

Atribución de creencias e irracionalidad: El caso de los delirios

Belief ascription and irrationality: The case of delusions

VILATTA, EMILIA*
IDH- CONICET - UNC. emiliavilatta@gmail.com

RECIBIDO EL 5 DE FEBRERO DE 2021, APROBADO EL 1 DE MARZO DE 2021

RESUMEN ABSTRACT

En este artículo me ocuparé del debate en torno a la posibilidad de atribuir a los delirios el estatus de creencias, a partir de la tensión conceptual que aparece entre el requisito de racionalidad que exige la atribución de creencias y el carácter irracional que exhiben los delirios. Para ello, analizaré lo que denomino como el “argumento de la irracionalidad” en las dos variantes que identifico: la variante trascendental y la variante empírica. En contra de este argumento, mostraré que, pese a la irracionalidad que manifiestan los estados delirantes, podemos seguir atribuyéndoles el estatus de creencias.

In this paper I will focus on the discussion about the possibility to ascribe the status of beliefs to delusions, considering that there is a conceptual tension between the requirement of rationality for belief ascription and the irrationality that delusions exhibit. For this purpose, I will analyse what I call as the “irrationality argument” in the two variants that I identify: the transcendental variant and the empirical variant. Against this argument, I will show that, despite the irrationality that delusional states exhibit, we can continue ascribing to such states the status of beliefs.

PALABRAS CLAVE KEY WORDS

Delirios, creencias, racionalidad, atribución intencional.

Delusions, beliefs, rationality, intentional ascription.

*  orcid.org/0000-0001-7878-0885 **Google Scholar**



1. Introducción

En años recientes, los delirios han atraído particularmente la atención de numerosos filósofos analíticos, ya que, entendidos como un caso paradigmático de irracionalidad, suponen un desafío significativo a buena parte de las teorías filosóficas clásicas sobre la atribución de estados intencionales, en particular a los enfoques de corte interpretativista que asumen estándares de racionalidad. Así, una de las discusiones filosóficas centrales en torno a los delirios ha sido la de si es posible atribuir a las emisiones lingüísticas calificadas como “delirantes” el estatus de estados doxásticos (i.e. creencias).

El enfoque estándar, derivado de la Psiquiatría, considera que un delirio es una “creencia falsa basada en deducciones incorrectas sobre la realidad externa que se mantiene firmemente a pesar de lo que el resto de las personas creen y a pesar de las pruebas evidentes e indiscutibles de lo contrario (APA 824). Como puede apreciarse, este es un enfoque de tipo doxástico que reconoce el carácter irracional de los delirios al mismo tiempo que los considera como creencias. Sin embargo, en los debates filosóficos recientes (Bayne & Pacherie, Bortolotti & Miyazono, Diez, López-Silva) este enfoque ha sido criticado por evidenciar una supuesta tensión conceptual: dado que la atribución de creencias se halla vinculada a la satisfacción de determinados parámetros de racionalidad, se ha puesto en cuestión que los delirios –por su irracionalidad– puedan ser considerados como creencias.

La crítica central, esbozada por quienes defienden enfoques no doxásticos, aquellos cuyo punto de convergencia es que los delirios no serían creencias sino otro tipo de estado mental, reside en lo que denominaré aquí como “*el argumento de la irracionalidad*”. En pocas palabras, este argumento señala que debido a las características de irracionalidad que los delirios manifiestan, no pueden ser considerados creencias. En el presente artículo me ocuparé de este argumento con el objetivo de mostrar que no resulta contundente en contra del enfoque doxástico de los delirios ni en contra de los enfoques interpretativistas que suponen que la atribución de creencias implica atribución de racionalidad. Para ello, en primer lugar (sección 2), mostraré que el argumento puede leerse de dos maneras. Por un lado, como un argumento *trascendental* según el cual la racionalidad es una *condición de posibilidad* para atribuir creencias. Por otro lado, como un argumento *empírico* de acuerdo con el

cual las creencias, como *output*, deberían exhibir ciertas características de racionalidad que los delirios no exhibirían. Luego (sección 2.1), analizaré la primera variante y para ello recuperaré el “*argumento del trasfondo*” elaborado por Donald Davidson en sus escritos sobre la irracionalidad. A partir de ello, argumentaré que en la medida en que podamos comprender la emisión lingüística del sujeto delirante y atribuirle el contenido de su delirio, las condiciones mínimas de racionalidad necesarias para la interpretación se encuentran garantizadas. Por lo tanto, no hay razones para pensar que la irracionalidad de los delirios impida la atribución doxástica ni para cuestionar la tesis de que la atribución de creencias implica racionalidad. Posteriormente (sección 3), argumentaré que, incluso si el argumento es leído en su variante empírica, éste no funciona como una crítica sustantiva al enfoque doxástico. Para ello mostraré que los delirios, por un lado: (i) no son tan irracionales como aparece *prima facie* y, por otro lado (sección 4), (ii) que la manera en que está desarrollado el argumento supone requisitos de racionalidad demasiado exigentes incluso para las creencias ordinarias. En conclusión, señalaré que es posible compatibilizar la tesis de que para atribuir creencias debemos atribuir racionalidad con el hecho de que los delirios, pese a mostrar irracionalidad en algunos aspectos, aún pueden ser vistos como creencias.

