

## LA IRRACIONALIDAD Y EL AUTO-ENGAÑO

SILVIA BELLO  
Instituto de Investigaciones  
Filosóficas, UNAM

El auto-engaño es un tema sumamente amplio, los casos reconocidos como de auto-engaño son, generalmente, muy variados y complejos. Consideraré solamente un subconjunto de éstos: aquel formado por casos simples en los que una creencia que va en contra de otras creencias ya establecidas en la mente del sujeto es deseada por éste y así llega a ser formada. En estos casos encontramos un proceso psicológico dirigido a promover y mantener una creencia que va en contra de la evidencia que el sujeto posee; dicho proceso encierra la formación de, digamos, la creencia de que P, la cual es causada por el deseo de que P sea (o fuese) el caso y de que uno creyese que P.

No todo caso de auto-engaño, que envuelve la formación de una creencia, va en contra de otras creencias ya aceptadas por el sujeto; por otro lado, hay casos de auto-engaño cuya causa motivadora, por así decirlo, no es un deseo sino, por ejemplo, una emoción como los celos, o una fobia, etcétera. Discutiré solamente casos donde la formación de la creencia en cuestión es causada por un deseo y va en contra de otras creencias que el sujeto sostiene.

La paradoja más familiar del auto-engaño es que éste requiere que una sola persona juegue dos papeles incompatibles: como engañador, la persona debe dejar de creer aquello que, como engañado, debe creer. ¿Cómo puede alguien convencerse a sí mismo de que P y aun así mantener su creencia original de que no-P? Hay una dificultad para entender cómo es que un hombre racional puede auto-engañarse; el aspecto más interesante es que efectivamente hay casos que reconocemos como de auto-engaño y debemos tratar de explicarlos ha-

ciendo inteligible la irracionalidad que ellos envuelven. Si queremos explicar cómo nos autoengañamos, tenemos que considerar cómo es que podemos formar una creencia que va en contra de nuestras propias creencias originales, nuestra propia evaluación de la evidencia, simplemente porque queremos tener esa creencia; tenemos que tratar de explicar las paradojas de irracionalidad que encierra el auto-engañó.

En este artículo presentaré los que parecen ser los principales problemas que alguien ha de afrontar si quiere explicar la formación de creencias irracionales motivadas. También introduciré una discusión en torno a las contribuciones de D. Davidson y D. F. Pears para una posible solución de estos problemas.<sup>1</sup>

## I

En su artículo "Paradoxes of Irrationality", D. Davidson señala que hay un conflicto entre la manera normal de describir y de explicar las actitudes proposicionales (al igual que otros estados y sucesos mentales) y la idea de que tales estados pueden ser irracionales. Las paradojas surgen porque estructuras que están sujetas a las demandas de racionalidad no cumplen con estas demandas completamente; el obstáculo para una explicación inteligible del proceso es la necesidad de coherencia, consistencia o racionalidad entre los elementos de una estructura compleja. Los elementos mentales envueltos en dicho proceso (la formación de una creencia irracional motivada) se encuentran relacionados en formas que se apartan de la manera natural de relacionarse de los estados mentales que forma la base de las explicaciones por medio de razones, las explicaciones características de los fenómenos descritos en términos mentales.

Una creencia se muestra como internamente irracional cuando es vista como *no* razonable en relación a lo que hay

<sup>1</sup> La frase 'conexiones lógicas', que aparece a lo largo del trabajo, debe tomarse en sentido débil, amplio, que incluya casos tales como inferencias inductivas, hacia la mejor explicación, probabilísticas, etc. Esto se aclarará a través del trabajo.

en la mente del creyente; por ejemplo, si la creencia va en contra de aquello que el creyente sostiene como verdadero. En este sentido, la irracionalidad interna crea problemas y paradojas, porque ella envuelve inconsistencia en el patrón de creencias (deseos, intenciones, etcétera) de la persona. La irracionalidad interna envuelve la ocurrencia de un fenómeno mental que no es visto como razonable a la luz de los otros estados y eventos mentales del sujeto; así pues, envuelve un fenómeno mental que carece de una explicación psicológica estándar.

