

ATRIBUCIONES DE AGENCIA MENTAL Y EL DESAFÍO DESDE LA PSICOPATOLOGÍA*

*Pablo López-Silva***

<https://orcid.org/0000-0001-7457-7724>

pablo.lopez.silva@gmail.com

RESUMEN *Una atribución de agencia mental es definida como el acto de asignar el inicio o autoría de un pensamiento en primera persona a un agente específico, esto, con el fin de generar grados suficientes de control sobre nuestra vida cognitiva. Si bien esta estrategia es fundamental para distinguir diferentes tipos de estados cognitivos – tales como deliberaciones, razonamientos, juicios, entre otros –, muchos autores han indicado que también es clave para entender experiencias cognitivas psicopatológicas en psicosis. Este artículo revisa la forma en que la estructura subjetiva de los delirios de inserción de pensamiento puede ser utilizada como una estrategia para evaluar los méritos argumentativos de las teorías que intentan explicar la arquitectura y naturaleza de las atribuciones de agencia mental. Luego de evaluar algunos aspectos fundamentales de las dos teorías dominantes en la literatura actual, se sugiere que, a pesar de sus importantes aportes al desarrollo de la discusión, tales enfoques no poseen los recursos explicativos suficientes para lidiar con el desafío propuesto.*

Palabras claves *Atribución de agencia mental, fenomenología cognitiva, delirios, psicosis, psicopatología.*

* Artículo recibido el 19/10/2019. Aceptado el 04/12/2019. Quisiera agradecer a Joel Smith, Tim Bayne, Tom McClelland y a Jöelle Proust por los comentarios provistos en versiones preliminares de este trabajo. Algunas ideas contenidas en este trabajo fueron presentadas en el Instituto Jean Nicod (Francia), Universidad de Warwick (Reino Unido), Universidad de Valparaíso (Chile) y Universidad de Génova (Italia). La escritura final de este trabajo se realizó en el marco del Proyecto FONDECYT No 11160544 'La Arquitectura Agencial del Pensamiento Humano' otorgado por la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) del Gobierno de Chile.

** Universidad de Valparaíso. Valparaíso, Chile. Instituto Milenio para la Investigación en Depresión y Personalidad (MIDAP). Chile.

ABSTRACT *An attribution of mental agency is defined as the act of assigning the initiation or authorship of a first personal thought to a specific agent in order to generate sufficient degrees of control over our own cognitive life. Although this strategy is crucial to distinguish different types of cognitive states – such as deliberations, reasonings, judgements, among others –, a number of authors have suggested that it is also crucial to understand psychopathological cognitive experiences in psychosis. This paper explores the way in which the subjective structure of thought insertion can be used as a strategy to evaluate the merits of theories aiming at explaining the nature and architecture of attributions of mental agency. After examining some of the most fundamental aspects of the two dominant theories in current literature, I suggest that, despite the contributions made by such theories to the discussion, none of them is able to face the challenge in a plausible way.*

Keywords *Attributions of mental agency, cognitive phenomenology, delusions, psychosis, psychopathology.*

1. Atribuciones de agencia: desde lo motor hacia lo mental

Existen al menos dos formas en las que los seres humanos podemos devenir conscientes de nuestros movimientos corporales (Daprati et al. 1997; Blackemore y Frith, 2003; Gallagher, 2007; Bayne, 2011). Por un lado, uno puede ser consciente de su propio cuerpo moviéndose de forma *voluntaria*, por ejemplo, cuando estamos nadando. Por lo general, este tipo de experiencias son fenomenológicamente complejas ya que van acompañadas de variaciones en las sensaciones de control, esfuerzo, voluntariedad, complejos monitoreos motores, entre muchos otros elementos (Bayne, 2008; Pacherie, 2008). Por otra parte, uno puede devenir consciente de su propio cuerpo moviéndose involuntariamente al momento de, por ejemplo, ser empujado por un oponente mientras rugamos *rugby*. Investigadores en filosofía, psiquiatría y ciencias cognitivas han convenido que aquello que distingue el primer modo de conciencia corporal del segundo es la existencia de una autoatribución de agencia motora en el caso voluntario.¹ Una ‘autoatribución de agencia motora’

1 Este concepto ha recibido distintas etiquetas en la literatura: ‘atribuciones de agencia’ (Stephens y Graham, 2000), ‘sentido de agencia’ (De Vignemont y Fournieret, 2004; Gallagher, 2000, 2014; Peacocke, 2003); ‘conciencia de control corporal’ (Mandik, 2010), y ‘conciencia de acción’ (Prinz, 2007). El asunto relevante en este contexto es que todos estos autores hacen referencia explícita a la existencia de un tipo de atribución – implícita o explícita – que distingue la conciencia de los movimientos corporales voluntarios de los involuntarios.

referiría al acto de atribuir el inicio del movimiento corporal en cuestión a uno mismo y, por lo tanto, experimentarse uno mismo como el autor o agente de tal movimiento. Mientras el acto de nadar sería experimentado como teniendo un inicio voluntario, el ser empujado no poseería tal componente (Gallagher, 2007; Proust, 2009; Bayne, 2011).²