2. El argumento de la irracionalidad

Como crítica al enfoque doxástico estándar, han proliferado distintos tipos de enfoques filosóficos que pueden denominarse “no-doxásticos”, porque pese a sus diferencias comparten la siguiente tesis: los delirios no serían creencias, sino algún tipo diferente de estado mental. Mientras que algunos consideran que se trata de otra clase de estado de los ya conocidos tales como actos de imaginación (Egan), actos de habla vacíos (Berrios) o meta-representaciones (Currie), otros filósofos consideran que se trataría de nuevos tipos de estados mentales (“*sui generis*”), tales como una mezcla entre creencias e imaginaciones (“*bi-imaginaciones*”) (Stephen & Graham), creencias intermedias (Schwitzgebel) o creencias de segundo nivel (Frankish). Pese a las divergencias que existen entre estos distintos enfoques¹, lo que subyace a todos ellos es lo que denominaré como “*el argumento de la irracionalidad*”.

¹ Aquí me interesa considerar no las características particulares de estos enfoques sino el argumento central que esgrimen en contra del enfoque doxástico estándar.

De acuerdo con este argumento, los delirios serían *demasiado irracionales* como para contar como creencias (Dub). Este argumento toma dos variantes. Por un lado, lo que denomino su variante *trascendental*, según el cual se afirma que el *proceso de interpretación* involucrado en la *atribución* de las creencias se halla gobernado por principios que presuponen la racionalidad del agente interpretado. Por otro lado, en su variante *empírica*, se señala que una característica del rol funcional de las creencias es ser racional: si un estado mental no se ha formado y se mantiene de modo racional, ni motiva la conducta en un modo racionalmente apropiado, entonces no juega el rol de una creencia. En la primera versión, la racionalidad es una presuposición de la práctica de atribución de creencias. En la segunda, la racionalidad es una característica del *output* del proceso de interpretación: esto es, del estado en sí mismo. Veamos en detalle cada variante de este argumento.

2.1 Constricciones de racionalidad en la atribución de creencias: el argumento del trasfondo

Según mencionamos, en la variante *trascendental*, el argumento consiste en señalar que el proceso de *atribución* de creencias se halla gobernado por principios que implican atribuir racionalidad al agente interpretado. Los críticos del enfoque doxástico esgrimen que, puesto que los sujetos delirantes son demasiado irracionales no podemos atribuirles entonces tal racionalidad ni considerar los estados como creencias. Esta lectura de las constricciones de racionalidad en la atribución de creencias se apoya en la tradición interpretativista², particularmente en el enfoque de Donald Davidson, quien ha recibido mayor atención y crítica en estos debates, quizás por haber enfatizado el hecho de que solo los *seres racionales* pueden ser intencionalmente interpretados. Es así como diversos filósofos han señalado –algunos para criticar al enfoque doxástico y otros para criticar al propio Davidson– que los sujetos delirantes no satisfarían el requisito de racionalidad que, de acuerdo con este filósofo, debe cumplir un agente para que sea legítimo atribuirle creencias (Bortolotti, Campbell, Currie, Klee, Reimer).

Ahora bien, las diferentes críticas parecen implicar que la interpretación se encontraría atada a la posibilidad (empírica) de que un sujeto satisfaga o no las constricciones de racionalidad. No obstante, el argumento de

² Englobamos aquí a aquellos filósofos que han considerado que el *contenido* de los estados mentales está determinado por el proceso de interpretación.

corte interpretativista, particularmente en su versión davidsoniana, si bien enfatiza que existe una restricción de racionalidad sobre la atribución de creencias, afirma que es una *condición de posibilidad* del proceso interpretativo que las creencias y acciones del agente sean racionales. Es decir, es un argumento de tipo *trascendental* aquel que afirma que la atribución doxástica supone atribución de racionalidad y debería ser leído en este sentido. Veamos en detalle en qué consiste el requisito de racionalidad establecido por Davidson y si es cierto que los sujetos con delirios no podrían satisfacerlo.

En el marco del proyecto de la interpretación radical, Davidson (*into truth*) en línea con Quine, postula el denominado "*principio de caridad*", el cual sostiene que, si es posible interpretar el significado de las emisiones de un hablante y conjuntamente sus creencias y otros estados mentales, esto se debe a que la mayor parte de sus creencias son verdaderas y el sujeto es, en amplia medida, racional. Este principio no se considera una suposición contingente u opcional, sino *constitutiva* del acto de interpretación (Ludwig). En palabras de Davidson: "En nuestro afán de hacer que tenga sentido lo que dice [el hablante], intentaremos con una teoría que lo encuentre consistente, un creyente de verdades y un amante del bien" (Davidson 222).

Así, en el proceso de interpretación, el intérprete se guiaría por tres supuestos: en primer lugar, habrá de asumir que los contenidos de las creencias más básicas del sujeto están constituidos por ciertos rasgos objetivos del entorno que causan dichas creencias en el sujeto. En segundo lugar, habrá de asumir que, en los casos más básicos, lo que el sujeto considera verdadero es lo que él mismo considera verdadero. En tercer lugar, habrá de atribuir al sujeto la capacidad de pensar, por lo general, de modo coherente. De acuerdo con Davidson (*Mente*), a menos que el intérprete acepte estos tres supuestos básicos acerca del sujeto a interpretar, no será capaz de dar sentido a las emisiones del hablante. Así, en virtud de estos principios, una interpretación intencional exitosa investirá *necesariamente* al agente interpretado de una *racionalidad básica*.

Ahora bien, si esta racionalidad básica es una condición de posibilidad de la atribución de estados intencionales a otro agente, cabe preguntarse qué sucede con aquellos casos en los cuales identificamos irracionalidad en el hablante. ¿Podemos detectar irracionalidad y al mismo tiempo, según el principio de caridad, considerar al agente como racional? El mismo Davidson ("*Paradoxes*", "*Incoherence*", *Problems*), consciente

de esta tensión, en distintos escritos sobre la irracionalidad ha brindado un argumento denominado como “*el argumento del trasfondo*”, que busca compatibilizar el principio de caridad con la posibilidad de atribuir irracionalidad.