Queremos explicar el fracaso deseado, dentro de una misma mente, de coherencia o consistencia en el patrón de creencias; el problema es que el tipo de explicaciones que se aplica a lo mental demanda un patrón coherente de creencias, deseos y otras actitudes a la luz de los cuales el fenómeno mental que ha de ser explicado es visto como razonable —desde un punto de vista del sujeto.

## II

La paradoja general de la formación de creencias irracionales motivadas (es decir, causadas por un deseo) es: ¿Cómo puede alguien formar una creencia que no está apoyada por la evidencia que se posee (apoyada por otras creencias) sino que está basada en un deseo? O peor aún, ¿cómo puede alguien formar una creencia sobre la base de un deseo, incluso cuando ésta va en contra de la evidencia que se posee?

Debemos diferenciar dos partes de esta paradoja: primero, ¿cómo puede un fenómeno mental, un deseo en este caso, causar otro fenómeno mental, una creencia, sin ser el primero una razón para el segundo? Y segundo, ¿cómo pueden algunos fenómenos mentales, algunas creencias, fallar en causar aquello para lo que ellos constituyen una razón suficiente?

Podemos identificar la paradoja más familiar del autoengaño con la forma más extrema de la paradoja general: ¿cómo puede alguien formar la creencia de que P cuando también sostiene la creencia de que no-P, simplemente por-

que quiere creer que P? Pero hay otras formas menos extremas de esta paradoja. Siguiendo a Pears,<sup>2</sup> podemos distinguir cuatro tipos de casos, según los elementos que encontramos en la mente del sujeto en relación a las premisas de su evaluación (potencial) de la evidencia: primero, el sujeto puede poseer evidencia balanceada entre P y no-P; segundo, él puede poseer evidencia inductiva en favor de no-P; tercero, la evidencia que él posee puede implicar lógicamente no-P; finalmente, el sujeto puede haber formado ya la creencia de que no-P. Si, dado cualquiera de estos casos, un hombre mentalmente sano forma la creencia de que P, siendo competente para no hacerlo, entonces nos encontramos con un caso paradójico; porque consistencia y relaciones de razón se rompen de una manera tal que necesita de una explicación.

Consideremos el siguiente ejemplo. Alguien está tratando de averiguar si su hijo ha dejado ya de fumar o aún no. Esta persona evalúa la evidencia que posee y concluye que, por todo lo que ha considerado, P ('mi hijo ha dejado ya de fumar') es tan plausible como no-P. Supongamos que él sostiene el principio de que cuando uno tiene tan buenas razones para pensar que algo es el caso como para pensar que no lo es, uno debe suspender creencia acerca del asunto en cuestión. Pero nuestro sujeto quisiera que su hijo ya hubiese dejado de fumar y que él tuviese la creencia de que su hijo ha dejado ya de fumar. Así pues, él forma la creencia deseada a pesar de que sus deseos no son razones para pensar que P es el caso.

Aquí el estado de suspensión de creencia es bloqueado por la producción de la creencia deseada. Muy a menudo, cuando nuestra evidencia es neutral o balanceada, aventuramos una adivinanza y, en general, nuestra elección de una opción sobre otra(s) es motivada; pero el sujeto de nuestro ejemplo sostiene que P es verdadera y, al hacerlo, está ignorando o está pasando por alto sus principios de formación (racional) de creencias. Por simple y fácil que parezca, el hecho de que nuestro ejemplo, y casos similares, envuelva la formación de una creencia que es causada por un deseo y que va en contra

<sup>2</sup> D.F. Pears, "The Goals and Strategies of Self-Deception" (inédito, 1982), p. 2.

de un principio que el sujeto sostiene, crea dificultades que hacen que el proceso parezca paradójico.