Ahora bien, transportando este debate al caso de la conciencia que tenemos de nuestros ‘propios pensamientos’, diversos autores han sugerido que es posible también hablar de ‘atribuciones de agencia mental’, estas, definidas como el acto de atribuir el inicio o ‘autoría’ de un pensamiento al cual uno tiene acceso en 1^{era} persona a un agente específico con el fin de producir grados suficientes de control sobre nuestra vida cognitiva (Frith, 1992; Graham y Stephens, 1994; Campbell, 1999, 2002; Stephens y Graham, 2000; Gallagher, 2000, 2014; Proust, 2009; Vosgerau y Voss, 2014).³ Si bien parece natural pensar desde un punto de vista experiencial que este tipo de atribuciones tendería a ser autorreferencial en el caso de los pensamientos [yo soy el autor de mis pensamientos], este no siempre parece ser el caso, es más, las discusiones específicas sobre la naturaleza y arquitectura de las atribuciones de agencia mental se han nutrido principalmente del estudio filosófico y empírico de síntomas psicóticos que parecen implicar diversas alteraciones en la forma en que sujetos identifican al agente de algunos de sus propios pensamientos (Stephens y Graham, 2000; Campbell, 1999, 2002; Gallagher, 2000, 2014; López-Silva, 2016, 2017, 2018). Los fenómenos más discutidos en este contexto han sido los ‘delirios de control del pensamiento’, donde pacientes indican que un agente externo controla la creación de sus pensamientos, y los ‘delirios de inserción de pensamiento’, donde sujetos reportan que un agente externo de distinta naturaleza (humanos, mesas, animales, etc.) ha puesto ciertos pensamientos en su cabeza (Frith, 1992; Mullins y Spence, 2003); ‘el robo del pensamiento’, ‘los delirios de exposición del pensamiento’ y los ‘pensamientos obsesivos’ (O’Brien y Soteriou, 2009; Martin y Pacherie, 2013; Bayne, 2013), todos ejemplos de alteraciones en la sensación del ser el autor y estar en control de la vida cognitiva privada.

2 Es importante notar que tal distinción no conlleva a la diferencia entre ‘algo que yo hago’ y algo que ‘otro hace’ ya que movimientos como los tics o los espasmos no conllevan conciencia de un agente desde un punto de vista experiencial. Es más, la atribución de un movimiento corporal a un otro (en ausencia de ese otro actuando causalmente sobre el propio cuerpo) es un rasgo característico de ciertos delirios con base corporal (Ver nota 3; Stephens y Graham, 2000).

3 El debate analizado en este artículo depende fuertemente de algunos elementos del debate acerca de la naturaleza experiencial de los pensamientos i.e. la fenomenología cognitiva. Si bien esta discusión está actualmente abierta, nuestra examinación acepta la idea de que los pensamientos poseen algún tipo de *fenomenalidad* independiente de las descripciones de su origen. Para mayor información sobre este debate, ver: Bayne y Montague (2013).

Una de las ideas fundamentales en el seno de esta discusión es que una teoría sobre las atribuciones de agencia mental que logre explicar las instanciaciones patológicas y no-patológicas del fenómeno sería preferible a una que solo logra darle sentido a sus instancias no patológicas (o no patológicos). Por lo tanto, la existencia de alteraciones en el proceso de atribución de agencia mental es tomada en la literatura como una oportunidad para evaluar los méritos argumentativos de las teorías que intentan explicar el fenómeno (Bayne, 2013; López-Silva, 2014). A esta estrategia quisiera llamarle ‘el desafío desde la psicopatología’.

Luego de evaluar algunos aspectos fundamentales de las dos teorías de agencia mental dominantes en la literatura actual, en este artículo sugiero que, a pesar de sus importantes aportes al desarrollo de la discusión en filosofía y psiquiatría, tales teorías no son capaces de lidiar con el desafío desde la psicopatología, sobre todo cuando intentan aplicar sus recursos explicativos al caso de los delirios de inserción de pensamiento, y por lo tanto, concluyo, la discusión se encuentra abierta a la construcción de nuevas alternativas explicativas.

2. Entendiendo las atribuciones de agencia mental

Las actuales teorías que intentan darle sentido a las atribuciones de agencia mental intentan explicar, por una parte, su naturaleza, y por otra, su arquitectura. Por naturaleza, estas teorías refieren fundamentalmente al origen del contenido del estado atribucional, mientras que, por arquitectura, refieren a la forma en que los diversos procesos implicados en la producción del estado atribucional se organizan dependiendo de su naturaleza. En la literatura actual se distinguen dos propuestas fundamentales, a saber, la teoría *top-down* con arquitectura cognitiva y la *bottom-up* con arquitectura experiencial.⁴

4 Quisiera hacer tres precisiones respecto de este punto. Primero, las dos teorías que se revisarán en este trabajo son las más populares en la literatura actual en filosofía de la mente y psicopatología, pero no son las únicas. Un análisis de todas las teorías disponibles es un proyecto que no puede ser contenido en un solo artículo. Segundo, si bien en la actualidad las teorías *top-down* son cognitivas y las *bottom-up* son experienciales, queda abierta la posibilidad de elaborar potenciales propuestas híbridas o cíclicas, por lo que tales categorías quedan abiertas a reestructuración. En tercer lugar, una de las principales interrogativas en ese contexto es la pregunta por el ‘autor de los propios movimientos corporales/pensamientos’. Esta pregunta posee dos dimensiones. Para entenderlas, se deben distinguir aquellas premisas *necesariamente verdaderas* sobre la mente (metafísica de la mente) de aquellas que serían *experiencialmente verdaderas* (naturaleza experiencial de la mente humana). Si bien, *necesariamente* es verdad que uno es el autor de sus propios movimientos corporales, esto no es siempre el caso al nivel experiencial ya que existen instancias en las cuales sujetos reportan ciertos movimientos corporales como si no fuesen originados por ellos mismos a voluntad (espasmos, tics, etc.) o incluso, como si fuesen generados por otros agentes que no existen

