

Influencias no pertinentes, desacuerdos, y razonamiento desacreditador: un análisis a partir del autoengaño

*Irrelevant Influences, Disagreements, and Debunking Reasoning:
An Analysis Based on Self-Deception*

Agostina Vorano

Sociedad Argentina de Análisis Filosófico / Consejo Nacional de Investigaciones Científicas y Técnicas
/ Universidad de Buenos Aires / Universidad Nacional de Mar del Plata; Argentina
vorano.agostina@gmail.com

Resumen

En el presente trabajo, objetaremos la hipótesis según la cual el desafío epistémico generado por las influencias no pertinentes se reduce al suscitado por los desacuerdos entre pares epistémicos que se reconocen como tales. Desarrollaremos nuestra objeción considerando como caso paradigmático de influencia no pertinente al deseo que obra como factor motivacional en casos típicos de autoengaño. Asimismo, examinaremos el rol del razonamiento desacreditador en los casos de desacuerdos en los cuales una de las personas discordantes se encuentra autoengañada.

Palabras clave: desacuerdos entre pares epistémicos reconocidos, influencias no pertinentes direccionales, ansiedad de contingencia, irracionalidad motivada, sesgos.

Abstract

In this paper, we will object to the hypothesis according to which the epistemic challenge posed by irrelevant influences can be reduced to the one posed by acknowledged epistemic peer disagreement. We will develop our objection considering as a paradigmatic example of an irrelevant influence the desire that acts as a motivational factor in typical cases of self-deception. Furthermore, we will examine the role played by the debunking reasoning in the cases of disagreement in which one of the conflicting persons is self-deceived.



Received: 19/09/2022. Final version: 27/10/2023

eISSN 0719-4242 – © 2023 Instituto de Filosofía, Universidad de Valparaíso

This article is distributed under the terms of the

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Internacional License



CC BY-NC-ND

Keywords: acknowledged epistemic peer disagreement, directional irrelevant influence, contingency anxiety, motivated irrationality, biases.

1. Influencias no pertinentes y ansiedad de contingencia

Todos estamos familiarizados con observaciones como las siguientes: “crees que sus cuadros son los peores de la exposición porque tienes una pésima relación personal con él y deseas que fracase en su carrera artística”, “crees en el karma porque tus padres son budistas” o “aceptas el psicoanálisis lacaniano porque estudiaste en la Universidad de Buenos Aires”. Claramente, observaciones de este tipo encierran una crítica epistémica contra aquel a quien se dirigen: el núcleo de dicha crítica radica en que las creencias en cuestión no parecen estar basadas en evidencia pertinente, sino que parecen responder a factores carentes de valor epistémico. En los debates contemporáneos sobre esta temática, dichos factores suelen denominarse “influencias no pertinentes” o “factores arbitrarios de trasfondo” (de ahora en más, emplearemos sólo la primera designación, a fin de facilitar la exposición).

Vavova (2018, p. 136) ofrece la siguiente descripción acerca de dichas influencias:

Una influencia sobre una creencia es no pertinente en relación con una creencia y una persona. En otras palabras, una influencia no pertinente para mí con respecto a mi creencia de que p es una que (a) ha influido sobre mi creencia de que p y (b) no es pertinente en relación con la verdad de p .

Por su parte, Mogensen (2015, p. 592) las describe en los siguientes términos:

(...) [a] una condición C es *arbitraria* en relación con la creencia de S de que p si y sólo si el que S sepa que C ocurre no le provee (o no le proveería) razón alguna a S para creer que p . [...] [b] una condición C es un *factor de trasfondo* en relación con el hecho de que S crea que p si y sólo si C contribuye a la explicación del hecho de que S crea que p , pero C no forma parte de las razones sobre las cuales S basa su creencia de que p . (Cursivas en el original.)

Avnur y Scott-Kakures (2015) ofrecen una conceptualización acerca de las influencias no pertinentes que se solapa en buena medida con las descripciones que ya hemos recapitulado. Pero, además, introducen una distinción ausente en dichas caracterizaciones: diferencian las “influencias posicionales” de las “influencias direccionales”. Las primeras “(...) nos colocan en una posición en particular, en vez de otra, para poseer ciertas razones [...] causan que poseamos, o nos ponen en una posición para apreciar, cierta evidencia (...)” (Avnur & Scott-Kakures, 2015, pp. 10-12). En cambio, las influencias direccionales “(...) causan que nuestro *manejo de la evidencia* favorezca un resultado particular, predeterminado (...)” (Avnur & Scott-Kakures, 2015, p. 12; cursivas en el original). Estos dos autores se focalizan en los

deseos como -potenciales- influencias no pertinentes¹. Ahora bien, ellos sostienen que los deseos que ejercen una influencia de índole posicional no son epistémicamente problemáticos, mientras que los que ejercen una influencia de índole direccional sí lo son. Para dar cuenta de esta diferencia, retomaremos un ejemplo propuesto por los autores citados: supongamos que A detesta al director del instituto en el que trabaja, y por ello le complacería enterarse de que aquél está involucrado en algún escándalo que amenace su puesto laboral. Por otro lado, A desea comer una pizza, y por eso concurre a cierto local para comprarla. Mientras espera, se encuentra con un colega confiable, quien le informa que ha visto al director del instituto involucrado en alguna actividad reprochable. De no haber sido por el deseo de comer pizza, A no creería que el director del instituto ha realizado algo inadecuado; el deseo de comer pizza lo ha colocado a A en una posición en la que pudo obtener evidencia testimonial pertinente. Así, la creencia que A se forma es contingente en relación con su deseo de comer pizza; pero esta contingencia no pone en duda la justificación de dicha creencia. Supongamos, en cambio, que A obtiene información por parte de un colega confiable acerca del director del instituto, según la cual éste no parece haberse visto involucrado en ninguna acción cuestionable. Sin embargo, el deseo de A de que el director del instituto sea despedido incide en la manera en que A evalúa la información que ha obtenido, de modo tal que A arriba sesgadamente a la creencia de que el director estuvo involucrado en acciones reprochables. En este caso, que A obtenga evidencia acerca de su deseo de que el director del instituto sea despedido sí pone en jaque la justificación de su creencia de que el director ha hecho algo inapropiado; aquí, dicho deseo actúa como una influencia direccional. Puede advertirse, pues, que las influencias no pertinentes posicionales son epistémicamente inocuas. Por el contrario, las direccionales resultan ser problemáticas.