En estos textos, Davidson nos recuerda que ningún estado particular *per se*, por ejemplo una creencia determinada –no importa lo extraña que pueda resultarle a los demás–, puede ser considerada en sí misma como irracional. Es decir, las creencias “nunca son irracionales en sí mismas, sino dentro de un patrón más amplio” (*Problems* 142), es decir, en relación con otros estados. Somos capaces de “dar sentido a las aberraciones [de la racionalidad] cuando éstas son vistas justamente dentro de un trasfondo de racionalidad” (190).

Así, el argumento del trasfondo sostiene que solo podemos interpretar a un agente como teniendo creencias irracionales si dicho agente mantiene un *trasfondo de racionalidad* en el resto de sus estados mentales. Esto es así, ya que sólo podremos dar sentido a tales creencias, e incluso a la caracterización de estas como “irracionales”, por referencia a los demás estados mentales y acciones del sujeto, los cuales –según el principio de caridad– deben poder ser vistos como siguiendo las normas de racionalidad. Esto no solo es indispensable para preservar la inteligibilidad de la interpretación, sino, además, como señala Davidson, porque “sin el elemento de racionalidad, nos negamos a aceptar la explicación como apropiada para un fenómeno mental” (196). En síntesis, las creencias irracionales pueden ser identificadas sólo por referencia y en contraste con la red general de creencias, otros estados mentales y acciones del agente, que exhiben un trasfondo de racionalidad y garantizan así el acto interpretativo.

Este principio, como he mencionado, no se considera como una sugerencia metodológica para la interpretación, sino como *condición de posibilidad* de ésta. O, en otros términos: si no asumimos el principio de caridad, no podríamos siquiera reconocer el comportamiento intencional de los agentes y, por lo tanto, tampoco podríamos reconocerlos como *agentes* ni podríamos atribuirles creencias y deseos en absoluto. Así, la justificación central del principio de caridad toma la forma de un argumento *trascendental*, en la medida en que pretende mostrar que las prácticas de atribución de estados mentales a agentes intencionales presuponen *necesariamente* la aplicación del principio de caridad. De modo tal que, el argumento de la irracionalidad cuando

apela a las constricciones de racionalidad de la *interpretación* no puede ser leído sino en este sentido trascendental y, bajo este sentido, la respuesta que cabe esgrimir es simple: si somos capaces *de hecho* de interpretar un delirio, léase, de comprender su contenido intencional (por ej. comprendemos que un sujeto crea ser perseguido), entonces el principio de caridad se encuentra operando. De otra manera, la interpretación se hubiera visto socavada³.

En el marco de este debate, algunos críticos del enfoque davidsoniano, como Bortolotti (“Intentionality”, “Delusions”, “In defence”) han intentado defender la tesis de que podemos atribuir creencias sin atribuir racionalidad. Es decir, bajo el mismo cometido de argumentar a favor del enfoque doxástico de los delirios, han empleado una estrategia inversa. En concreto, Bortolotti ataca la tesis davidsoniana de que debemos atribuir racionalidad para atribuir creencias y utiliza a los delirios de modo instrumental, como caso que, a su criterio, mostraría que podemos atribuir creencias sin atribuir racionalidad. Sin embargo, su crítica se basa en una lectura empírica del principio davidsoniano y no trascendental, a la vez que descansa en una inflación de los requisitos de racionalidad que demanda el acto de interpretación⁴ haciendo ver de este modo a los delirios como completamente irracionales (Vilatta).

Ahora bien, la suposición de racionalidad del principio de caridad, al estar justificada en un argumento trascendental, representa una condición *necesaria* para la interpretación y la atribución de estados mentales y no una afirmación empírica sobre el comportamiento de los agentes. Si se tratase de una afirmación de tipo empírica, ésta podría falsearse mediante ejemplos tal como intenta Bortolotti, no obstante, como bien señala Quintanilla, al ser un argumento trascendental, “ningún descubrimiento antropológico podría mostrar que es falso, sólo un análisis más fino de los conceptos de interpretación” (86). O puesto de otro modo, el argumento del trasfondo solo puede objetarse si se muestra que, en contra de Davidson, la interpretación es posible aunque no apliquemos el principio de caridad. Ninguna de las críticas ni

³ Esto podría suceder, por ejemplo, si nos encontramos ante una persona con un daño neurológico que profiera emisiones incoherentes sintáctica y semánticamente que no logren tener sentido para un intérprete. Esto no es, sin embargo, lo que sucede en el caso de los delirios, pues justamente se trata de una emisión lingüística semántica y sintácticamente comprensible.

⁴ No estoy interesada aquí hacer un análisis particular del enfoque de Bortolotti, sino antes bien, en presentar su caso como un ejemplo que ilustra cómo el argumento del trasfondo davidsoniano ha sido mal interpretado en los recientes debates filosóficos sobre los delirios.

a Davidson ni al enfoque doxástico ha logrado tal cometido, antes bien, se han dedicado a señalar características de irracionalidad que exhibirían los estados delirantes. Pese a ello, por mor del argumento, concedamos este punto y veamos si, interpretado bajo la versión empírica puede imposibilitarnos considerar a los delirios como creencias.

3. La irracionalidad de los delirios: la lectura empírica

Esta segunda variante del argumento parte de afirmar que un estado mental cuenta como una creencia sí y solo si tiene determinadas características funcionales. Los delirios no exhibirían estas características, por lo tanto, no serían creencias. La forma general del argumento sería la siguiente:

1. Una creencia tiene la propiedad P.
2. Los delirios no tienen la propiedad P.
3. Por lo tanto, los delirios no son creencias.