Aquí, el sujeto tiene razones que no causan aquello para lo que ellas son razones suficientes, a saber, el estado de suspensión de creencia. La suspensión de creencia debiera seguirse dado que el sujeto juzga que su evidencia está balanceada entre P y no-P y sostiene el principio de que cuando tiene evidencia balanceada debería de suspender la creencia sobre el asunto. Pero a estos estados y sucesos no les sigue el estado de suspensión de la creencia y no les sigue a causa del deseo que el sujeto tiene de tener la creencia de que P. Esto es irracional; encontramos una distorsión en las conexiones lógicas de los contenidos de los elementos mentales envueltos, pero la deseada falla para suspender creencia y la deseada formación de la creencia de que P en su lugar no encierran, ni llevan a, una contradicción lógica: la creencia de que P, o el juicio de que P es verdadera, no contradice el juicio del sujeto de que su evidencia está balanceada y que, dada su evidencia, él debería suspender la creencia.

Debemos notar lo siguiente: en los casos donde la evidencia que el sujeto posee está balanceada, una adivinanza, en sí, no presenta problema. Lo que crea paradojas aquí no es el adivinar, o el escoger una opción como nuestra adivinanza, sino el hecho de que el sujeto forma la creencia de que lo que él ha escogido como su adivinanza es en realidad lo que es el caso. Consideremos un ejemplo. Un amigo pide a otro que adivine qué número va a ganar en el próximo giro de la ruleta, y que apueste a ese número. Supongamos que a nuestro sujeto le gusta más el 7 que ningún otro número, piensa que el 7 es su número de la suerte, y así apuesta al 7 esperando que éste sea el número ganador. Esto no está mal, aun cuando el apostador tiene tantas razones para creer que el 7 será el número que salga en el próximo giro, como para pensar que saldrá cualquier otro número. Pero si el apostador forma la creencia de que es realmente el número 7 el que va a ganar (y sostiene esta creencia fuertemente) sólo porque él quiere, entonces surge un problema para explicar la formación de esta creencia: el apostador no tiene ninguna razón para creer

que su adivinanza es (o será) realmente el caso. El siguiente ejemplo puede ayudar a aclarar esto. Supongamos que alguien quiere ir a X lugar; en su camino hacia X se topa con una bifurcación de la calle; él tiene tan buenas razones para pensar que el camino de la derecha es el que lleva a X como para pensar que es el camino de la izquierda el que lleva a X; pero juzga que es mejor continuar el viaje que regresar a casa y decide aventurar una adivinanza. Nuestro sujeto elige seguir por el camino de la izquierda porque se ve más agradable. Ahora bien, desear que el camino de la izquierda lleve a X y actuar de acuerdo con su adivinanza está bien, pero una adivinanza sobre la que uno actúa no es una creencia; si el sujeto forma la creencia de que el camino de la izquierda es el que lleva a X, entonces la irracionalidad entra conjuntamente con la creencia y con ella emerge un problema para explicar la formación de esa creencia. El sujeto tiene una razón para aventurar una adivinanza y también tiene una razón para elegir el camino de la izquierda, pero no tiene ninguna razón para creer que lo que él elige como su adivinanza es efectivamente el caso: el deseo de que P sea el caso no es una razón para creer que P es el caso. Racionalmente, él debería creer que es tan plausible que el camino que eligió sea el que lleva a X como lo es el que sea el otro camino el que lleva a X. Naturalmente, él puede esperar que haya elegido correctamente, pero no tiene ninguna razón para creer que efectivamente su elección fue la correcta. Si forma la creencia de que es el camino de la izquierda el que lleva a X (y la sostiene fuertemente) sólo porque quiere, entonces nos encontramos con una dificultad: ¿cómo puede el deseo de que P sea verdadera causar la creencia de que P es verdadera?<sup>3</sup> El deseo es gratificado por la formación de la creencia, pero el deseo consigue su gratificación de una manera desviada, porque la creencia es causada por el deseo y no, por ejemplo, por una inferencia a partir de premisas que apoyen la verdad de P. El problema