2.1. Naturaleza sensorial y arquitectura *bottom-up*

Imitando la estructura explicativa de la teoría dominante en el debate sobre las atribuciones de agencia motora, los defensores de la teoría *bottom-up* (o ‘desde abajo hacia arriba’) de las atribuciones de agencia mental proponen que la naturaleza del fenómeno es primordialmente experiencial. En esta teoría, una atribución de agencia mental surge de la mera adopción del contenido experiencial de un pensamiento consciente de primer orden al cual se tiene acceso en primera persona (Gallagher, 2000, 2007, 2014; Zahavi, 2005; Gallagher y Zahavi, 2008). Por esto, esta teoría también podría ser denominada teoría tipo *endorsement* ya que la totalidad del contenido del estado atribucional es adoptado desde un estado sensorial de primer orden. Dentro de esta teoría, un pensamiento es autoatribuido agencialmente simplemente porque el pensamiento es experimentado como algo que uno ha creado. Así, esta teoría asume que la fenomenología de los pensamientos es fundamentalmente activa por lo que la dirección causal de tal atribución en términos arquitectónicos va desde la experiencia (*bottom-abajo*) hacia la atribución (*up-arriba*). Esta teoría no necesariamente niega la existencia de juicios cognitivos más robustos acerca de la agencia de un pensamiento, sino que, como señala Gallagher (2007, p. 348), esta teoría señala que: “atribuciones de propiedad o agencia de segundo nivel dependerán de una experiencia de propiedad o agencia de primer nivel” (traducción del autor). Para los defensores de teorías del estilo *bottom-up*, habría algo especial en la fenomenología de los pensamientos que nos haría atribuirlos a nosotros mismos en términos de agencia; este componente clave sería una ‘sensación de agencia mental’ (*sense of mental agency*) contenida en la estructura fenoménica de aquellos pensamientos que serían finalmente autoatribuidos de forma agencial (De Hann y De Bruin, 2010).⁵ Por lo tanto, se asume que cada pensamiento (que luego será autoatribuido) representa al sujeto como autor del mismo, i.e. el autor es parte del contenido fenoménico de un pensamiento de primer orden (Zahavi, 2005, 2018).

(delirios de control externo). Este proyecto se enmarca en las discusiones sobre la naturaleza experiencial de la mente humana. Para una revisión más profunda de este asunto, revisar: O'Brien y Soteriou (2009).

5 Es importante señalar que, si bien existirían diversas sensaciones de primer nivel asociadas a la actividad cognitiva - tales como la sensación de fluidez, control, integración, coherencia, entre otras - las teorías *bottom-up* dominantes exigen como criterio específico la existencia de una sensación de agencia específica que nutre juicios de agencia mental de segundo orden (ver por ejemplo, Gallagher, 2000 y 2014). Esto podría verse, sin duda, como una potencial debilidad del modelo al tener una exigencia demasiado específica para el cumplimiento de sus parámetros explicativos.

2.2. Naturaleza cognitiva y arquitectura *top-down*

Para los defensores de una arquitectura *top-down* de las atribuciones de agencia mental, la naturaleza del fenómeno es esencialmente cognitivo e inferencial. Este tipo de atribución surgiría luego de un análisis introspectivo de nuestros propios pensamientos conscientes que conduciría a explicar su ocurrencia retrospectivamente en términos causales (Graham y Stephens, 1994; Campbell, 1999; Stephens y Graham, 2000; Vosgerau y Voss, 2014). Por esto, esta teoría también podría ser denominada ‘explicacionista’, esto, porque el contenido final del estado atribucional surgiría como una explicación de la ocurrencia de un pensamiento de primer orden. Mediante una inferencia (*top-arriba*) basada en información contextual sobre el contenido del pensamiento, información sobre el contexto de surgimiento del pensamiento e información acerca de la naturaleza fenoménica del pensamiento en cuestión (*down-abajo*), el sujeto lograría referirse a éste como su ‘autor’. En este sentido, la dirección causal de la producción de una atribución de agencia mental va desde el set de conocimiento previo y creencias sobre la realidad (*top-arriba*) del sujeto, hacia la evaluación de un pensamiento con características fenoménicas específicas pero pasivas a la luz de la pregunta por su agencia (*down-abajo*). En esta teoría no habría nada especial respecto del carácter experiencial-representacional de un pensamiento en términos de agencia. El agente de un pensamiento surgiría como una operación de segundo orden de carácter evaluativa, inferencial y retrospectiva en tanto imposición de la categoría de ‘agencia’ a los pensamientos a los cuáles se tiene acceso en 1^{era} persona basado en cierta información privilegiada respecto de su ocurrencia (en otras palabras, no habría una sensación de agencia mental a la base de una atribución de agencia mental, los pensamientos poseerían una naturaleza pasiva desde un punto de vista experiencial). Finalmente, se sigue que, dentro de esta teoría, la agencia no es algo que pertenece experiencialmente al pensamiento, sino que es una forma en la cual se le da sentido a su ocurrencia experiencialmente pasiva i.e. es una forma de explicar la ocurrencia de nuestros pensamientos.