Así pues, la evidencia de que una creencia ha sido formada bajo una influencia no pertinente de índole direccional constituye un derrotador. Avnur y Scott-Kakures (2015) apelan a dos argumentos para defender esta idea. El primero de ellos es relativo a la confiabilidad del proceso de formación de creencias: si uno obtiene evidencia según la cual dicho proceso no ha sido confiable, o lo ha sido en menor medida de lo que debería, dicha evidencia constituye un derrotador. Si el proceso de formación de una creencia se encuentra sujeto a una influencia direccional, es menos confiable que si no se encontrara bajo dicha influencia; por ese motivo, evidencia según la cual el proceso por el cual uno adquirió cierta creencia se encontraba efectivamente sujeto a una influencia direccional constituye evidencia de que dicho proceso ha sido menos confiable de lo que habría sido en caso contrario. En este punto, es relevante señalar que los deseos no son anti- o contra-confiables: desear que p no es un indicador de que $\neg p$. Cabe preguntarse, entonces, por qué la influencia direccional ejercida por el deseo de que p hace que el proceso de formación de la creencia de que p sea menos confiable. Lo que ocurre es que (a menos que se disponga de evidencia según la cual existe una correlación significativa

¹ Estos autores aclaran que interpretan la noción de “deseo” de manera amplia, incluyendo todo tipo de intereses y aversiones.

entre el que uno desee que p y el que p sea verdadera) creer en consonancia con aquello que uno desea es tan confiable como adoptar creencias al azar. El segundo argumento esgrimido por Avnur y Scott-Kakures es relativo a consideraciones generales acerca de las normas epistémicas que rigen la formación de creencias: en términos generales, disponer de razones para creer que un tercero o uno mismo manipuló la evaluación de la evidencia con la que contaba constituye razón para creer que el manejo de dicha evidencia no respondió exclusivamente a metas estrictamente epistémicas. Por esta razón, se torna necesario reconsiderar la legitimidad de la transición que conduce desde el juicio de que uno posee buenas razones para creer que p hacia la efectiva adopción de la creencia de que p .

Como puede apreciarse a partir de los fragmentos previamente citados, las caracterizaciones ofrecidas por Vavova (2018) y por Mogensen (2015) acerca de las influencias no pertinentes abarcan tanto las de índole posicional como las de índole direccional. En el presente trabajo, nos centraremos en las influencias de este último tipo. Ello responde, por un lado, a que consideramos adecuada la tesis de Avnur y Scott-Kakures (2015) según la cual sólo las influencias direccionales son epistémicamente problemáticas. Y, por otro lado, a que el fenómeno que analizaremos (el deseo que obra como factor motivacional en los casos de autoengaño directo) constituye, por razones que aclararemos en la cuarta sección, una influencia de esta índole.

Así pues, cuando advertimos que algunas de las creencias que sostenemos son el producto de influencias no pertinentes direccionales, deberíamos experimentar cierta “intranquilidad” epistémica, que Mogensen (2015) denomina “ansiedad de contingencia”. En otras palabras, advertir la incidencia de dichas influencias sobre nuestras creencias debería resultarnos preocupante. Es en este sentido que obtener evidencia según la cual nos encontramos bajo cierta influencia no pertinente direccional suscita alguna clase de desafío epistémico. Dicho desafío nos inclina (o, al menos, debería inclinarnos) a reducir nuestra confianza en las creencias en cuestión, a revisarlas y, en consecuencia, a modificarlas de alguna manera. Ahora bien, ¿qué clase de desafío epistémico es éste? ¿Se trata de un tipo de desafío específico, *sui generis*? ¿O, por el contrario, dicho desafío podría ser reducido a un desafío de alguna otra clase?

2. La hipótesis del desacuerdo

Los cuatro autores recién citados (Avnur & Scott-Kakures, 2015; Mogensen, 2015; Vavova, 2018) han examinado la posibilidad de que el desafío suscitado por las influencias no pertinentes se reduzca al suscitado por los desacuerdos. En este terreno, no hay consenso: Avnur y Scott-Kakures y Vavova juzgan que las influencias no pertinentes generan un desafío epistémico *sui generis*; por el contrario, Mogensen defiende lo que él denomina “la hipótesis del desacuerdo”, según la cual, precisamente, el primer tipo de desafío es reductible al segundo. A continuación, examinaremos sintéticamente los argumentos esgrimidos a favor y en contra de dicha hipótesis.

Comenzaremos analizando la propuesta de Mogensen (2015). Este autor, tras evaluar algunas alternativas a la hipótesis del desacuerdo, concluye que ésta es la opción más robusta para dar cuenta de la importancia epistémica de los casos que involucran influencias no pertinentes. Mogensen argumenta a favor de la hipótesis del desacuerdo haciendo frente a dos críticas dirigidas en su contra. Según la primera de estas críticas, es cierto que en muchos casos de influencias no pertinentes existe desacuerdo entre las partes involucradas, pero es asimismo cierto que esto no ocurre en muchos otros casos. Mogensen enfrenta esta objeción señalando que en los casos de influencias no pertinentes en los que no se registra desacuerdo alguno, la ausencia de éste es meramente arbitraria; a esto, añade la idea según la cual la relevancia epistémica de los desacuerdos arbitrariamente ausentes es equivalente a la de los desacuerdos efectivamente existentes. De acuerdo con la segunda de las críticas dirigidas en contra de la hipótesis del desacuerdo, parecería inapropiado sostener que en los casos de ansiedad de contingencia lo único que reviste importancia epistémica son los desacuerdos (actuales o potenciales) en juego, en vez de las influencias no pertinentes. Mogensen asevera que la hipótesis del desacuerdo es compatible con la intuición que motiva esta objeción; esto se debe a que la significatividad epistémica del desacuerdo se encuentra modulada por el conocimiento que se posea acerca de las causas de dicho desacuerdo. Ahora bien, de acuerdo con este autor, lo fundamental es la existencia del desacuerdo, mientras que disponer de conocimiento sobre sus causas reviste una importancia meramente derivada.

