allá de ese rol causal. La comprensión de este hecho nos permite quedarnos con las explicaciones de la psicología popular aunque nos desprendamos de su lenguaje sobre inobservables. La naturaleza de estos inobservables no tiene un sustrato definido. Su naturaleza es un rol causal.

Pasar a dar cuenta de las mismas explicaciones de la psicología popular en un lenguaje íntegramente observacional. Tal es la virtud de las oraciones Ramsey.

Inferimos, por otro lado, que si como afirma la oración Ramsey de la psicología popular, hay por lo menos una realización de esta teoría, entonces la psicología popular es verdadera. Lo que se representa por la oración Carnap de P:

$$(Carnap P) \exists F_1 \dots \exists F_n P(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m) \rightarrow P(E_1 \dots E_n, I_1 \dots I_k, O_1 \dots O_m)$$

A efectos de brindar un esquema de definición de un término mental a partir de la técnica expuesta, piénsese que en el contexto de (P) en lugar del término mental ‘E₁’ tenemos al término mental familiar ‘dolor.’ De este modo, la variable ‘F₁’ reemplaza a ‘dolor’ en (Matriz P). Lo mismo ocurre en (Ramsey P). Esto basta para ofrecer una definición funcional del término ‘dolor’ que prescindiera de cualquier término mental. La definición es la siguiente:

¹¹ A propósito de este razonamiento, recuérdese el eslogan de Quine (1958) “no entidad sin identidad.” También véase Braddon-Mitchell & Jackson (2007).

(Dolor) Algún x tiene (se encuentra en estado de) dolor = *def* ser un x tal que $\exists F_1 \dots \exists F_n T[(F_1 \dots F_n, I_1 \dots I_k, O_1 \dots O_m) \& x \text{ está en } F_1]$

El extremo izquierdo de esta definición con el uso del término ‘dolor’ expresa un estado mental y el extremo derecho con el uso de la conjunción ‘(F₁... F_n, I₁... I_k, O₁... O_m)’ expresa un estado funcional. Se trata, no obstante, de dos modos de referirse a una y la misma cosa. Exhiben diferentes sentidos, pero la misma denotación; no en vano la relación de identidad (=). El estado mental es numéricamente idéntico al estado funcional. El dolor es un rol causal.

Adaptemos la figura 3 dada en la sección anterior a los fines presentes. Allí se nos ofrecía una teoría mínima acerca del dolor. Podemos formularla como sigue:

(D) D(dolor, alivio, piedra cayendo sobre el pie, hielo colocado sobre el pie, ¡Au!, ¡Uh!)

Siendo más específicos con las relaciones causales de las que da cuenta la teoría mínima del dolor:

(D’) D(Para cualquier x, si x tiene dolor y recibe el *input* sensorial de una piedra cayendo sobre su pie, x sigue teniendo dolor y exclama “¡Au!” como *output* conductual; si x tiene dolor y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, x pasa a tener alivio y exclama “¡Uh!” como *output* conductual; si x tiene alivio y recibe el *input* sensorial de una piedra cayendo sobre su pie, x pasa a tener dolor y exclama “¡Au!” como *output* conductual; si x tiene alivio y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, x sigue teniendo alivio y no emite *output* conductual alguno)

Reemplazando los términos mentales ‘dolor’ y ‘alivio’ por las variables predicativas ‘F₁’ y ‘F₂,’ respectivamente, la oración Ramsey de la teoría es:

(Ramsey D’) $\exists F_1 \exists F_2$ D(Para cualquier x, si x está en F₁ y recibe el *input* sensorial de una piedra cayendo sobre su pie, x sigue estando en F₁ y exclama “¡Au!” como *output* conductual; si x está en F₁ y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, x pasa a estar en F₂ y exclama “¡Uh!” como *output* conductual; si x está en F₂ y recibe el *input* sensorial de una piedra cayendo sobre su pie, x pasa a estar en F₁ y exclama “¡Au!” como *output* conductual; si x está en F₂ y recibe el *input* sensorial de un trozo de hielo colocado sobre su pie, x sigue estando en F₂ y no emite *output* conductual alguno)

En consecuencia la definición de dolor es:

(Dolor D') Algún x tiene (se encuentra en estado de) dolor = *def* $\exists F_1 \exists F_2 D[(\text{Para cualquier } x, \text{ si } x \text{ está en } F_1 \text{ y recibe el } \textit{input} \text{ sensorial de una piedra cayendo sobre su pie, } x \text{ sigue estando en } F_1 \text{ y exclama "¡Au!" como } \textit{output} \text{ conductual; si } x \text{ está en } F_1 \text{ y recibe el } \textit{input} \text{ sensorial de un trozo de hielo colocado sobre su pie, } x \text{ pasa a estar en } F_2 \text{ y exclama "¡Uh!" como } \textit{output} \text{ conductual; si } x \text{ está en } F_2 \text{ y recibe el } \textit{input} \text{ sensorial de una piedra cayendo sobre su pie, } x \text{ pasa a estar en } F_1 \text{ y exclama "¡Au!" como } \textit{output} \text{ conductual; si } x \text{ está en } F_2 \text{ y recibe el } \textit{input} \text{ sensorial de un trozo de hielo colocado sobre su pie, } x \text{ sigue estando en } F_2 \text{ y no emite } \textit{output} \text{ conductual alguno}) \& x \text{ está en } F_1]$

Referencias

- Block, N. (1978/1980). Troubles with functionalism. En su *Readings in philosophy of psychology*, Vol. I (pp. 268-305). Cambridge, Massachusetts: Harvard University Press.
- Block, N. (1996). What is functionalism? En D. M. Borchert (Ed.), *The encyclopedia of philosophy supplement*. URL: <http://cogprints.org/235/1/199712008.html>
[Consulta: 19 de enero de 2016].
- Braddon-Mitchell, D. & Jackson F. (2007). *Philosophy of mind and cognition*. Oxford: Blackwell.
- Byrne, A. (2011). Functionalism [Material 8 del curso "Minds and machines"]. Departamento de Lingüística y Filosofía, Massachusetts Institute of Technology. URL: http://ocw.mit.edu/courses/linguistics-and-philosophy/24-09-minds-and-machines-fall-2011/study-materials/MIT24_09F11_functionalism.pdf
[Consulta: 25 de enero de 2016].
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2), 67-90.
- Churchland, P.S. (1986). *Neurophilosophy*. Cambridge, Massachusetts: The MIT Press.
- García-Carpintero, M. (1995). El funcionalismo. En F. Broncano (Ed.), *La mente humana*, Enciclopedia Iberoamericana de Filosofía, Vol. 8 (pp. 43-76). Madrid: Trotta.
- Heil, J. (2000). *Philosophy of mind*. Londres: Routledge.
- Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.

- Kim, J. (1996). *Philosophy of mind*. Boulder: Westview.
- Levin, J. (2013). Functionalism. En E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. URL: <http://plato.stanford.edu/archives/fall2013/entries/functionalism/> [Consulta: 30 de enero de 2016].
- Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 3(50), 249-258.
- Maloney, J. C. (1999). Functionalism. En R. A. Wilson & F. C. Keil, *The MIT encyclopedia of the cognitive sciences* (pp. 332-335). Cambridge, Massachusetts: The MIT Press.
- Moya, C. (2006). *Filosofía de la mente*. Valencia: Universitat de València.
- Putnam, H. (1960). Minds and machines. En S. Hook (Ed.), *Dimensions of mind* (pp. 148-180). Nueva York: New York University Press.
- Putnam, H. (1967). Psychological predicates. En W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37-48). Pittsburgh: University of Pittsburgh Press.
- Quine, W. V. O. (1958/2002). Hablando de objetos. En su *La relatividad ontológica y otros ensayos* (pp. 13-41). Madrid: Tecnos.
- Rabossi, E. (2004). Acerca de la autonomía de la psicología y de la universalidad de sus tipos básicos. En R. A. Martins, L. A. C. P. Martins, C. C. Silva & J. M. H. Ferreira (Eds.), *Filosofia e história da ciência no Cone Sul: 3º Encontro* (pp. 159-164). Campinas: AFHIC.
- Ramsey, F. P. (1929/1978). Theories. En su *Foundations: Essays in philosophy, logic, mathematics and economics* (pp. 101-125). Londres: Routledge & Kegan Paul Ltd.
- Schwarz, W. (2015). Analytic functionalism. En B. Loewer & J. Schaffer (Eds.), *A companion to David Lewis* (pp. 504-518). Oxford: Wiley Blackwell.
- Shagrir, O. (2005). The rise and fall of computational functionalism. En Y. Ben-Menahem (Ed.), *Hilary Putnam* (pp. 220-250). Cambridge: Cambridge University Press.
- Shoemaker, S. (1981). Some varieties of functionalism. *Philosophical Topics*, 1(12), pp. 93-119.
- Weatherson, B. (2014). David Lewis. En E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. URL: <http://plato.stanford.edu/archives/win2014/entries/david-lewis/> [Consulta: 30 de enero de 2016]