En concreto, las características que se ha argumentado que deberían tener las creencias son: veracidad/plausibilidad, sensibilidad a la evidencia y adecuada integración con otros estados mentales. Este argumento en la literatura ha tomado al menos tres formas centrales, las que denominaré como: “*el argumento de falsedad y los contenidos bizarros*”, el de la “*insensibilidad a la evidencia*” y el de la “*circunscripción*”. Veamos en detalle estos argumentos, considerando críticamente la segunda premisa, es decir, si los delirios efectivamente carecen de la propiedad P.

En primer lugar, hallamos el (1) “*argumento de la falsedad y los contenidos bizarros*”, el cual consiste en señalar que, puesto que las creencias tienen por objetivo la verdad, son plausibles y coherentes, los delirios no lo son, pues son falsos, con contenidos bizarros y altamente implausibles, por lo tanto, los delirios no serían creencias (Campbell, *et al.*). Sin embargo, en lo que respecta a la falsedad e implausibilidad del contenido, estos no parecen ser buenos criterios para distinguir a los delirios de otras creencias. Primero, porque algunos delirios pueden ser de hecho verdaderos, por ejemplo, alguien con un delirio celotípico que cree que está siendo engañado por su esposa, puede estar de hecho en lo correcto (Jaspers 109). Las personas con delirios no sufren por la falsedad de sus creencias, ni son catalogadas como personas con trastornos mentales por esto, sino antes bien por otras cualidades disfuncionales asociadas a sus delirios. Segundo, porque una creencia ordinaria puede ser falsa o bizarra y poco plausible, y no por

eso la catalogamos de delirante, pensemos por ejemplo en cualquier creencia supersticiosa cotidiana o en las teorías conspiracionistas. Es decir, no empleamos esos criterios ni en la atribución cotidiana de creencias ni se emplean en la clínica para evaluar a ciertas emisiones lingüísticas como delirantes.

En segundo lugar, tenemos el (2) “*argumento de la insensibilidad a la evidencia*”, cuya objeción se centra en señalar que los delirios no se vinculan con la evidencia tal como lo hacen las creencias. Estas se formarían con base en evidencia apropiada y se mantendrían o revisarían en función de la evidencia disponible. Luego, se señala que los delirios se formarían a partir de evidencia insuficiente y que se mantendrían a pesar de la evidencia en contra, por lo cual, no podrían ser considerados como creencias (Dub). Frente a este argumento, cabe decir que no es cierto que los delirios no responden *en absoluto* a la evidencia. Primero, porque los pacientes sí parecen tomar en consideración cierta evidencia que podría interpretarse como a favor de sus delirios. Un claro ejemplo es el del famoso paciente Schreber (10) quien afirma: “para mí existe un cúmulo en verdad abrumador de razones probatorias, que desearía que usted conociese en detalle por el contenido de mis Memorias”⁵. El problema, en todo caso, parece radicar más bien en que esas sean *buenas* o las *mejores* evidencias para formarse tal o cual creencia. Segundo, porque si tomamos en cuenta los resultados de efectividad de las terapias cognitivo-conductuales (CBT) para el tratamiento de los delirios (Freeman, Landa) la evidencia sugiere que los sujetos sí responden a la evidencia, aunque por supuesto con menor flexibilidad que sujetos no delirantes. Este enfoque, basado en la idea de que los delirios son un tipo de creencias (Alford & Beck), propone como una de las técnicas principales cuestionar los delirios a la luz de contra-evidencia, de modo progresivo y gradual. Uno de los procedimientos frecuentes en este tipo de terapia es el análisis de la evidencia y la generación de explicaciones alternativas (Read, Mosher & Bentall). El hecho de que el peso dado a la evidencia se modifique durante el tratamiento mostraría que los sujetos delirantes no necesariamente mantienen una convicción inquebrantable, sino que ésta puede variar gradualmente y, en este sentido, a medida que avanza el proceso terapéutico, los sujetos se vuelven cada vez menos reacios a someter a escrutinio sus creencias delirantes, es decir, aumentan su sensibilidad a la evidencia.

⁵ Las *Memorias* constituyen un libro en el cual presenta lo que él considera como evidencia justificatoria de sus delirios.

Consideremos ahora al tercer argumento, “(3)” el *argumento de la circunscripción*”, el cual destaca el hecho de que los delirios se encontrarían circunscriptos, refiriéndose con ello a la inercia de los delirios para afectar los otros estados mentales y la conducta (Campbell). Se destacan al respecto tres tipos de circunscripción. La circunscripción *inferencial* que refiere a que los sujetos delirantes fallarían en extraer las consecuencias lógicas de sus delirios y no resolverían las contradicciones existentes entre el delirio y el resto de sus creencias. La circunscripción *conductual* según la cual los sujetos delirantes no se comportarían como si creyeran en el contenido de sus delirios; y la circunscripción *emocional*, según la cual los sujetos delirantes no exhibirían las respuestas afectivas apropiadas.

Veamos aquí si esta caracterización de los delirios es justa. En cuanto a la circunscripción *inferencial*, la primera cuestión es: ¿realmente hay fallas de integración entre los delirios y las otras creencias del sujeto? La respuesta es a veces, pero no siempre ni de modo sistemático ni generalizado. La literatura clínica muestra que, en la mayoría de los casos, los delirios pueden fallar en integrarse con algunas creencias particulares, por lo general de alto valor emocional, pero se encuentran bien integrados con el resto de las creencias y actitudes intencionales del sujeto. Un buen ejemplo es el caso de un paciente hospitalizado con síndrome de Capgras que afirmaba que su padre era un impostor y cuando se le preguntaba por qué creía que éste fingía, daba respuestas como la siguiente: “esto es tan sorprendente doctor, ¿por qué alguien pretendería ser mi padre? Quizás mi padre empleó a alguien para que me cuide y le dio algún dinero para poder pagar mis cuentas” (Hirstein & Ramachadran 438). Su creencia delirante estaba integrada con otras creencias vinculadas a la necesidad de que alguien lo cuide en el hospital, a la probabilidad de que alguien lo hubiera contratado, etc. Del mismo modo que los pacientes con síndrome de Cotard (el delirio de estar muerto) afirman, por ejemplo, que creen que deben ser enterrados o que deben realizar un velorio en su honor, mostrando así la integración de su delirio a otras creencias.