Proseguiremos ahora con el análisis de los argumentos en contra de la hipótesis del desacuerdo. Comenzaremos con los planteos de Avnur y Scott-Kakures (2015), quienes apelan a dos consideraciones a fin de rechazar dicha hipótesis. En primer lugar, observan que la extensión del desafío epistémico generado por las influencias no pertinentes es diferente de la extensión del desafío suscitado por los desacuerdos: hay casos de influencias no pertinentes en los que no está involucrado desacuerdo alguno, y hay casos de desacuerdos en los que no hay influencias no pertinentes en juego. En segundo lugar, Avnur y Scott-Kakures señalan que difiere la naturaleza de los derrotadores involucrados en cada caso. Estos autores contemplan dos tipos de derrotadores frente a una creencia de que p : los derrotadores refutantes [*outweighing defeaters*], que constituyen evidencia a favor de la creencia de que $\neg p$; y los derrotadores socavantes [*undermining defeaters*], que constituyen evidencia según la cual uno carece de justificación para creer que p . Avnur y Scott-Kakures aseveran que en los casos de desacuerdos pueden estar involucrados derrotadores de cualquiera de estos dos tipos, pero que, en cambio, los casos de influencias no pertinentes únicamente involucran derrotadores socavantes (puesto que el que una persona advierta que, por ejemplo, su deseo de que p ha sesgado sus creencias al respecto no constituye evidencia a favor de $\neg p$; por el contrario, constituye evidencia según la cual la persona carece de justificación para creer que p). Una diferencia relevante es la que existe entre los derrotadores socavantes y los derrotadores constituidos por evidencia de segundo orden (Lasonen-Aarnio, 2014). Aún luego de que una persona obtenga un derrotador del primer tipo, puede seguir afirmándose que su evaluación inicial de la evidencia de la que disponía fue racional. Por el contrario, la evidencia de segundo orden pone de manifiesto

que la actitud doxástica en juego nunca fue racional; en otras palabras, la evidencia de segundo orden no es evidencia que vuelva injustificada la creencia en cuestión -inicialmente racional- sino que es evidencia de que dicha creencia nunca estuvo justificada. Esta diferencia, que no debe ser perdida de vista, es desestimada por Avnur y Scott-Kakures (2015, p. 34). Ahora bien, teniendo presente la importancia de dicha distinción, cabe remarcar que los casos de influencias no pertinentes direccionales involucran derrotadores constituidos por evidencia de segundo orden.

A diferencia de Avnur y Scott-Kakures (2015), Vavova (2018) se aproxima a la concepción de Mogensen (2015) puesto que concede la posibilidad de que todo caso de influencia no pertinente involucre desacuerdo al menos de manera potencial (de modo tal que sería meramente contingente el hecho de que en ciertos casos de influencias no pertinentes no se registre desacuerdo alguno). Ahora bien, Vavova señala que, aún si esto fuera cierto, no es claro que un desacuerdo meramente potencial sea epistémicamente problemático: en primer lugar, parece razonable sostener que la carga de la prueba debería recaer sobre quien quisiera defender aquella tesis; además, visto y considerando que la mera posibilidad de desacuerdo es ubicua, sería menester mostrar que dicha línea argumental no conduce en última instancia a un escepticismo global. Por otro lado, Vavova observa que, incluso si todo caso de influencia no pertinente supusiera desacuerdo (aunque no fuera sino un desacuerdo meramente potencial), de aquí no se sigue que sea dicho desacuerdo lo que suscite el desafío epistémico en juego. Por el contrario, el desacuerdo en cuestión podría ser sencillamente un fenómeno emergente, un mero indicador de aquellos factores que genuinamente generan el desafío epistémico. Por último, Vavova señala que no todo caso de desacuerdo es, a la vez, un caso de influencia no pertinente; así, por ejemplo, desacuerdos sobre problemas aritméticos sencillos muy frecuentemente se explican debido a que una de las personas involucradas ha cometido un error de cálculo.

A pesar de que se contraponen, tanto las propuestas de Avnur y Scott-Kakures (2015) y Vavova (2018) como la concepción de Mogensen (2015) se erigen sobre una misma confusión. Ésta consiste en que dichos investigadores no contemplan específicamente desacuerdos entre *pares epistémicos que se reconocen como tales*, sino que abordan la noción de desacuerdo en términos generales, sin ulteriores cualificaciones.

Ahora bien, ¿cualquier tipo de desacuerdo es epistémicamente problemático? Parece existir un amplio consenso según el cual la respuesta a este interrogante es negativa: los desacuerdos que revisten importancia epistémica son aquellos que se desarrollan entre pares epistémicos que se reconocen como tales -o “pares epistémicos reconocidos”, para abreviar- (Kelly, 2005; Feldman, 2006; Christensen, 2007, 2009; Elgin, 2010). Para que quienes participan de un desacuerdo puedan ser considerados como pares epistémicos reconocidos, deben satisfacerse las siguientes condiciones: el desacuerdo debe ser genuino; ambas partes deben poseer la misma evidencia; ambas partes deben exhibir igual disposición a responder a dicha evidencia de

una manera epistémicamente apropiada; y ambas partes deben poseer buenas razones para creer que las condiciones previas se satisfacen. En la próxima sección, analizaremos cada una de estas condiciones.

No todo desacuerdo, entonces, reviste importancia epistémica. Así, por ejemplo, Kelly (2005, p. 168) afirma que:

[p]or supuesto, el mero hecho del desacuerdo no necesita ser problemático: si, por ejemplo, yo estuviera convencido de que todos aquellos que desacordaron conmigo simplemente eran tontos, o no se habían molestado en pensar acerca del asunto con suficiente cuidado, o no estaban familiarizados con evidencia que yo poseo (evidencia que, si se les presentara, resultaría en que cambien su perspectiva), entonces yo simplemente podría desestimar ese desacuerdo.

No todo desacuerdo se desarrolla, pues, entre pares epistémicos reconocidos. Por el contrario, obviamente existen numerosos escenarios en los cuales una de las personas discordantes es epistémicamente superior respecto de la otra, en lo que concierne al tópico que es objeto del desacuerdo. Si dicha persona epistémicamente superior adquiere la creencia justificada de que aquél con quien desacuerda es inferior epistémicamente en lo que concierne al tópico en juego, esto no le proporciona razón alguna para modificar su actitud doxástica acerca de dicho tópico. Por el contrario, la persona que es superior epistémicamente está justificada en retener su actitud doxástica original. En cambio, si la persona epistémicamente inferior adquiere la creencia justificada de que aquél con quien desacuerda es su superior epistémico respecto del tópico en juego, entonces dicha persona debería modificar su actitud doxástica relativa a dicho tópico. Como puede apreciarse, no parece haber espacio para controversias en cuanto a cómo deben proceder en este caso la persona epistémicamente superior y la persona epistémicamente inferior: precisamente, cuando quienes desacuerdan difieren en sus jerarquías epistémicas, el desacuerdo en cuestión no reviste carácter problemático.

Ahora bien, cuando un desacuerdo se gesta entre pares epistémicos reconocidos, no resulta sencillo establecer cómo deberían dichas personas ajustar sus actitudes doxásticas respecto de la proposición en disputa. ¿Podría ser racional que las personas involucradas en un desacuerdo de esta índole conserven sus actitudes doxásticas originales en cuanto a la proposición en juego? ¿O, por el contrario, es menester que alguna de estas personas, o ambas, modifiquen dichas actitudes? En caso de que esto último sea lo correcto, ¿qué modificación sería racional llevar a cabo? A diferencia de lo que ocurría en el caso que consideramos previamente -en el cual eran disímiles las jerarquías epistémicas de las personas discordantes-, aquí no es sencillo responder estas preguntas. Es por ello que los desacuerdos entre pares epistémicos reconocidos suscitan un genuino desafío epistémico.