3. El desafío desde la psicopatología: teorías sobre la atribución de agencia mental y el caso de los delirios de inserción de pensamiento

El desafío desde la psicopatología implica la idea de que la observación y descripción fenomenológica detallada de síntomas psicopatológicos que suponen alteraciones en el proceso de atribución de agencia mental puede ser ocupada como estrategia para evaluar los méritos argumentativos de las

teorías que intentan explicar el fenómeno (López-Silva, 2017).⁶ En el resto de este trabajo me centraré en como los delirios de inserción de pensamiento presentan un desafío insoslayable para tales teorías.⁷ Este tipo de delirio implica la creencia de que ciertos agentes externos de distinta naturaleza tienen la capacidad de poner pensamientos en la cabeza de un paciente (Mullins y Spence, 2003; López-Silva, 2015). Estos son tres reportes clásicos del síntoma:

Miro hacia afuera por la ventana y pienso que el jardín se ve bonito y que el pasto se ve muy bien. Pero los pensamientos de Eammon Andrews vienen a mi mente. No hay otros pensamientos ahí, solo los suyos. Él usa mi mente como si fuese una pantalla y proyecta sus pensamientos en ella de la misma forma en que proyectas una fotografía (Mellor, 1970, p. 17, *traducción del autor*).

No escuchaba esas palabras como sonidos literales, como si las casas estuviesen hablando y yo las escuchase. En vez de eso, las palabras solo venían a mi cabeza – eran ideas que estaba teniendo. Pero instintivamente sabía que no eran mis ideas. Ellas pertenecían a las casas, y las casas las habían puesto en mi cabeza (Saks, 2007, p. 29, *traducción del autor*).

Ciertos pensamientos son puestos en mi mente, como ‘Mata a Dios’. Es solo mi mente funcionando, pero no lo es. Ellos vienen de este sujeto, Chris. Son sus pensamientos (Frith, 1992, p. 66, *traducción del autor*).

Este fenómeno psicopatológico ha sido interpretado como la alteración más compleja del proceso de atribución de agencia mental, por lo que ofrece una excelente oportunidad para evaluar los méritos de las teorías que intentan explicar el fenómeno (Stephens y Graham, 2000; Gallagher, 2000, 2014; Zahavi, 2005; Campbell, 1999, 2002; Frith, 1992; Proust, 2009).

Comenzamos nuestro análisis crítico con la teoría *top-down*. Una de las dificultades iniciales más claras de este enfoque es que, para que su estructura explicativa funcione, éste descansa en una teoría psicológica sobre los mecanismos atribucionales altamente problemática.⁸ Defensores de esta teoría indican que un pensamiento P de un sujeto S es finalmente autoatribuido en términos de agencia, si y solo si, P es consistente y coherente con las

6 El foco principal del debate acerca de la aplicabilidad de estas teorías en el campo de la psicopatología estará en la evaluación de las explicaciones que ofrecen principalmente para estos dos tipos de delirios. Si se discuten otros fenómenos psicopatológicos, será solamente en forma tangencial.

7 Si bien el ‘desafío desde la psicopatología’ puede tomar muchas formas (dependiendo del fenómeno psicopatológico que se describa), existe acuerdo en este contexto al considerar a los delirios de inserción de pensamiento como el fenómeno más desafiante para las teorías de agencia mental dada su compleja estructura fenomenológica y narrativa (Stephens y Graham, 2000; Gallagher, 2000, 2004, 2014; Zahavi, 2005; Campbell, 1999, 2002; Proust, 2009; López-Silva, 2017, 2018, entre muchos otros).

8 Al proponer una fenomenología cognitiva más pobre en términos fenoménicos, este enfoque parece verse forzado a descansar en mecanismos psicológicos para explicar el contenido y producción de las atribuciones de agencia mental.

expectativas que S tiene de sí mismo en función de su autoimagen (Graham y Stephens 1994; Stephens y Graham 2000; Campbell 1999, 2002; cf. Gallagher, 2007). Aquí, una condición necesaria para la autoatribución retrospectiva de agencia mental es que P posea un contenido consistente con la forma en que S piensa sobre sí mismo i.e. que P sea algo que S ‘podría llegar a pensar’ en un contexto específico. Sin embargo, muchas veces uno autoatribuye agencialmente pensamientos que claramente no son consistentes con la propia autoimagen y las expectativas que uno posee de sí mismo. Esto es exactamente lo que se observa, por ejemplo, en la autoatribución agencial de pensamientos moralmente reprochables⁹. En estos casos, un pensamiento P no parece ser consistente con la imagen que tengo de mí mismo y aún así, lo autoatribuyo agencialmente. Es más, es exactamente esta atribución la que explica el carácter conflictivo que los pensamientos moralmente reprochables poseen; si tales pensamientos no fuesen autoatribuidos, estos no poseerían carácter conflictivo en lo absoluto. En estos casos, S piensa P y P no es consistente con la autoimagen de S, y a pesar de esto, P termina siendo autoatribuido. Claramente, este tipo de casos no puede ser explicado por un enfoque *top-down*. Casos como este son comunes en la vida mental humana y muestran que simplemente no es cierto de que las autoatribuciones de agencia dependen de si podemos, o no, darle sentido al contenido de cierto pensamiento en coherencia con nuestra autoimagen y expectativas; la verdad es que muchos de los pensamientos que autoatribuimos en términos agenciales ni siquiera poseen esta condición.