Respecto a la supuesta circunscripción *conductual*, este resulta ser el argumento más difícil de sostener ya que es bien conocido que un gran número de pacientes actúan guiados por sus creencias delirantes. Por ejemplo, una paciente con síndrome de Cotard, negándose a almorzar en el hospital, afirma: “ya no necesito comer, pues estoy muerta” (Miralles 31). Otra paciente que también se

rehusaba a alimentarse, y que presentaba delirios corporales según los cuales le faltaban partes de su cuerpo, afirmaba: “no puedo comer porque no tengo boca” (Miralles 31). Casi todos los pacientes muestran alguna forma de conducta relacionada al delirio, es común por ejemplo que los pacientes con delirios persecutorios tengan conductas relacionadas a ello (tapar sus ventanas, no salir de su hogar, esconderse, etc.) (Freeman). De hecho, si no fuera el caso que los delirios afecten la conducta observable, no habría demasiado de lo cual preocuparse en términos de la salud mental del sujeto.

En cuanto a la supuesta circunscripción *afectiva*, se afirma que los sujetos delirantes fallarían en dar una respuesta emocional apropiada al contenido de sus delirios. Por ejemplo, esgrimen que cuando los pacientes que sufren de delirio de Capgras y afirman tener la creencia “mi cónyuge es un impostor”, fallan en reaccionar del modo en que se esperaría si los delirios fueran creencias, por ejemplo, sintiendo miedo y huyendo. Sin embargo, por una parte, no es cierto que los pacientes nunca actúen de modo consecuente con el contenido de sus delirios, como ya he señalado. Otro ejemplo claro de ello es el caso “BC” una mujer que cree que está siendo vigilada por dispositivos de alta tecnología. En terapia, la paciente relata que el sentir que escuchan todas sus conversaciones le genera una gran angustia, y tiene conductas compensatorias relacionadas a ello: gastó £300 para que eliminaran de su hogar estos dispositivos contratando a personas para que busquen los mismos detrás de las paredes. En otra ocasión, creyendo que su vecino la espiaba, rompió la ventana de éste y tuvo una crisis emocional que hizo que tuviera ser hospitalizada (Fulford, Thorton & Graham).

Hasta aquí podrá apreciarse que las tres versiones del argumento de la irracionalidad, en su variante empírica, realizan una caracterización de los delirios que exagera su irracionalidad al presentar a estos estados como careciendo por completo de una propiedad P (veracidad/plausibilidad, sensibilidad a la evidencia, integración). Ahora bien, si nuestro análisis plausible, cabría decir que la segunda premisa (que los delirios no tendrían la propiedad P) –al menos no siempre– resultaría cierta. Pues bien, consideremos ahora la primera premisa del argumento es decir si las creencias deben tener necesariamente esas propiedades para ser consideradas como tales.

4. Racionalidad óptima vs. racionalidad limitada

En esta sección pretendo mostrar que el argumento empírico, en sus distintas variantes, es esgrimido bajo una óptica de racionalidad óptima que parece tomar a las distintas propiedades racionales P como fenómenos de todo/nada que se presentan o están ausentes en las creencias. Las tres versiones del argumento implican que estas propiedades son un requisito necesario para que el estado sea considerado una creencia. Sin embargo, muchas de nuestras creencias ordinarias no presentan siempre estas propiedades. Son numerosas las críticas que se han esgrimido a la idea de una racionalidad óptima, es decir a la tendencia a asumir que los agentes humanos a los que atribuimos creencias poseen una racionalidad “máxima”. El supuesto de una racionalidad óptima resulta demasiado exigente y optimista respecto a los estándares de racionalidad humana reales (Martín, Kolodny, Cherniak, *Minimal*; Cherniak “Minimal rationality”).

La evidencia reciente en psicología cognitiva apunta al hecho de que la racionalidad de los humanos es *limitada* (Bortolotti, *Delusions*; Mandelbaum). En esta línea, Cherniak ha propuesto una idea de *racionalidad mínima*, la cual supone una capacidad limitada que está lejos de ser perfecta o ideal, pero que se ajusta mejor a las actuales investigaciones en psicología cognitiva experimental sobre el razonamiento humano que muestran que poseemos capacidades cognitivas acotadas y que frecuentemente cometemos algunos fallos en el razonamiento.

Así, desde un análisis de la racionalidad que no solo tenga justificación conceptual sino también *empírica*, lo “exigible” solo puede fijarse teniendo en cuenta cómo son *de hecho* los sujetos reales y cuál es su estructura y funcionamiento psicológico. Entonces, siguiendo el *argumento del trasfondo*, un mínimo de racionalidad es necesario para poder interpretar a otros agentes, sus deseos, creencias y acciones. Es decir, la racionalidad es una condición de posibilidad de la atribución de creencias. Pero a la vez, también es un hecho que los agentes reales (y no ideales) tienen capacidades cognitivas finitas y limitadas (computacionales y de memoria). Como señala irónicamente Cherniak (*Minimal* 51): “es realmente irracional aspirar a un ideal de racionalidad”, antes bien, la evidencia empírica sugiere que es más acorde caracterizar a la racionalidad humana como sub-óptima.