En síntesis, consideramos que los autores cuyos aportes hemos analizado (Avnur y Scott-Kakures, 2015; Mogensen, 2015; Vavova, 2018) han abordado inadecuadamente la problemática referida a las relaciones entre el desafío epistémico suscitado por las influencias no pertinentes y el generado por los desacuerdos. Esto se debe a que se han referido a los

desacuerdos de manera genérica, sin contemplar una distinción que -por las razones que acabamos de exponer- consideramos que es crucial: aquella que existe entre, por un lado, los desacuerdos entre pares epistémicos reconocidos y, por otro lado, los desacuerdos entre superiores epistémicos e inferiores epistémicos. Sobre la base de lo que hemos planteado hasta aquí, consideramos que es menester reformular la hipótesis que debe someterse a discusión. De acuerdo con esta nueva hipótesis, que denominaremos “hipótesis del desacuerdo entre pares epistémicos reconocidos”, el desafío epistémico generado por las influencias no pertinentes se reduce al suscitado por el desacuerdo *entre pares epistémicos reconocidos*. A continuación, presentaremos argumentos para objetar esta hipótesis.

3. Desacuerdos entre pares epistémicos reconocidos

Como señalamos en el apartado anterior, la literatura predominante acerca de los desacuerdos (Kelly, 2005; Feldman, 2006; Christensen, 2007, 2009; Elgin, 2010) se ha concentrado sobre un tipo específico de ellos: a saber, los desacuerdos entre pares epistémicos reconocidos. King (2012, pp. 252-253) realiza una detallada revisión de esta literatura, y concluye que la mayoría de los autores juzga necesario el cumplimiento de las siguientes cuatro condiciones para que una disputa pueda ser identificada como un desacuerdo entre pares epistémicos reconocidos:

1. “Condición de desacuerdo: S cree que p , en tanto que T cree que $\neg p$ ”. Unas pocas aclaraciones básicas son necesarias en relación con esta condición: en primer lugar, el desacuerdo en cuestión debe ser genuino; es decir, las actitudes doxásticas de los sujetos involucrados deben dirigirse a la misma proposición. En segundo lugar, la mención de sólo dos sujetos, S y T, responde meramente a una simplificación expositiva; es decir, podría haber más sujetos involucrados en el desacuerdo. En tercer lugar, recurrir a un marco de actitudes plenas, en vez de apelar a uno de actitudes graduadas, responde asimismo a consideraciones de simplificación expositiva (para el desarrollo de nuestra argumentación, es indistinto qué tipo de marco se adopte).

2. “Condición de misma evidencia: S y T tienen la misma evidencia, E, pertinente en relación con p ”. Cuán viable sea la satisfacción de esta condición dependerá, obviamente, de qué concepción se tenga acerca de la evidencia. De esta manera, si se adopta una perspectiva “estrecha” acerca de la evidencia, tal como alguna variante de concepción dialéctica (de acuerdo con la cual sólo se concibe como evidencia aquello que puede ser articulado discursivamente y, consecuentemente, compartido con otros, de modo tal que no cuentan como evidencia las experiencias sensoriales puramente subjetivas o las intuiciones), la satisfacción de este requisito resultaría relativamente sencilla.

3. “Condición disposicional: S y T tienen igual disposición a responder a E de una manera epistémicamente apropiada”. Distintos autores consideran disímiles virtudes epistémicas en relación con este punto: entre otras, mencionan la inteligencia general, la honestidad intelectual,

tual, la pericia lógica, la ausencia de sesgos... A pesar de su variedad, puede considerarse que aquello a lo que remiten en último término estas virtudes epistémicas es a la confiabilidad de la persona en lo que respecta a la evaluación de la evidencia.

4. “Condición de reconocimiento: S y T tienen buenas razones para pensar que las condiciones 1, 2 y 3 se satisfacen”. Como puede apreciarse, el rasgo que distingue a esta condición de las tres anteriores es que aquéllas son condiciones de primer orden, mientras que ésta es de segundo orden. Exploraremos la importancia de esta condición en la quinta sección del presente trabajo.

King (2012) argumenta que, aun reduciendo el umbral de satisfacción de cada una de estas condiciones al mínimo, su cumplimiento sólo se logra en raras ocasiones. Consecuentemente, asevera, los genuinos desacuerdos entre pares epistémicos reconocidos son muy poco frecuentes. Independientemente de cuán acertada sea la tesis defendida por este autor, la idea que pretendemos defender aquí es la siguiente: las influencias no pertinentes *necesariamente* introducen disparidad epistémica puesto que impiden la satisfacción de la condición disposicional. De esta manera, cualquier desacuerdo que eventualmente ocurriera entre un sujeto que se encontrara bajo el influjo de una influencia no pertinente y otro que no lo estuviera no consistiría en un desacuerdo entre *pares epistémicos* reconocidos. Consecuentemente, debe rechazarse la hipótesis del desacuerdo entre pares epistémicos reconocidos. A continuación, desarrollaremos esta idea mediante la consideración de un caso paradigmático de autoengaño² directo³.

4. Autoengaño, influencias no pertinentes, y desacuerdo

Consideremos el siguiente ejemplo paradigmático de autoengaño directo, descrito por Fernández Acevedo (2018, p. 148):

² En el presente trabajo, presupondremos una conceptualización deflacionista acerca del autoengaño. Específicamente, seguiremos la propuesta elaborada por Mele (2001). Un aspecto nodal del deflacionismo consiste en rechazar la tesis según la cual es apropiado modelar los casos típicos de autoengaño sobre la base del engaño interpersonal. Por consiguiente: 1. los deflacionistas niegan la idea de que el autoengaño sea un proceso intencional: por el contrario, consideran que la conjunción de diversos sesgos cognitivos y factores motivacionales es suficiente para explicar satisfactoriamente este fenómeno; 2. los deflacionistas rechazan la idea de que el autoengaño implique la coexistencia de creencias contradictorias: en cambio, aseveran que para que exista autoengaño sólo se requiere que la persona adopte o mantenga una creencia en presencia de evidencia contraria a ésta (la persona, según sostienen, no posee ni adquiere, en ningún momento, la creencia que es apoyada por dicha evidencia).

³ La cualificación “directo” se emplea para aludir a casos de autoengaño en los cuales la creencia autoengañosa refiere a aquello que es deseado por la persona (es decir, la persona desea que p , y adopta o mantiene la creencia autoengañosa de que p). Los casos de autoengaño directo se diferencian de los de autoengaño “retorcido” (que han sido discutidos en la literatura con menor frecuencia); el aspecto distintivo de esta última índole de casos radica en que la persona desea que p , pero adquiere o mantiene la creencia autoengañosa de que $\neg p$.