Ahora bien, además de poseer problemas para explicar las instancias no patológicas del fenómeno, el enfoque *top-down* tampoco parece estar en condiciones de lidiar con el desafío propuesto por los delirios de inserción de pensamiento. Basado en estudios de desempeño cognitivo en pacientes psicóticos, el enfoque *top-down* propone que pacientes que sufren de este síntoma simplemente no son capaces de rastrear las causas de ciertos pensamientos (Martin y Pacherie, 2013); esto les llevaría a concluir que si tales pensamientos no son de ellos (no ha sido creado por ellos), necesariamente deben ser de alguien más (debe haber sido creado por alguien más). Esto, finalmente, precipitaría la externalización del contenido del pensamiento en cuestión como una especie de racionalización (Synofzik, Vosgerau y Newen, 2008; Martin y Pacherie, 2013; Vosgerau y Voss, 2014). El problema principal

9 Se asumirá que un pensamiento moralmente reprochable posee, paradigmáticamente, un carácter conflictivo para el sujeto quien posee acceso en 1^{era} persona al contenido del pensamiento, tal que, cuando S piensa que P y el contenido de P inmoral, la ocurrencia de P en S *necesariamente* implica un conflicto moral que se enfrenta a la autoimagen que S tiene de sí mismo.

con esta explicación es que la inhabilidad de un sujeto S para rastrear las causas de un pensamiento P no explica necesariamente el tipo de atribución externa que caracteriza los delirios de inserción de pensamiento; no explica el elemento más importante del síntoma. Esta inhabilidad solo explicaría el hecho de que un sujeto S experimente un pensamiento P como sorpresivo o ‘descontextualizado’, lo cual no es diferente de la forma en que experimentamos varios de los pensamientos que comúnmente tenemos día a día. Muchas veces, uno mismo no logra rastrear las causas específicas de un pensamiento P que aparece repentinamente en nuestro flujo de la conciencia. Sin embargo, uno no termina externalizando tales pensamientos.

Pensemos en el caso de estar hablando sobre los actuales cambios políticos en Latinoamérica con un grupo de colegas. En medio de ese contexto uno podría proyectar, pensar o recordar pensamientos con cierto tipo de contenido relacionado. Súbitamente, un pensamiento descontextualizado con el contenido ‘necesito comprar una pinza para mi cocina’ aparece en el flujo de la conciencia. Su carácter descontextualizado muchas veces incluirá sensaciones de sorpresa o incluso bochorno por no poder mantener cierta tarea cognitiva como el eje de la propia cognición consciente. Sin embargo, este tipo de episodios son muy comunes y se les denomina como ‘pensamientos repentinos’ (*unbidden thoughts*, Frankfurt, 1976). La mejor forma de explicar este tipo de pensamientos sería aludiendo a la incapacidad *in-situ* de rastrear e incorporar en la experiencia consciente de tal pensamiento parte de su información causal. El punto que quiero estresar acá es que el enfoque *top-down* no logra explicar cómo autoatribuimos pensamientos de los cuales no poseemos información causal en el momento, y tampoco por qué los pacientes que sufren delirios de inserción de pensamiento terminarían externalizando tales pensamientos.

Continuemos nuestro análisis con la teoría de arquitectura *bottom-up*. Tal como se puede observar, este enfoque imita la estructura explicativa de la teoría dominante en el contexto de la discusión sobre la agencia motora, lo que levanta las primeras sospechas. Las acciones motoras son muy diferentes de los pensamientos en términos estructurales y fenomenológicos. Las acciones motoras obedecen a experiencias sensoriales, y si bien los pensamientos podrían adquirir algunas características fenoménicas a partir de su asociación a ciertos contenidos de origen sensorial, estos no son experiencias sensoriales propiamente tales, son experiencias cognitivas (Pitt, 2004; Bayne y Pacherie, 2013). Es claro que caminar es diferente de pensar en caminar desde un punto de vista experiencial. Pues bien, por estas razones, no es claro si debiésemos establecer paralelismos entre enfoques que descansan fuertemente en las características sensoriales que ciertos estados mentales poseen (movimientos

corporales), esto, porque es posible indicar que los movimientos corporales adquieren su fenomenología y estructura atribucional exactamente dada la arquitectura sensorial retroalimentaria que poseen (*feedback* aferente y eferente), y este tipo de arquitectura no parece ser propia de las experiencias cognitivas como los pensamientos. Si bien este no es un punto suficientemente fuerte como para abandonar la teoría, llama la atención a mirar cualquier paralelismo entre acciones motoras y pensamientos con un grado de reserva (Proust, 2009).

Un problema más fundamental dentro del enfoque *bottom-up* se observa al examinar la fenomenología del pensamiento que este asume para que su estructura explicativa funcione. Tal como en el caso motor, este enfoque asume la existencia de una sensación de agencia mental que forma parte de la estructura fenoménica de todos los pensamientos que son finalmente autoatribuidos agencialmente por un sujeto. Así, las autoatribuciones de agencia mental surgen de la mera adopción del contenido fenoménico de esta sensación de agencia mental de primer orden tal que, para los defensores de esta teoría, la fenomenología más fundamental de los pensamientos es activa (Gallagher, 2014). Pues bien, el problema es que esta premisa fundamental no parecer ser correcta, o al menos, no está garantizada. Proust (2009) ha señalado que, mientras para algunos, los reportes de sus propios pensamientos se asemejan a algo que ellos ‘hacen’, para otros, los pensamientos son algo que simplemente ‘ocurre’ en el flujo de la conciencia. Esto genera dos problemas. Desde un punto de vista experiencial, la emergencia de un pensamiento no parece estar acompañada de una sensación de agencia remotamente comparable con el caso motor, esto, porque los pensamientos no se experimentan como algo que uno está haciendo activamente, sino que como algo que nos ocurre (Strawson, 2003; López-Silva, 2017). El acto de pensar parece poseer una fenomenología receptiva, que no es lo mismo que el requerimiento fundamental del enfoque *bottom-up*. Tal como sugieren de Hann y de Bruin (2010), lo que caracteriza la aparición de un pensamiento en mi campo de la conciencia es una especie de concentración, una especie de embudo mediante el cual uno puede direccionar los pensamientos que están por aparecer. Algunos podrían asociar sensaciones agenciales a este período previo a la emergencia de un pensamiento, lo cual no me parece problemático. Sin embargo, la aparición de un pensamiento propiamente tal se experimenta de forma receptiva. Algunos críticos podrían indicar que, si bien no hay una sensación de agencia asociada a la aparición de un pensamiento, este fenómeno parece estar precedido por una sensación de esfuerzo o control, por ejemplo, como cuando estamos intentando resolver un problema matemático, o simplemente intentamos encontrar la respuesta para

algún asunto (Frith, 1992). Si bien esta sugerencia parece correcta, solo implica que uno posee algo de control para iniciar ciertos trenes de pensamiento, lo cual no es lo mismo que controlar la emergencia de un pensamiento específico en el mismo sentido en que uno controla un movimiento corporal, por lo tanto, la crítica no parece ser dañina.