<sup>3</sup> En realidad, el deseo en cuestión debe ser el deseo de tener la creencia de que P, o el deseo de que P fuese verdadera y que uno debiese de tener la creencia de que P.

es que el deseo no le da ningún apoyo racional a la creencia y así, cuando el deseo causa la formación de la creencia, lo hace de una manera que se desvía de la manera estándar en que los fenómenos mentales que ocurren en una misma mente se encuentran relacionados.

### III

El aspecto lógico o racional de las actitudes proposicionales, las conexiones lógicas entre los contenidos de varias actitudes proposicionales, lo que ellas causan, y lo que las causó, las caracterizan como los estados y sucesos que son. Así, la idea de una creencia irracional resulta paradójica. Los fenómenos mentales se explican por medio de racionalizaciones; la descripción psicológica de un estado (como una actitud proposicional) lo hace sujeto de explicaciones por medio de razones. Los fenómenos mentales son descritos en términos de contenidos y éstos pueden ser especificados por sus conexiones lógicas con muchos otros contenidos proposicionales. Cuál sea el contenido específico de un estado o de un suceso mental, depende de las relaciones lógicas que éste guarda con el contenido proposicional de muchos otros estados y sucesos mentales del sujeto. Como Davidson formuló el problema: “La existencia de explicaciones de razones (. . .) es un aspecto integrado a las intenciones (. . .) y a muchas otras actitudes y emociones (. . .). Un aura de racionalidad, de ajustarse dentro de un patrón racional, es pues inseparable de estos fenómenos, al menos en tanto que ellos sean descritos en términos psicológicos.”<sup>4</sup>

No podemos atribuir una única creencia independientemente de muchos otros estados y sucesos mentales; la atribución de una creencia conlleva la adscripción de muchas otras actitudes proposicionales cuyos contenidos guardan, entre sí y con la creencia atribuida, conexiones lógicas apropiadas. Alguien cree que debiera llamar al mecánico para arreglar su

<sup>4</sup> D. Davidson, “Paradoxes of Irrationality”, en Ernest Jones Memorial Lectures: reprinted in R.A. Wolheim (ed.), 2 ed., Cambridge, 1982, p. 289.

carro porque cree que un mecánico puede arreglar carros, cree que tiene un carro y que su carro está descompuesto, quiere que su carro sea compuesto, etcétera. Puede haber otros conjuntos de creencias y deseos engarzados con base en los cuales alguien puede llegar a creer que debiera llamar a un mecánico para arreglar su carro, pero la atribución de esta creencia tendrá que hacerse junto con la atribución de muchas otras creencias y otras actitudes a la luz de las cuales la primera se torna inteligible.

El rompimiento de las relaciones de razón, entre elementos mentales que se encuentran relacionados causalmente, presenta un problema que resulta paradójico, porque la ocurrencia de relaciones de razón adecuadas, entre elementos mentales relacionados causalmente, está impuesta sobre los trabajos de la mente por nuestra manera de describir y de explicar estos estados y sucesos. La descripción mental de un estado o un suceso lo hace un candidato para ser una razón; las razones guardan relaciones lógicas entre sí y están relacionadas lógicamente con lo que ellas explican; la descripción mental de un estado se sigue lógicamente de sus razones.

Tenemos pues que explicar ambas irregularidades: a) ¿Cómo es que una razón puede fallar en causar aquello para lo que ella es una razón suficiente? y b) ¿Cómo es que un elemento mental puede causar otro elemento mental con el cual no guarda ninguna relación lógica apropiada y, sin embargo, siendo su causa, lo explica mas no lo justifica (racionalmente)?