Como consecuencia, si un sujeto tiene un conjunto particular de creencias y deseos, el sujeto será racional “si realiza *algunas*, pero no todas, de las acciones que tiendan a satisfacer sus deseos a la luz de sus creencias” (5) y si realiza “*algunas* acciones que tiendan a satisfacer sus deseos tal como para él es el mundo” (5). Se denomina a esto “*condición de racionalidad mínima*”. Diferentes creencias pueden integrarse y responder a la evidencia disconfirmatoria en diferentes grados, dependiendo del tipo de información, los patrones cognitivos, las emociones asociadas, el contexto social, etc. Esta caracterización positiva se completa con una caracterización negativa: no debería ser el caso que el sujeto persistente y sistemáticamente muestre contradicciones, crea falsedades o realice acciones que no tiendan a satisfacer sus deseos a la luz de sus creencias. Es decir, no sería racional un sujeto que realice sistemáticamente estas acciones. Así, la racionalidad *mínima* debería ser su manera general de actuar, aceptándose la irracionalidad siempre que sea local y no global, de modo tal que no se socave la posibilidad de interpretar al agente.

Hay dos consecuencias importantes de esta idea de racionalidad mínima que merecen atención. La primera refiere al tipo de conocimiento que se debería tener al momento de formular una buena teoría de la racionalidad para la atribución de estados intencionales: no basta con tener una teoría del significado aplicada a las aserciones del hablante y una teoría general sobre la estructura conceptual de esos estados intencionales, sino que, además, es preciso conocer la estructura cognitiva real del sujeto. En este sentido, es importante que las teorías sobre cómo atribuimos estados intencionales recurran a evidencia empírica específica sobre las capacidades psicológicas de los agentes para poder delimitar, por un lado, los umbrales de racionalidad mínima que calificarían a un sujeto como un agente racional, y por otro, los umbrales de racionalidad que podrían serle normativamente exigibles (Cherniak). La segunda lección por extraer es la existencia de algún grado de compartimentalización de lo mental⁶ y el hecho de notar que la falla concomitante para delinear ciertas inferencias o resolver ciertas inconsistencias, no necesariamente cuenta como evidencia de que un agente es irracional o de que sus estados no sean calificados como creencias. Más bien, dadas nuestras capacidades limitadas, alguna irracionalidad local parece ser el precio que pagar por la racionalidad global. Esto es precisamente lo que parece

⁶ Que existan compartimentos en la mente que sean coherentes a su interior pero que puedan generar inconsistencias con otros en la medida en que se encuentran –al menos parcialmente– separados o encapsulados.

suceder no solo en los casos de delirios sino también con muchas de nuestras creencias ordinarias. Veamos esto en detalle.

En cuanto a la verdad/plausibilidad, una creencia ordinaria puede ser falsa, implausible o bizarra y no por eso la catalogamos de delirante. Por ejemplo, la creencia falsa de un estudiante de que si utiliza una estampita le irá bien en un examen, o las creencias conspiracionistas de que hay agentes del gobierno que controlan el clima. Como mencionamos previamente, la falsedad o rareza de una idea no parece ser un buen criterio para distinguir ideas delirantes de no-delirantes y en la medida en que las creencias ordinarias también pueden ser falsas, implausibles o bizarras esta característica no parece ser un impedimento para considerar a los delirios como creencias.

Respecto a la sensibilidad a la evidencia, la idea de que las creencias deben formarse siempre a la luz de buena evidencia y de responder adecuadamente ante la evidencia en contra es demasiado exigente incluso para las creencias ordinarias y diversos estudios empíricos han mostrado que buena parte de nuestras creencias no responden a la evidencia disconfirmatoria (Bortolotti, *Delusions*; Mandelbaum). Por ejemplo, en los últimos años numerosas personas dicen creer que la tierra es plana, contra toda evidencia científica. Otro grupo de personas afirma que beber jugo de limón previene el cáncer, nuevamente contra toda evidencia. A la vez que sabemos que es difícil para las personas considerar evidencia asociada a contenidos emocionalmente cargados, por ejemplo, alguien puede ignorar toda la evidencia que le indica que su pareja le engaña, solo porque desearía que eso no fuera verdad. Asimismo, debe tenerse en consideración, que la mayor parte de las personas en muchas ocasiones creen cosas solo por haber visto o escuchado algún relato sin ningún testeo de confiabilidad apropiado. Alguien, por ejemplo, puede creer supersticiosamente —pero con mucha convicción— que su humor actual se debe a la posición de la luna, aunque no haya evidencia sólida que justifique tal afirmación.

Es decir, existen muchas creencias ordinarias que carecerían del grado apropiado de relación con la evidencia que demandaba el requisito analizado. Nuevamente aquí, nadie pone en duda que tales afirmaciones sean creencias, sino que simplemente evaluamos la verdad o falsedad de estas. Es decir, podemos cuestionar el valor de dicha evidencia y su peso, pero no cuestionamos que la persona esté *creyendo eso*, pues al fin

de cuentas es una aseveración formada con base en —probablemente no la mejor— evidencia (pero evidencia al fin) que se integra al proceso de la elaboración de la creencia. Si aplicamos el mismo criterio para el caso de los delirios, no parece haber razones contundentes entonces para negarles de inmediato su carácter doxástico.