El segundo problema asociado a este requerimiento explicativo, surge al observar que la idea de una sensación de agencia mental como condición necesaria para una autoatribución de agencia mental es inconsistente con la atribución de agencia de los pensamientos repentinos (*unbidden thoughts*). Este tipo de pensamiento es experimentado fenomenológicamente como emergiendo en el flujo de la conciencia de forma involuntaria (Frankfurt, 1976). El mismo Shaun Gallagher (2014) - defensor de una teoría *bottom-up* - ha reconocido que la mejor forma de darle sentido a los pensamientos repentinos es explicándolos como casos de pensamientos que no poseen sensación de agencia asociada. Ahora bien, a pesar de esto, los sujetos que experimentan este tipo de pensamientos igual terminan autoatribuyéndolos agencialmente, lo que prueba que la presencia de una sensación de agencia mental en la fenomenología más fundamental de los pensamientos no es una condición necesaria para su autoatribución en términos agenciales. Finalmente, incluso si existiese una *sensación de agencia mental*, el enfoque *bottom-up* no es capaz de explicar su origen o la forma en que lograría estar presente en la estructura fenoménica de todos los pensamientos.

Finalmente, el enfoque *bottom-up* tampoco parece estar en condiciones para enfrentar el desafío propuesto por los delirios de inserción de pensamiento. Según los defensores de este enfoque, este síntoma se produciría por la ausencia de una sensación de agencia mental en ciertos pensamientos, lo cual precipitaría su externalización en términos de agencia (Gallagher, 2014; Zahavi, 2005). Sin embargo, tal como se puede advertir, esta explicación no lograría distinguir entre la explicación ofrecida para los pensamientos repentinos y los delirios de inserción de pensamiento. Ambos fenómenos cognitivos son explicados como casos en donde pensamientos específicos carecen de sensación de agencia. Sin embargo, mientras los pensamientos repentinos no son externalizados, los otros sí lo son, constituyéndose posteriormente en pensamiento insertados. ¿Por qué unos pensamientos son externalizados y otros no si ambos carecen de sensación de agencia? El enfoque *bottom-up* no parece poder explicar esto.

Recientemente, defensores del enfoque *bottom-up* han intentado resolver este desafío indicando que, en el caso de los pensamientos externalizados que caracterizan los delirios de inserción de pensamiento, a la falta de sensación de agencia mental, se le debe sumar una *sensación de inserción* (Billon y

Kriegel, 2015). Superficialmente, esta sugerencia parecería resolver el debate, sin embargo, este no parece ser el caso. En estricto rigor, todo lo que esta sugerencia implica es que un pensamiento podría experimentarse sin sensación de agencia mental (lo cual, a mi opinión, obedece a todos los casos normales) y, además, como si fuese insertado. Sin embargo, esta sugerencia no explica la especificidad de los reportes de los pacientes psicóticos. Si bien, muchas veces estos sujetos no pueden siquiera recordar el contenido de un pensamiento insertado en su mente, los reportes son claros en la identificación del agente externo que, supuestamente, realizó la inserción (Mullins y Spence, 2003). Es más, la identificación del agente externo es una de las dos condiciones fundamentales para que el síntoma sea diagnosticado. Sin esta identificación, el diagnóstico del síntoma sería ininteligible. Pues bien, estos casos no implican solamente una sensación de pasividad cognitiva, mas una sensación de inserción – como plantean Billon y Kriegel –, los pensamientos insertados en este caso patológicos son específicamente experimentados como insertados por un agente específico (Frith, 1992; Stephens y Graham, 2000; Mullins y Spence, 2003; Saks, 2007; Payne, 2013). Así, si bien esta sugerencia parece complejizar el análisis fenomenológico que los defensores de la teoría *bottom-up* poseen de los pensamientos, no resuelve el desafío ya que no explica como los pacientes pasan de sentir un pensamiento P como si fuese insertado a experimentarlo como si un agente específico lo hubiese insertado. Claramente, uno podría indicar que sentir un pensamiento P como insertado no es suficiente para externalizar completamente su contenido, ni menos para explicar la especificidad de los agentes que los pacientes psicóticos reportan.