En los casos de la formación de creencias irracionales motivadas, la irracionalidad entra dos veces. Primero, entra cuando el deseo de tener la creencia de que P hace que el sujeto ignore o deje de lado algunas de sus creencias (y principios), a saber, aquellas que van en detrimento de la creencia deseada, puesto que este deseo *no* es una razón en contra de esas creencias. Segundo, la irracionalidad entra cuando ese mismo deseo hace que el sujeto crea que P, porque el deseo de tener una creencia no da apoyo racional a la creencia en ninguna forma, no podemos pasar del contenido del deseo al contenido de la creencia de ninguna manera lógicamente apropiada.

#### IV

Davidson ha señalado una manera de reconciliar una explicación, que muestra que una creencia es irracional, con el elemento de racionalidad inherente en la descripción de las actitudes proposicionales. El elemento de racionalidad lo encontramos en el contenido de estos estados y los contenidos proposicionales tienen, entre sí, los tipos de relaciones lógicas “de los que depende la razón o su fracaso”.<sup>5</sup>

El contenido proposicional de ambos, la causa y el efecto, es central para una explicación de la formación de creencias irracionales motivadas. La creencia se explica por medio del deseo de tener esa creencia; por medio de una causa que, teniendo propiedades lógicas y siendo así un candidato para una razón, no es una razón para lo que ella explica en esta ocasión. La irracionalidad es una falla en la racionalidad. Dado que las propiedades lógicas de la causa no guardan una relación lógica apropiada con las propiedades lógicas del efecto, y así no muestran a la primera como una razón para el segundo, pero muestran la pertinencia de relaciones de razón aquí, el segundo aparece como irracional. Sin embargo, el contenido del deseo explica por qué el sujeto llegó a creer que P y no que Q o cualquier otra creencia. Hay una relación explicativa entre el hecho de que el contenido del deseo es ‘tener la creencia de que P’ y tanto: a) el hecho de que el sujeto forma la creencia de que P y así gratifica su deseo (esto puede expresarse diciendo que el sujeto tiene un motivo para formar la creencia, pero que esto no es una razón para pensar que P es verdadera), como b) el hecho de que el sujeto ignora justamente aquellas creencias y principios que van en detrimento de la creencia de que P, esto es, justamente aquellas creencias que amenazan la posible gratificación del deseo en cuestión. (Esto último puede expresarse diciendo que el sujeto tiene un motivo para ignorar, o dejar de lado, sus creencias, a saber, él quiere creer que P; pero que este motivo no fue una razón en contra de ellas mismas y así, fue irrelevante como una razón para su no-intervención.)

<sup>5</sup> *Op. cit.*, p. 299.

Ahora bien, la causa mental, el deseo de tener la creencia de que P, tiene ella misma una explicación normal de razones, e.g., el sujeto quiere tener esa creencia porque el tenerla le proporcionará placer, o *comfort* (o por alguna otra razón). La gente busca alcanzar la verdad en la formación de sus creencias; cuando este objetivo no juega su papel normativo usual en la formación de creencias, sino que es ignorado *con ese propósito*, se necesita un motivo para buscar tener la creencia formada de esa manera anormal, un motivo para que el sujeto ignore sus creencias relevantes, sus principios de formación de creencias y su objetivo de alcanzar la verdad; un obvio candidato, al menos en muchos casos, es el deseo de obtener placer, o *comfort*, mediante la gratificación de un deseo. Por otro lado, la creencia de que P, una vez formada, seguramente será una razón sobre la que el sujeto realizará algunas acciones intencionales y formará otros estados mentales. Sin embargo, la relación de razones entre el deseo y la creencia está rota y, por tanto, aún queda el problema de explicar la formación de la creencia irracional.