W es un médico competente y bien formado, que ha estado experimentando desde hace bastante tiempo síntomas cada vez más claramente compatibles con la enfermedad de Huntington [...] Por otra parte, posee información proveniente de sus familiares que sostiene que su padre, con quien ha perdido contacto muchos años atrás, sufrió la enfermedad, lo cual, en caso de ser cierto, implicaría que la probabilidad de que él la padeciera fuese ciertamente elevada. W, no obstante, niega vehementemente ante sus íntimos la posibilidad de padecer ese mal. Desestima la información proveniente de sus familiares y conjetura explicaciones que, si bien no completamente inverosímiles, son claramente menos plausibles que la alternativa más indeseable.

Sea p la proposición “W no padece la enfermedad de Huntington”. Es el caso que $\neg p$; y la evidencia pertinente (que es poseída por W) provee mayor apoyo para $\neg p$ que para p ⁴. Ahora bien, dado que W desea que p , y considerando que su deseo condiciona la manera en que evalúa la evidencia pertinente que posee, W adquiere la creencia autoengañosa de que p . Así, el deseo en cuestión constituye una influencia no pertinente sobre la creencia en juego. Esto puede apreciarse retomando cualquiera de las conceptualizaciones acerca de las influencias no pertinentes que describimos anteriormente (y que, como señalamos, se superponen en buena medida). Así, siguiendo la propuesta elaborada por Vavova (2018), puede notarse que: (a) por un lado, el deseo de W de que p influye sobre el hecho de que W adquiera la creencia de que p (debido, como señalamos, a que dicho deseo condiciona la manera en que W evalúa la evidencia pertinente con la que cuenta; por ejemplo, ocasiona que W malinterprete negativamente⁵ la evidencia relativa a los síntomas que experimenta, de manera tal que W juzga que dicha evidencia no cuenta en contra de p). Y (b) por otro lado, el deseo de W de que p no es relevante en relación con la verdad de p . Por su parte, retomando la conceptualización sobre las influencias no pertinentes elaborada por Mogensen (2015), puede notarse que: (a) en caso de que W supiera que desea que p , eso no le proveería razón alguna a W para creer que p ; y (b) el deseo de W de que p contribuye a la explicación del hecho de que W crea que p (debido a que, como acabamos de señalar, el deseo de W de que p condiciona la manera en que este sujeto evalúa la evidencia que posee, ocasionando que, por caso, la malinterprete negativamente, y de esa manera dicho deseo contribuye causalmente con que W adquiera la creencia de que p), pero dicho deseo claramente no forma parte de las razones sobre las cuales W basa su creencia de que p . Finalmente, retomando la propuesta de Avnur y Scott-Kakures (2015), puede aseverarse que el deseo en juego es una influencia no pertinente de índole direccional puesto que éste, como ya hemos señalado, incide sobre el modo en que W evalúa la evidencia con la que cuenta, de manera tal que se favorece un resultado particular (precisa-

⁴ Cabe aclarar que, en principio, los casos paradigmáticos de autoengaño no son permisivos (Schoenfield, 2014): se trata de casos en los que hay una única manera epistémicamente racional de responder a la evidencia pertinente. Esta idea se encuentra presupuesta en la gran mayoría de las discusiones acerca del autoengaño.

⁵ La expresión “malinterpretación negativa de la evidencia” debe entenderse aquí en el sentido técnico propuesto por Mele (2001, p. 26).

mente, la adquisición o mantenimiento de la creencia de que p). No se trata de que el deseo en cuestión meramente coloque a W en una posición tal que pueda acceder a cierta evidencia; por el contrario, dicho deseo incide sobre la manera en que W maneja la evidencia en juego.

Ampliamos ahora este ejemplo paradigmático de autoengaño directo, ideando un escenario hipotético en el que W dialogara detalladamente sobre su estado de salud con alguna otra persona. Supongamos que, debido a la insistencia de sus preocupados allegados, W realiza una consulta con el médico X . Este otro médico es tan competente y se encuentra tan bien formado como W . Por otro lado, a diferencia de W , X no tiene ningún deseo relativo al estado de salud de W ; específicamente, no desea que p ni desea que $\neg p$. Por último, W le ofrece a X una descripción minuciosa de los síntomas que experimenta; además, le muestra videos grabados por sus allegados, en los cuales pueden observarse con claridad algunos de esos síntomas (por ejemplo: movimientos faciales anómalos, marcha inestable, movimientos espasmódicos en distintas zonas del cuerpo, deterioro del habla, comportamientos antisociales, desorientación). Asimismo, le comenta que sus familiares sostienen que su padre padeció la enfermedad de Huntington. Analicemos esta situación hipotética considerando las cuatro condiciones de los desacuerdos entre pares epistémicos reconocidos enumeradas por King (2012):

1. Condición de desacuerdo: como vimos, W cree que p . En cambio, X creería que $\neg p$, precisamente en virtud de que no se encuentra autoengañado (y en virtud de que de hecho es el caso que $\neg p$ y de que la evidencia pertinente provee mayor apoyo para $\neg p$ que para p).

2. Condición de misma evidencia: W y X poseen la misma evidencia pertinente en relación con p ; o, al menos, existe un solapamiento importante entre la evidencia pertinente en relación con p poseída por W y aquella poseída por X . Si adoptamos una concepción estrecha acerca de la evidencia, la satisfacción de esta condición resulta relativamente sencilla: W describe con minuciosidad sus síntomas a X , le muestra videos en los cuales varios de los síntomas relevantes son claramente visibles, y le brinda la misma información que le han proporcionado sus familiares acerca del hecho de que su padre ha padecido la enfermedad de Huntington. Otra cuestión que contribuye a que pueda lograrse la satisfacción de esta condición radica en que la evidencia en juego en los casos paradigmáticos de autoengaño no es excesivamente compleja ni extensa (en este caso, la evidencia pertinente es relativa a cierto conjunto de síntomas bien definidos y a la existencia de antecedentes familiares)⁶.

Consideramos que, para que sea legítimo sostener que una persona se encuentra autoengañada, es menester que se satisfaga el requisito de evidencia propuesto por Fernández Acevedo (2018, p. 149):

⁶ Esto se aprecia más nítidamente si se comparan los casos de autoengaño usualmente discutidos en la literatura sobre el tema (relativos al padecimiento de cierta enfermedad, a cualidades personales positivas, etc.) con los casos de desacuerdos entre (presuntos) pares epistémicos que suelen ser discutidos en la literatura sobre ese tópico (a modo de ejemplo, considérese el desacuerdo contemplado por Kelly -2005- en torno al rol que la intención de intimidar a la Unión Soviética pudo haber jugado en la decisión de H. Truman de bombardear Japón en 1945).