En este contexto, alguien podría sugerir que los pacientes pasan de experimentar un pensamiento sin agencia y con una sensación de inserción a la identificación del agente de forma inferencial.¹⁰ Sinceramente, no tendría problemas con esta solución, es más, creo tener una opinión similar (que aún estoy elaborando). El problema con esta opción es que no parece ser consistente con la estructura explicativa del enfoque. Los defensores del enfoque *bottom-up* han recientemente sido enfáticos en la elaboración de la comprensión del contenido de los delirios de inserción de pensamiento desde una modalidad de *endorsement* (Gallagher, 2014; Zahavi, 2018; Henriksen, Parnas & Zahavi, 2019). Esto quiere decir que las experiencias conscientes delirantes adquieren su contenido representacional de la riqueza experiencial de los estados de primer orden que subyacen a estos. El enfoque *bottom-up* sugiere que un

10 Agradezco a uno de los revisores por este punto.

reporte doxástico con el contenido ‘X insertó P en mi cabeza’ solo es inteligible si el sujeto posee una experiencia de X insertando P en su cabeza¹¹ y, por lo tanto, la presencia de un agente en el reporte debería ser parte de la experiencia primaria a la base de éste. El enfoque *bottom-up* no deja espacio para que el agente sea el producto de un acto inferencial i.e. no presente en la experiencia primera que sustenta un reporte delirante.

4. ¿*Bottom-up* o *top-Down*?

Este artículo ha intentado avanzar en la comprensión de la naturaleza y arquitectura de las atribuciones de agencia mental mediante un examen crítico de las dos teorías dominantes en el debate. Si bien este trabajo posee antecedentes similares en el contexto de la discusión de la agencia motora, ningún trabajo sistemático y exclusivo para el caso mental ha podido ser encontrado en la literatura actual. Tal como se puede advertir, el análisis nos lleva a concluir que las actuales versiones del enfoque *top-down* y del enfoque *bottom-up* no son capaces de lidiar con una serie de problemas conceptuales y fenomenológicos, pero, por sobre esto, ninguno de estos enfoques parece estar en buen pie para lidiar de forma plausible con el desafío desde la psicopatología, especialmente con aquel propuesto por la fenomenología de los delirios de inserción de pensamiento. A todo lo ya señalado se le suma el siguiente problema: ambos enfoques parecen carecer de evidencia empírica directa que apoye sus premisas. Para ambos enfoques, la evidencia ha sido establecida extrapolando los datos disponibles en casos de alteraciones neuropsicológicas que estarían a la base de delirios motores como la somatoparafrenia y los delirios de control externo i.e. alteraciones en experiencias de carácter sensorial y motor. Tal como ya hemos señalado, no es claro que uno pueda extrapolar directamente la evidencia disponible en los casos motores para los casos cognitivos sin levantar algunas sospechas y, por lo tanto, ambos enfoques parecen carecer de evidencia empírica directa, lo que añade otro desafío a la lista. Ahora bien, todo esto llama claramente a la exploración de nuevas alternativas que intenten darle sentido a la pregunta central de este artículo. El desafío acá es establecer el camino que estas exploraciones podrían tomar. Por ejemplo, algunos podrían intentar

11 Otra forma de elaborar el problema sobre el origen del contenido de los delirios es el explicacionismo (ver el inicio del artículo). Este enfoque indica que el contenido final del enfoque doxástico es una explicación para una experiencia de primer nivel que es poco clara o menos rica en contenido, por lo que existen altos grados inferenciales en el proceso de elaborar un delirio. Sin embargo, esta modalidad no parece ser consistente con el enfoque *bottom-up* sino que con la teoría *top-down*. Para una revisión detallada de estas dos modalidades de emergencia del contenido de los delirios, ver: Bayne y Pacherie (2007).

redefinir una teoría *top-down*, mientras que otras podrían intentar reconstruir una teoría del estilo *bottom-up*. Cual sea el caso, incluso en el caso de la exploración de propuestas híbridas, los futuros enfoques deberán ser capaces de superar los problemas que nuestro análisis ha notado en las dos teorías dominantes en la actualidad.

Referencias

- BAYNE, T. “The disunity of consciousness in psychiatric disorders”. En K. W. M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, y T. Thornton. *The Oxford handbook of philosophy and psychiatry*. Oxford: Oxford University Press, 2013.
- _____. “The sense of agency”. En F. Macpherson. *The senses*. Oxford: Oxford University Press, 2011.
- _____. “The phenomenology of agency”. *Philosophy Compass*, Nr. 3, pp. 182-202, 2008.
- BAYNE, T. y PACHERIE, E. “Narrators and comparators: The architecture of agential self-awareness”. *Synthese*, Nr. 159, pp. 475-491, 2007.
- BILLON, A. “Does consciousness entail subjectivity? The puzzle of thought insertion”. *Philosophical Psychology*, Vol. 26, Nr. 2, pp. 291-314, 2013.
- BILLON, A. y KRIEGER, U. “Jaspers’ Dilemma: The Psychopathological Challenge to Subjectivity Theories of Consciousness”. En R. Gennaro. *Disturbed Consciousness*. USA: MIT Press, 2015.
- BLACKEMORE, J. y FRITH, C. “Self-awareness and action”. *Curr Opin Neurobiol*, Nr. 2, pp. 219-224, 2003.
- CAMPBELL, J. “The ownership of thoughts”. *Philosophy, Psychiatry y Psychology*, Vol. 9, Nr. 1, pp. 35-39, 2002.
- _____. “Schizophrenia, the space of reasons, and thinking as a motor process”. *The Monist*, pp. 609-625, 1999.
- DAPRATI, E. *et al.* “Looking for the agent, an investigation into self-consciousness and consciousness of the action in schizophrenic patients”. *Cognition*, Nr. 65, pp. 71-86, 1997.
- DE HANN, S. y DE BRUIN, L. “Reconstructing the minimal self, or how to make sense of agency and ownership”. *Phenomenology and the Cognitive Sciences*, Nr. 9, pp. 373-396, 2010.
- DE VIGNEMONT, F., y FOURNERET, P. “The sense of agency: A philosophical and empirical review of the “who” system”, *Consciousness and Cognition*, Nr. 13, pp. 1-19, 2004.
- FRANKFURT, H. “Identification and externality”. En A. O. Rorty. *The Identities of Persons*. Berkeley, University of California Press, 1976.
- FRITH, C. “The cognitive neuropsychology of schizophrenia”. Hillsdale: Erlbaum, 1992.
- GALLAGHER, S. “Relations between agency and ownership in the case of schizophrenic thought insertion and delusions of control”. *Review of Philosophy and Psychology*, Vol. 6, Nr. 4, pp. 865-879, 2014.