Como ya señalé, las explicaciones psicológicas normales explican por medio de racionalizaciones; una causa mental ayuda a explicar otro fenómeno mental al mostrar que la primera es una razón para el segundo —y la causa del segundo. El principio de relaciones lógicas, o de razones, apropiadas, que ocurren entre los contenidos proposicionales de los fenómenos mentales que están relacionados causalmente, es impuesto sobre lo mental por nuestra manera más básica de describir y de explicar estos fenómenos. Las explicaciones de razones (aplicadas a la formación de creencias) demandan, entre otras cosas, (i) que el contenido proposicional de varias creencias y otras actitudes guarden relaciones lógicas apropiadas entre sí y con el contenido de las creencias que ellos ayudan a explicar, y (ii) que los estados actuales de creencia, etcétera, causen la creencia que ha de ser explicada.<sup>6</sup>

Así, las explicaciones de razón requieren la ocurrencia de una causa mental, una causa con un contenido proposicional

<sup>6</sup> *Op. cit.*, p. 298.

que la haga un candidato para una razón, y también requieren que la causa sea una razón para el estado o suceso mental que ella causa. El patrón de explicaciones psicológicas pide, de los trabajos funcionales de la mente, que los fenómenos mentales interactúen causalmente de acuerdo con el requerimiento de relaciones lógicas apropiadas ocurriendo entre sus respectivos contenidos proposicionales. De aquí que el hecho de que algunas de las creencias del sujeto no produzcan el estado mental correcto, de acuerdo con los trabajos funcionales bien integrados de la mente, ni tampoco frenen la formación de una creencia que va en contra de ellas, junto con el hecho de que esta falla funcional es causada por el deseo de tener dicha creencia y de que este deseo causa la creencia en cuestión, presentan un problema para la explicación de la formación de creencias irracionales motivadas.

En los casos que estamos considerando, hay una causa mental con un contenido proposicional que explica la formación de la creencia, pero la relación lógica entre los contenidos de la causa y el efecto está distorsionada o se ha perdido; la causa mental no es una razón para el estado mental que ella causa. Por otro lado, hay razones que no causan aquello para lo que son razones suficientes, a saber, impedir la formación de la creencia que va en contra de ellas y suspender la creencia (en el primer tipo de casos), o la formación de la creencia opuesta (en el segundo y el tercer tipo de casos), o la conjunción de dos o más creencias (en el cuarto tipo de casos). Aún más, esto último es causado por un deseo y ese deseo no racionaliza la falla.

Ahora, para explicar el elemento de irracionalidad presente en todos estos casos, debemos describir la causa en términos mentales; pero si nos aferramos a la descripción mental de la causa, entonces nos apartamos del único patrón de explicación claro que se aplica a lo mental, porque este patrón requiere que la causa sea más que un candidato para una razón; la causa debe ser una razón apropiada.<sup>7</sup>

¿Cómo podemos explicar la relación causal no-racional que

<sup>7</sup> *Op. cit.*, p. 299.

ocurre entre el deseo de tener la creencia de que P y: (i) la creencia de que P, y (ii) la falla funcional de aquellas creencias y principios que van en detrimento de la creencia de que P?

La causa debe guardar su carácter de causa mental, debe encajar dentro de un sistema de creencias, deseos, emociones, etcétera, engarzadas entre sí; y así, dentro de un patrón racional. El efecto, aunque se ha mostrado que es irracional, debe también encajar en un sistema de creencias y deseos engarzados que lo haga inteligible.

Necesitamos una explicación unificada de los siguientes puntos: (i) Cómo y por qué las creencias relevantes para la ocasión, e.g. creencias acerca de la evidencia, fallan en producir el estado mental correcto de acuerdo con el trabajo funcional de un sistema mental bien integrado. En el primer caso, suspensión de la creencia; en el segundo y el tercero, la creencia de que no-P; en el cuarto caso encontramos una pregunta más difícil: ¿cómo puede el sujeto creer que P y creer que no-P sin sostener como verdadera una contradicción?<sup>8</sup> (ii) En general, cómo y por qué las creencias originales fallan en impedir la formación de una creencia que va en contra de ellas. Finalmente, (iii) cómo es que un deseo burla a la razón y causa la creencia de que P.