En cuanto a la integración, también cabe señalar que las creencias no siempre están bien integradas unas con otras. Es muy común, por ejemplo, que alguien crea que fumar es perjudicial para la salud, pero también crea que fumar es una actividad placentera y que no la abandone pese a que diga que quiere cuidar su salud. También sucede esto usualmente con los prejuicios racistas y sexistas. Así, por ejemplo, en cierto contexto alguien puede decir que el sexismo es éticamente incorrecto y, en otro, manifestar creer que las mujeres son menos capaces de realizar ciertos trabajos. Del mismo modo, no escasean ejemplos de personas que se dedican a la ciencia y creen en sus métodos, aunque, al mismo tiempo, también creen, de manera inconsistente, en procedimientos sin evidencia como en terapias alternativas, chacras o lectura del tarot. En ninguno de estos ejemplos ordinarios las creencias parecen estar adecuadamente integradas, antes bien, se mantienen separadas y se activan en diferentes contextos. Cabe aclarar que esto no implica negar que los delirios estén menos integrados que las creencias ordinarias. Antes bien, lo que pretendo es señalar que la integración doxástica parece ser una cuestión de grados.

Por último, tener una creencia es diferente de actuar guiado por esta. No siempre actuamos de acuerdo con nuestras creencias. Pensemos nuevamente en el ejemplo previo de los fumadores, alguien puede creer que debería dejar de fumar y no lo hace. O alguien podría considerar que comer comida ultra procesada es perjudicial y seguir haciéndolo. Son numerosos los casos de *akrasia* (cuando actuamos en contra de nuestro mejor juicio) en la vida cotidiana que evidencian que no siempre nuestras creencias guían nuestras acciones. También debe considerarse que a veces las creencias son causalmente inertes, si las pensamos como disposiciones, las mismas se activan solo en algunos contextos, tal como señalamos que sucedería en algunos casos de prejuicios sexistas.

En síntesis, el argumento de la irracionalidad se apoya en una suposición de racionalidad óptima (de agentes racionalmente ideales), pasando por alto el cúmulo de evidencia psicológica reciente que indica que nuestras creencias ordinarias muchas veces no se ajustan a ciertos requerimientos de racionalidad y que los agentes humanos reales ordinariamente manifiestan

una racionalidad mínima o limitada. En otras palabras, este argumento no solo parece asumir una visión de todo/nada de estos requisitos respecto a las propiedades de racionalidad exigibles, sino que además supone constricciones sobre el fenómeno delirante que, con frecuencia, ni siquiera muchas de nuestras creencias ordinarias logran satisfacer.

5. Consideraciones finales

En el presente artículo me he ocupado del debate filosófico en torno a la posibilidad de atribuir a los delirios el estatus de creencias, considerando la irracionalidad que estos estados manifiestan y el requisito interpretativo de atribuir racionalidad para atribuir estados doxásticos. Según he señalado, la crítica central al modelo doxástico reside en lo que he denominado como “*el argumento de la irracionalidad*”, el cual afirma que los delirios son demasiados irracionales para ser considerados como creencias. Por lo anterior, este argumento se presenta bajo dos lecturas posibles en la literatura reciente: la variante trascendental y la variante empírica.

Por un lado, en su variante *trascendental*, la racionalidad es una *condición de posibilidad* para la atribución de creencias. Por otro lado, en su variante *empírica* apunta a señalar que el rol funcional de las creencias implicaría cumplir determinados requisitos de racionalidad que los delirios no satisfarían. En cuanto a la primera variante, he argumentado que, bajo la lectura interpretivista y apoyándome en el “*argumento del trasfondo*” elaborado por Davidson, en la medida en que podamos comprender la emisión lingüística del sujeto delirante y así atribuir el contenido del delirio a una persona, las condiciones mínimas de racionalidad necesarias para la interpretación se encuentran garantizadas. En consecuencia, no hay razones de peso para pensar que la irracionalidad de los delirios socava de inmediato la posibilidad de la atribución doxástica, ni para cuestionar la tesis de que la atribución de creencias implica racionalidad. O, en otras palabras, es posible atribuir a la persona la *creencia* delirante irracional al mismo tiempo que sostenemos la tesis de que la atribución de creencias implica racionalidad.

Finalmente, he argumentado que, incluso si el argumento de la irracionalidad es leído de manera *empírica* no representa un argumento contundente y decisivo en contra del enfoque doxástico ni contra la tesis de que la atribución de creencias implica racionalidad. Por un lado, porque los delirios no son tan irracionales como aparece *prima facie* y, por otro, porque el modo en que ha elaborado este argumento supone

requisitos de racionalidad demasiado exigentes incluso para las creencias ordinarias. De este modo, he mostrado cómo es posible compatibilizar la tesis de que para atribuir creencias debemos atribuir racionalidad con la posibilidad de atribuir a los delirios un carácter doxástico.

REFERENCIAS

Alford, Brad A., & Aaron T. Beck. "Cognitive therapy of delusional beliefs". *Behaviour Research and Therapy*, 32 (3), 1994: 369-380. Print.

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington: American Psychiatric Pub, 2013. Print.

Bayne, Tim, & Pacherie, Elisabeth. "In defence of the doxastic conception of delusions". *Mind & Language*. 20 (2), 2005: 163-188. Print.

Berrios, Germán E. *Hacia una nueva epistemología de la psiquiatría*. Buenos Aires: Polemos. 2011. Impreso.

Bortolotti, Lisa, and Miyazono, Kengo. "Recent work on the nature and development of delusions". *Philosophy Compass*. 10 (9), 2015: 636-645. Print.

"In defence of modest doxasticism about delusions." *Neuroethics*. 5 (1), 2012: 39-53. Print.

Delusions and other irrational beliefs. Oxford: Oxford University Press, 2010. Print.

"Intentionality without rationality". *Proceedings of the Aristotelian Society*. 105 (1). Oxford: Oxford University Press, 2005. Print.

Campbell, John. "Rationality, meaning, and the analysis of delusion". *Philosophy, Psychiatry, & Psychology*. 8 (2), 2001: 89-100. Print.