- _____. “The Natural Philosophy of Agency”. *Philosophy Compass*, Vol. 2, Nr. 2, pp. 347-357, 2007.
- _____. “Philosophical conceptions of the self: Implications for cognitive science” *Trends in Cognitive Sciences*, Vol. 4, Nr. 1, pp. 14-21, 2000.
- GALLAGHER, S. y ZAHAVI, D. “The Phenomenological Mind”. London: Routledge, 2008.
- GRAHAM, G. y STEPHENS, G. L. “Mind and mine”. En G. Graham y G. Stephens. *Philosophical Psychology*. Cambridge, MA, MIT Press, 1994.
- HENRIKSEN, M., PARNAS, J. y ZAHAVI, D. Schizophrenia and disturbed for-*me*-ness in schizophrenia. *Consciousness & Cognition*, 74, 2019, 102770.
- JASPERS, K. “General Psychopathology” (7th edition). Manchester: Manchester University Press, 1963.
- LÓPEZ-SILVA, P. “Mapping the psychotic mind: A Review on Thought Insertion”. *Psychiatric Quarterly*, [Online First], pp. 1-17, 2018.
- _____. “Me and I are not friends, just Acquaintances: On thought Insertion and Self-Awareness”. *The Review of Philosophy y Psychology*, 2017.
- _____. The Unity of Consciousness in Pre-Psychotic States. A Phenomenological Analysis. *Studies in Psychology*. 37, pp. 1-34, 2016.
- _____. “Schizophrenia and the place of egodystonic states in the aetiology of thought insertion” *The Review of Philosophy y Psychology*, Vol. 7, Nr. 3, pp. 577-594, 2015.
- _____. “La relevancia filosófica del estudio de la esquizofrenia. Cuestiones metodológicas y conceptuales”. *Revista Colombiana de Psiquiatría*, Vol. 43, Nr. 3, pp. 168-174, 2014.
- MANDIK, P. “Control consciousness”. *Topics in Cognitive Science*, Nr. 2, pp. 643-657, 2010.
- MARTIN, J.M. y PACHERIE, E. “Out of Nowhere: Thought Insertion, Ownership and Context-Integration”. *Consciousness and Cognition*, Vol. 22, Nr. 1, pp. 111-122, 2013.
- MELLOR, C.S. “First rank symptoms of schizophrenia”. *The British Journal of Psychiatry*, Nr. 117, pp. 15-23, 1970.
- MULLINS, S. y SPENCE, S. “Re-examining thought insertion”. *British journal of psychiatry*, Nr. 182, pp. 293-329, 2003.
- MYLOPOULOS, M. “Agentive awareness is not Sensory awareness”. *Phil Studies*, Online First, 2015.
- O’BRIEN, L. y SOTERIOU, M. “Mental Actions”. UK, OUP, 2009.
- PACHERIE, E. “The phenomenology of action: A conceptual framework”, *Cognition*, Nr. 107, pp. 179-217, 2008.
- PEACOCKE, C. “Awareness, ownership, and knowledge”. En J. Roessler y N. Eilan. *Agency and self-awareness: Issues in philosophy and psychology*. Oxford, Oxford University Press, 2003.
- PAYNE, E. “Speaking to my madness”. USA, CreateSpace, 2013.
- PITT, D. “The phenomenology of cognition Or What is it like to think that P?”. *Philosophy and Phenomenological Research*, Nr. 69, pp. 1-36, 2004.

- PRINZ, J. J. “All consciousness is perceptual”. En B. P. McLaughlin y J. Cohen. *Contemporary debates in philosophy of mind*. Malden, MA: Blackwell, 2007.
- PROUST, J. “Is there a Sense of Agency for Thoughts?” En L. O’Brien y M. Soteriou. *Mental Actions*. UK, OUP, 2009.
- SAKS, E.R. “The centre cannot hold. My journey through madness”. New York: Hyperion, 2007.
- SCHNEIDER, K. “Clinical Psychopathology” (trans. By M.W. Hamilton). New York: Grune y Stratton, 1959.
- STEPHENS, G. L. y G. GRAHAM. “When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts”. Cambridge MA, MIT Press, 2000.
- STRAWSON, G. “Mental Ballistics or the Involuntariness of Spontaneity”. *Proceedings of the Aristotelian Society*, N.S., Nr. 103, pp. 227-256, 2003.
- SYNOFZIK, M., VOSGERAU, G. y NEWEN, A. “I move, therefore I am: a new theoretical framework to investigate agency and ownership”. *Consciousness and Cognition*, Nr. 17, pp. 411-424, 2008.
- VOSGERAU, G. y VOSS, M. “Authorship and Control over Thoughts”. *Mind y Language*, Vol. 29, Nr. 5, pp. 534-565, 2014.
- ZAHAVI, D. Consciousness, Self-Consciousness, Selfhood: a Reply to some Critics. *Review of Philosophy & Psychology*, Vol. 9, Nr. 3, pp. 703-718. 2018.
- _____. “Subjectivity and selfhood: Investigating the first-person perspective”. Massachusetts: MIT Press, 2005.