## V

Hemos notado que las creencias, y los principios, relevantes no intervienen en los trabajos funcionales de la mente cuando debieran hacerlo, ellas fallan en impedir la formación de una creencia que se les opone y en producir el estado de creencia correcto. En casos del primer tipo, la formación de la creencia de que P va en contra del principio de que uno debe suspender la creencia cuando la evidencia está balanceada y del hecho de que la evidencia que el sujeto posee está

<sup>8</sup> En casos del tercer tipo la pregunta es: ¿Cómo puede el sujeto creer que Q, creer que Q implica no-P, ser capaz de, y competente para, inferir no-P, y aun así fallar en inferir no-P? Nótese que en casos donde el sujeto no es una persona mentalmente sana, la explicación se dará en términos de su locura.

balanceada entre P y no-P, lo cual, conjuntamente, debiera llevarlo a suspender la creencia. Sin embargo, la formación de la creencia de que P no hace que nos encontremos con un peligro de contradicción lógica aquí. Similarmente, los casos del segundo tipo envuelven irracionalidad pero no hay ningún peligro de contradicción lógica en ellos. Aquí, las creencias relevantes, originales del sujeto, e.g. creencias acerca de la evidencia, apoyan la plausibilidad de no-P sobre P; pero la evidencia que el sujeto posee no es concluyente: ésta no excluye la posibilidad de que P sea verdadera, aunque sí va en detrimento de su verdad.

Por ejemplo, supongamos que el hombre que deseaba que su hijo ya hubiese dejado de fumar encuentra unas colillas escondidas en el cuarto de su hijo. Ahora la evidencia que él posee señala inductivamente hacia no-P. Podemos suponer que el sujeto sostiene el principio de que debiese de creer aquello hacia lo que su evidencia señala, o que debe de sostener creencias sobre la base de su evidencia. Ahora él tiene una tendencia racional a creer que no-P; pero su evidencia, no siendo concluyente, le permite pensar que él ha considerado solamente un lado de los hechos, que puede haber evidencia relevante que él desconoce y que inclina la balanza hacia P. Quizá las colillas pertenecen a uno de los amigos de su hijo, quizá son ya muy viejas. El sujeto puede darse cuenta de que cree que P sólo porque quiere; que dada su evidencia, debería creer que no-P, y puede decir que sin embargo sostendrá que P hasta que descubra que P realmente no puede ser verdadera. El puede pensar que, al final, encontrará que la evidencia apoya a P y empezar a buscar evidencia de una manera tendenciosa. Nuevamente, el sujeto tiene razones que no causan aquello para lo que ellas son razones suficientes; la creencia de que no-P debiera ser producida dado que el sujeto juzga que su evidencia señala inductivamente hacia no-P y sostiene el principio de que debería de tener creencias sobre la base de su evidencia; por otro lado, su evidencia debiera bastar para que él rechazara la creencia de que P dado su principio de formación de creencias. La irracionalidad interna aparece cuando el sujeto no forma una creencia sobre la base de su eviden-

cia sino que sostiene la creencia opuesta sobre la base de su deseo de tener esa última creencia; porque cuando él hace esto, él va en contra de su principio y deja de lado ese principio no porque tenga una razón en contra del principio mismo, o en contra de su evidencia, sino porque desea tener la creencia de que P. Así, por un lado encontramos que las creencias y los principios relevantes fallan en funcionar como una causa de aquello para lo que ellas constituyen razones; y por el otro, encontramos que las relaciones de razón se rompen cuando el deseo de tener la creencia de que P causa tanto la falla funcional de las creencias relevantes —falla en su intervención en los trabajos de la mente como ellas debieran—, como la producción de la creencia de que P.