Cermolacce, Michel; Sass, Louis & Parnas, Joseph. "What is bizarre in bizarre delusions? A critical review". *Schizophrenia Bulletin*. 36 (4), 2010: 667-679. Print.

Cherniak, Christopher. *Minimal rationality*. Massachusetts: MIT Press, 1986. Print.

"Minimal rationality and optimal brain wiring". *Proceedings 13th International Congress of Logic, Methodology & Philosophy of Science*, 2009. Print.

Currie, Gregory. "Imagination, delusion and hallucinations". *Mind & Language*. 15 (1), 2000: 168-183. Print.

Díez, Patricio Antonio. "Creencia y delirio". *Revista de la Asociación Española de Neuropsiquiatría*. 31 (1), 2011: 71-91. Impreso.

Donald, Davidson. "Paradoxes of irrationality". *Philosophical Essays on Freud*. Cambridge: Wollheim, Richard & Hopkins, Cambridge University Press, 1982. Print.

Mente, mundo y acción: claves para una interpretación. Barcelona: Gedisa, 1992. Impreso.

"Incoherence and irrationality". *Dialéctica*. 39(4), 1985: 345-354. Print.

Inquiries Into Truth and Interpretation: Philosophical Essays Volume 2. Oxford: Oxford University Press, 2001. Print.

Problems of rationality. Volume 4. Oxford: Oxford University Press, 2004. Print.

Dub, Richard. "The rationality assumption". *Content and consciousness revisited*. Berlin: Muñoz-Suárez, Brigard, Springer Cham, 2015: 93-110. Print.

Egan, Andy. "Imagination, Delusion and Self - deception". *Delusion and Self - deception: Affective and Motivational Influences on Belief Formation*. New York: Bayne, Hova, 2008: 263-80. Print.

Frankish, Keith. "Delusions, levels of belief, and non-doxastic acceptances". *Neuroethics*. 5 (1), 2012: 23-27. Print.

Freeman, Daniel. "Persecutory delusions: a cognitive perspective on understanding and treatment". *The Lancet Psychiatry*. 3 (7), 2016: 685-692. Print.

Fulford, Kenneth; Thornton, Tim & Graham, George. *Oxford textbook of philosophy and psychiatry*. Oxford: Oxford University Press, 2006. Print.

Hirstein, William & Ramachandran, Vilayanur. "Capgras syndrome: a novel

probe for understanding the neural representation of the identity and familiarity of persons". *Proceedings of the Royal Society of London*, 1997: 437-444. Print.
 Jaspers, Karl. *Psicopatología general*. Barcelona: Beta, 1963. Print.

Klee, Robert. "Why some delusions are necessarily inexplicable beliefs". *Philosophy, Psychiatry, & Psychology*. 11 (1), 2004: 25-34. Print.
 Kolodny, Niko. "Why be rational?" *Mind*. 114 (455), 2005: 509-563. Print.

Landa, Yulia, Silverstein, Steven, Schwartz, Fred & Savit, Adam. "Group cognitive behavioral therapy for delusions: helping patients improve reality testing". *Journal of Contemporary Psychotherapy*. 36 (1), 2006: 9-17. Print.

López-Silva, Pablo. "The typology problem and the doxastic approach to delusions". *Filosofía Unisinos*. 17 (2), 2016: 202-211. Print.

Ludwig, Kirk. *Donald Davidson*. Cambridge: Cambridge University Press, 2003. Print.

Mandelbaum, Eric. "Troubles with Bayesianism: An introduction to the psychological immune system." *Mind & Language*. 34 (2), 2019: 141-157. Print.

Martín, María Isabel; Pérez Chico, David; Liz, Manuel; López-Flrido, Germán; Fernández, Nuria & Pérez Torres, Lorenzo. "Aspectos internos y externos de la racionalidad: ¿cuáles son los umbrales de la racionalidad exigible y cómo situarlos?" *Contextos*. 29 (1), 1997: 153-178. Impreso.

Miralles, Inmaculada; González Caballero, Gloria; Piñol, Belén; Morales, Concepción; Navajas, Mar y Sáenz, María Josefa. "Síndrome de Cotard asociado a demencia degenerativa. Presentación de tres casos". *Alzheimer Real Invest Demenc*, 56 (1), 2014: 30-36. Impreso.

Read, John, Mosher, Loren & Bentall, Richard. *Modelos de locura: aproximaciones psicológicas*. Madrid: Herder, 2005. Impreso.

Reimer, Marga. "A Davidsonian perspective on psychiatric delusions". *Philosophical Psychology*. 24 (5), 2011: 659-677. Print.

“Davidsonian holism in recent philosophy of psychiatry”. *Donald Davidson on Truth, Meaning, and the Mental*. Oxford: Preyer, Oxford University Press, 2012. Print.

Schreber, Daniel Paul. *Memorias de un enfermo de nervioso*. Buenos Aires: Perfil, 2008. Impreso.

Schwitzgebel, Eric. “Mad belief?” *Neuroethics*. 5 (1), 2012: 13-17. Print.

Stephens, Lynn & Graham, George. “Reconceiving delusion”. *International Review of Psychiatry*. 16 (3), 2004: 236-241. Print.

Quintanilla, Pablo. “Comprensión, imaginación y transformación”. *Areté*. 20 (1), 2008: 111-136. Impreso.

Vilatta, Emilia. “Un enfoque davidsoniano de los delirios: el caso del delirio de Capgras”. *Areté*. 29 (1), 2017: 183-212. Impreso.

Como citar:

Vilatta, Emilia. “Atribución de creencias e irracionalidad: el caso de los delirios”. *Discusiones filosóficas*. Jul. 21 (39), 2021: 15-34. <https://doi.org/10.17151/difil.2021.22.39.2>