(a) el agente debe estar en posesión del cuerpo de datos pertinente para la evaluación de su creencia; (b) este cuerpo de datos provee mayor justificación para $\neg p$ que para p ; (c) el agente debe ser capaz de interpretar estos datos como pertinentes para la evaluación de su creencia.

Como puede advertirse a partir de la descripción del ejemplo de W, este sujeto satisface las tres condiciones que conforman el requisito de evidencia. La aceptación de dicho requisito parece necesaria para la satisfacción de la condición de misma evidencia, puesto que ésta requiere explícitamente que los agentes discordantes estén en posesión de la evidencia pertinente para el asunto sobre el que desacuerdan.

3. Condición disposicional: mientras que X tiene la disposición a responder a la evidencia pertinente en juego de una manera epistémicamente apropiada, W carece de dicha disposición. Por el contrario, W tiene la disposición a evaluar de manera irracional la evidencia en cuestión. Así, si bien X y W comparten algunas virtudes epistémicas importantes para una adecuada evaluación de la evidencia en juego (a saber, ambos son igualmente competentes como médicos y se encuentran igualmente bien formados), difieren en un aspecto crucial: W desea que p , en tanto que X es “neutral” al respecto (es decir, no desea que p ni desea que $\neg p$). Como vimos, el deseo de que p contribuye causalmente a que la evidencia en cuestión sea evaluada de manera sesgada. Consecuentemente, se anula la posibilidad de satisfacer la condición disposicional. Es importante destacar que esta anulación es “sustantiva”. Es decir, no se trata de que no se pueda cumplir esta condición debido, por caso, a un circunscrito error de ejecución en la evaluación de la evidencia por parte de uno de los sujetos. Por el contrario, el que un sujeto desee que p “(...) conlleva un riesgo no trivial de error, debido a una tendencia a formar actitudes o evaluar la evidencia [pertinente respecto de p] de una manera no confiable” (Ballantyne, 2015, p. 145).

La tercera condición que compone el requisito de evidencia propuesto por Fernández Acevedo (2018, p. 149), que reza que “(...) el agente debe ser capaz de interpretar estos datos como pertinentes para la evaluación de su creencia (...)”, se vincula con la condición disposicional. La persona autoengañada efectivamente debe poder interpretar la evidencia que posee como pertinente para la evaluación de p ; la persona no autoengañada que hipotetizamos también cumpliría esta condición. Ahora bien, precisamente lo que las distingue es que la persona autoengañada, a diferencia de la que no lo está, evalúa sesgadamente la evidencia en juego debido a que desea que p .

4. Condición de reconocimiento: es plausible que esta condición no se cumpla. Específicamente, X podría acceder a buenas razones para pensar que la tercera condición no se satisface. Desarrollaremos este punto en el próximo apartado.

En síntesis, parece razonable afirmar que, en casos como el analizado, la condición de desacuerdo y la de misma evidencia podrían satisfacerse. Por el contrario, el hecho de que uno de los sujetos involucrados en el desacuerdo se encuentre bajo una influencia no pertinente direccional tornaría inviable la satisfacción de la condición disposicional. De esta manera, los

sujetos involucrados en un desacuerdo de este tipo no podrían ser pares epistémicos. Consecuentemente, puede rechazarse la hipótesis del desacuerdo entre pares epistémicos reconocidos: el desafío epistémico generado por las influencias no pertinentes no puede reducirse al suscitado por el desacuerdo entre pares epistémicos reconocidos.

Esto ciertamente no significa que no existan nexos interesantes entre influencias no pertinentes y desacuerdos. Así pues, considérese, a modo de ejemplo, el “test de los pares cognitivos imparciales” ideado por Mele (2001): supongamos que *S* cree que *p* debido a que se encuentra autoengañado, que *E* es el conjunto de evidencia pertinente para evaluar el valor de verdad de *p*, y que *E* resulta fácilmente accesible para *S*. Si *E* se pusiera a inmediata disposición de los pares cognitivos imparciales de *S*, serían muchos más quienes creerían que *p* es falso que quienes creerían que *p* es verdadero. Mele indica que un requisito mínimo para que los pares en cuestión sean imparciales consiste en que ellos no desean que *p* y tampoco desean que $\neg p$. Efectivamente, la paridad que requiere el test es únicamente evidencial. En lo que a los aspectos disposicionales respecta, claramente no puede haber una paridad entre las personas involucradas: una de ellas se encuentra bajo el influjo de una influencia no pertinente (el deseo de que *p*) mientras que la otra es neutral al respecto. A pesar de que, entonces, el desacuerdo que pone de manifiesto este test no se geste entre pares sino entre personas con disímil jerarquía epistémica, el aspecto interesante del mismo radica en la función de indicador que cumple: en caso de que la mayoría de las personas sometidas al test discrepe con la persona que se toma como target, esto nos alertará acerca de la incidencia de una influencia no pertinente sobre dicha persona. En otras palabras, el test de los pares cognitivos imparciales no reviste relevancia epistémica en sí mismo; por el contrario, su importancia radica en que funciona como indicador del genuino problema epistémico en juego (i.e., el accionar de una influencia no pertinente).

5. Autoengaño y razonamientos desacreditadores

En el apartado previo, señalamos que en los casos de desacuerdos en los que uno de los agentes discordantes se encuentra autoengañado mientras que el otro no lo está, no sólo no se cumple la condición disposicional, sino que es posible que tampoco se cumpla la condición de reconocimiento. Ésta, como indicamos anteriormente, es una condición de segundo orden consistente en que los sujetos involucrados en el desacuerdo tengan buenas razones para pensar que se satisface la condición de desacuerdo, la de misma evidencia, y la disposicional. Es probable que el agente discordante no autoengañado no sólo no tenga buenas razones para pensar que la condición disposicional se satisface, sino que obtenga, por el contrario, buenas razones para pensar que dicha condición no se satisface. Exploraremos este punto a través de las contribuciones de Ballantyne (2015), y retomaremos el ejemplo introducido en el apartado previo para facilitar la exposición.

Ballantyne (2015) examina con detalle lo que denomina “razonamiento desacreditador” [*debunking reasoning*], cuya estructura básica es la siguiente: “Premisa del sesgo: una persona

está sesgada respecto de la proposición p . Conclusión desacreditadora: tenemos una razón para reducir la confianza en el informe de esa persona acerca de p ” (Ballantyne, 2015, p. 144). Una vez que se logra establecer la premisa del sesgo, la aceptación de la conclusión desacreditadora no presenta inconvenientes. Lo que sí resulta problemático es identificar buenas razones para aceptar la premisa del sesgo.

El vínculo entre el razonamiento desacreditador y la condición de reconocimiento es claro: si un agente se encuentra sesgado respecto de p , es incapaz de responder de manera epistémicamente apropiada a la evidencia pertinente respecto de p con la que cuenta. En otras palabras, dicho agente tiene la disposición a formar de manera irracional su actitud proposicional dirigida a p . De esta manera -y prosiguiendo con el ejemplo presentado anteriormente- si X obtuviera razones que le permitieran aceptar la premisa del sesgo respecto de W , no se satisfaría la condición de reconocimiento (específicamente, porque X tendría buenas razones para considerar que no se cumple la condición disposicional).

¿Mediante qué estrategias, entonces, podría obtener X buenas razones para aceptar la premisa del sesgo en el caso de W ? Una estrategia viable es aquella que Ballantyne (2015) denomina “estrategia de carencia de respuesta frente a evidencia convincente”. Para que X pudiera apelar a esta estrategia, debería tener buenas razones para aceptar (1) que la evidencia en juego (consistente en los datos relativos al conjunto de síntomas experimentados por W , y en la información provista por sus parientes acerca de que su padre ha padecido la enfermedad de Huntington) racionalmente obliga a quien la posea a adoptar cierta actitud respecto de p (en este caso, creer que $\neg p$); (2) que W posee dicha evidencia, pero no adopta la actitud hacia p que tal evidencia exige adoptar; y (3) que la mejor explicación para dar cuenta de por qué W no adopta dicha actitud consiste en que W evalúa de manera sesgada la evidencia en cuestión. Ballantyne está interesado específicamente en desacuerdos de gran complejidad, en relación con los cuales no parece razonable afirmar que la evidencia pertinente sea convincente [*compelling*]. En este sentido, Ballantyne juzga que el rango de aplicación de la estrategia de carencia de respuesta frente a evidencia convincente es muy estrecho, y consecuentemente descarta dicha estrategia. Ahora bien, como ya notamos al analizar la condición de misma evidencia, la evidencia pertinente en los casos paradigmáticos de autoengaño (y, por consiguiente, en los casos de desacuerdos en los que un agente está autoengañado y otro no lo está) no sólo es considerablemente menos extensa y compleja que la involucrada en, por ejemplo, las disputas filosóficas, sino que además puede considerársela como convincente. Por ello, en principio sería legítimo y factible que X apelara a la estrategia de carencia de respuesta frente a evidencia convincente para desacreditar la actitud de W respecto de p .

Otra estrategia viable es aquella que Ballantyne (2015) denomina “estrategia de atribución del factor sesgante” [*biasing factor*]. Para que X pudiera apelar a esta estrategia, debería tener buenas razones para aceptar (1) que cierto factor (por ejemplo, el deseo de que p) tiende a sesgar los juicios acerca de p ; (2) que dicho factor sesgante afectó de hecho la actitud que W

formó respecto de p ; y (3) que ningún factor “iluminador” [*enlightening*] (es decir, un factor que anule la influencia del factor sesgante) tuvo incidencia en la formación de la actitud de W respecto de p .

Por último, otra estrategia viable es aquella que Ballantyne (2015) llama “estrategia de auto-exculpación”. Para que X pudiera apelar a esta estrategia, debería tener buenas razones para aceptar (1) que él no se encuentra sesgado en lo relativo a p ; y (2) que el error de W se debe a un sesgo (y no a carecer de evidencia con la que X cuenta, o a no disponer de las mismas virtudes epistémicas generales -como las relativas a su competencia y formación-, o a haber cometido un error puntual de ejecución en la evaluación de la evidencia en juego).

Ballantyne (2015) juzga que estas dos últimas estrategias son más prometedoras que la mencionada en primer lugar, puesto que su rango de aplicación es mucho más amplio. Empero, no se encuentran exentas de problemas. Éstos están relacionados, específicamente, con la “ceguera al sesgo” [*bias blind spot*], que consiste en “(...) una tendencia amplia y ubicua de las personas a notar la existencia y operación de sesgos mucho más en otros que en sí mismos” (Pronin, 2007, p. 38). Pronin y Ballantyne ofrecen una extensa recapitulación de la abundante evidencia provista por diversas investigaciones psicológicas acerca de la ceguera al sesgo. Considerando esto, se advierte la falibilidad de las razones con las que X podría creer contar para afirmar que W evaluó sesgadamente la evidencia que poseía (punto 2 de la estrategia de atribución del factor sesgante), o para aseverar que ningún factor “iluminador” tuvo incidencia en la formación de la actitud de W respecto de p (punto 3 de la estrategia de atribución del factor sesgante), o para afirmar que él (X) no se encuentra sesgado en lo relativo a p (punto 1 de la estrategia de la auto-exculpación).

Si bien, entonces, es innegable que tanto la estrategia de atribución del factor sesgante como la de la auto-exculpación son falibles, puede sostenerse que de todas maneras es viable recurrir a ellas en casos relativamente simples, tal como el que estamos considerando a modo de ejemplo paradigmático. Así, en lo que respecta a la estrategia de atribución del factor sesgante, (1) X podría estar adecuadamente informado acerca de los modelos científicos (por caso: Friedrich, 1993; Kunda, 1990) que versan sobre la manera en que diversos factores motivacionales contribuyen a sesgar el juicio de un agente sobre cierto tópico (por caso, la incidencia de los deseos relativos al estado de salud sobre la evaluación de la evidencia pertinente relativa a ese asunto);

(2) X podría tener evidencia sólida según la cual W desea que p (obtenida a partir del comportamiento verbal y no verbal de W , y del testimonio de sus allegados). A esto podría sumarse la eliminación de explicaciones alternativas menos plausibles (ver segundo punto de la estrategia de auto-exculpación);

(3) X podría indagar si W cuenta con algún tipo de entrenamiento relativo a regulación de sesgos, y obtener una respuesta negativa.

Por su parte, en lo que respecta a la estrategia de la auto-exculpación,

(1) X podría disponer del testimonio confiable de sus colegas acerca de su gran capacidad de evaluar a sus pacientes de manera neutral y desapegada, a lo cual se añadiría la inexistencia de un vínculo personal entre él y W;

(2) en virtud de que la evidencia pertinente en juego es bastante acotada, y de que habría dialogado explícitamente sobre ella con W, X podría descartar que el error de W responde a que carece de cierta evidencia relevante; X también podría obtener evidencia adecuada sobre la formación y competencia de W (por ejemplo, averiguando en qué institución y época se formó, cuál es su desempeño actual como médico, etc.), de modo tal que X pudiera descartar que el error de W responde a que carece de ciertas virtudes epistémicas generales; finalmente, X podría descartar que el error de W responde a un error puntual de ejecución en la evaluación de la evidencia en juego examinando la evidencia junto a él, y asegurándose de que, por ejemplo, preste la debida atención. Eliminando estas explicaciones alternativas, X podría concluir que es muy plausible que W se encuentre sesgado respecto de p ⁷.

Ciertamente, las estrategias que analizamos son falibles, débase esto a que su rango de aplicación es demasiado estrecho y/o a la posibilidad de que su utilización se vea entorpecida por fenómenos como la ceguera al sesgo. Empero, y como Ballantyne (2015, p. 158) mismo observa, es razonable y adecuado recurrir a ellas cuando las situaciones en juego son relativamente simples. Esto se aplica, como esperamos haber mostrado, a casos de desacuerdos en los cuales una de las personas involucradas se encuentra autoengañada mientras que la otra persona no lo está.

6. Conclusión

En el presente trabajo hemos, en primer lugar, indicado por qué no se ha planteado de manera apropiada la discusión en torno a la relación entre el desafío epistémico suscitado por las influencias no pertinentes y el desafío epistémico suscitado por los desacuerdos. Los autores que se han ocupado hasta ahora de esta problemática no han identificado adecuadamente qué tipo de desacuerdos debe ser tenido en cuenta al discutir este tópico. Considerar los desacuerdos en términos generales (es decir, abarcar indiscriminadamente tanto desacuerdos entre pares epistémicos reconocidos como desacuerdos entre superiores e inferiores epistémicos) obtura la posibilidad de discutir adecuadamente el tópico que nos interesa. Como resultado de esta crítica, hemos formulado la hipótesis que consideramos que debe ser discutida: a saber, la hipótesis del desacuerdo entre pares epistémicos reconocidos. Posteriormente, hemos, por un lado, argumentado en contra de dicha hipótesis y hemos, por otro lado, explorado el rol que el razonamiento desacreditador podría desempeñar en los desacuerdos que ocurren

⁷ Lo sostenido aquí también se aplicaría al tercer punto de la estrategia de carencia de respuesta frente a evidencia convincente.

entre una persona que se encuentra sujeta a una influencia no pertinente direccional (como el deseo de que p involucrado en casos de autoengaño directo) y una persona que no se encuentra bajo una influencia tal. La relevancia del desafío epistémico suscitado por las influencias no pertinentes direccionales es innegable, dada la ubicuidad de estas influencias y la amenaza para la racionalidad que representan. Dicha relevancia no decrece por el hecho de que el desafío generado por este tipo de influencias no pueda reducirse al suscitado por los desacuerdos entre pares epistémicos reconocidos; por el contrario, esto sólo demanda que se reconozcan sus especificidades.

Dicha importancia tampoco decrece en vista de la relativa facilidad con la que puede emplearse el razonamiento desacreditador contra el juicio de quien se encuentra bajo cierto tipo de influencia no pertinente direccional, ya que esta facilidad es, precisamente, relativa a la complejidad propia del caso que se encuentre bajo consideración. Aquí hemos examinado a modo de ilustración la desacreditación de una persona autoengañada. Ahora bien, los casos paradigmáticos de autoengaño directo revisten -como hemos remarcado reiteradamente- un nivel de complejidad epistémica comparativamente bajo; el accionar de influencias no pertinentes es notoriamente más complejo e intrincado en muchos otros casos. Por esa razón, el optimismo que se desprende de nuestro examen no debe ser sobredimensionado.

Agradecimientos

Agradezco las observaciones y críticas realizadas por Gustavo Fernández Acevedo, Diana I. Pérez y Ornella Bruna, las cuales me permitieron mejorar el presente trabajo. Asimismo, agradezco al Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina) por haberme otorgado la beca doctoral en cuyo marco realicé este trabajo.

Referencias bibliográficas

- Avnur, Y. y Scott-Kakures, D. (2015). How irrelevant influences bias belief. *Philosophical Perspectives*, 29(1), 7-39. <https://www.jstor.org/stable/26614559>
- Ballantyne, N. (2015). Debunking biased thinkers (including ourselves). *Journal of the American Philosophical Association*, 1(1), 141-162. <https://doi.org/10.1017/apa.2014.17>
- Christensen, D. (2007). Epistemology of disagreement: the good news. *Philosophical Review*, 116, 187-217. <https://doi.org/10.1215/00318108-2006-035>
- Christensen, D. (2009). Disagreement as evidence: the epistemology of controversy. *Philosophy Compass*, 4(5), 756-767. <https://doi.org/10.1111/j.1747-9991.2009.00237.x>
- Elgin, C. (2010). Persistent disagreement. In R. Feldman & T. Warfield (Eds.), *Disagreement* (pp. 53-68). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199226078.001.0001>
- Feldman, R. (2006). Epistemological puzzles about disagreement. In S. Hetherington (Ed.), *Epistemology futures* (pp. 216-236). Oxford University Press.

- Fernández Acevedo, G. (2018). Autoengaño y evidencia. *Manuscrito*, 41(3), 125-161. <https://doi.org/10.1590/0100-6045.2018.V41N3.GA>
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: a reinterpretation of confirmation bias phenomena. *Psychological Review*, 100, 298-319. <https://doi.org/10.1037/0033-295X.100.2.298>
- Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne & T. Gendler (Eds.), *Oxford Studies in Epistemology (volume I)* (pp.167-196). Oxford University Press.
- King, N. L. (2012). Disagreement: what's the problem? Or: a good peer is hard to find. *Philosophy and Phenomenological Research*, 85(2), 249-272. <https://doi.org/10.1111/j.1933-1592.2010.00441.x>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research*, 88(2), 314-345. <https://doi.org/10.1111/phpr.12090>
- Mele, A. R. (2001). *Self-deception unmasked*. Princeton University Press. <https://doi.org/10.1515/9781400823970>
- Mogensen, A. L. (2015). Contingency anxiety and the epistemology of disagreement. *Pacific Philosophical Quarterly*, 97(4), 590-611. <https://doi.org/10.1111/papq.12099>
- Pronin, E. (2007). Perception and misperception of bias in human judgement. *Trends in Cognitive Science*, 11(1), 37-43. <https://doi.org/10.1016/j.tics.2006.11.001>
- Schoenfeld, M. (2014). Permission to believe: why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48(2), 193-218. <https://doi.org/10.1111/nous.12006>
- Vavova, K. (2018). Irrelevant influences. *Philosophy and Phenomenological Research*, 96(1), 134-152. <https://doi.org/10.1111/phpr.12297>