En los casos del tercero y del cuarto tipo, las creencias relevantes del sujeto acerca de la evidencia no dejan latitud alguna para la posibilidad de que P sea verdadera, pues ellas descartan la verdad de P implicando lógicamente su falsedad. Así, de estos casos sí emerge un peligro de contradicción lógica. En ellos el sujeto no puede sostener que Q implica no-P, que Q es verdadera y que, sin embargo, P es verdadera; o bien sostener que P y no-P, a menos de que se contradiga a sí mismo. No consideraré casos de auto-contradicción, pero los casos que estamos considerando no nos llevan a atribuirle al sujeto creencias que son auto-contradictorias. En casos del tercer tipo, el sujeto sí cree que Q implica no-P, cree que Q y cree que P; en casos del cuarto tipo, el sujeto sí cree que no-P y sí cree que P; pero si él no sostiene una contradicción lógica como verdadera, entonces no debemos atribuirle la creencia conjuntiva de que no-P y P. ¿Cómo puede ser esto posible? ¿Cómo puede alguien sostener dos creencias mutuamente contradictorias sin sostener una contradicción al mismo tiempo? Pears sugiere lo siguiente: si un hombre sostiene creencias mutuamente contradictorias, pero no cree la conjunción de ese conjunto de creencias, entonces las creencias deben de estar separadas en su mente, el sujeto no las conjunta porque ellas se conservan separadas. El hecho de que creencias mutuamente contradictorias puedan co-existir en una misma mente sin que sean conjuntadas, y sin interactuar entre ellas

de la manera que nosotros esperamos, nos lleva a pensar que la mente está dividida en esos casos, una división conteniendo la creencia de que P y la otra la creencia de que no-P entre sus elementos.<sup>9</sup>

Si asumimos una división de la mente, despejamos el camino para acomodar de una manera inteligible todos los elementos envueltos en los casos bajo consideración.

Consideremos el siguiente ejemplo: Alguien que sabe que 8 multiplicado por 7 da 56 quiere saber cuántos cigarros se fumó en la última semana. Él sostiene que ha fumado 8 cigarros cada día, así, multiplica 8 por 7 y, equivocadamente, concluye que ha fumado 48 cigarros en la última semana. Supongamos que el error que el sujeto ha cometido fue causado por su deseo de haber fumado menos de 50 cigarros. 'Haber fumado 8 cigarros cada día en la última semana' (Q), implica lógicamente 'haber fumado más de 50 cigarros en la última semana'. Si el sujeto de nuestro ejemplo se diese cuenta del error que cometió y que 'haber fumado 56 cigarros' se sigue de sus premisas, él abandonaría su creencia de que fumó menos de 50 cigarros. Aquí el sujeto, siendo un hombre racional, no puede decir: 'mi evidencia implica que fumé 56 cigarros, pero yo creo que fumé 48 sólo porque quiero creer que fumé menos de 50 cigarros'. El sujeto no puede conjuntar sus creencias mutuamente contradictorias y continuar creyendo que P mientras que está consciente de su irracionalidad, a diferencia de los casos donde sí hay latitud: casos de los dos primeros tipos.

Ahora bien, si asumimos que la mente de este sujeto está dividida (temporalmente), que en su mente hay un subconjunto de atributos mentales donde la creencia de que 8 multiplicado por 7 da 56 se encuentra (momentáneamente) aislada del sistema mental donde la creencia de que P (la creencia de que el sujeto ha fumado menos de 50 cigarros) se produce, entonces podemos dar cuenta de la falla que el sujeto comete en los siguientes términos: La primera creencia no jugó el papel causal que debería haber jugado porque,

<sup>9</sup> D.F. Pears, *op. cit.*

debido al deseo del sujeto, ésta se conservó sin interactuar, y sin ser conjuntada, con ciertos elementos mentales del sistema donde la conclusión equivocada se produjo. Una división jugaría el papel del engañador y la otra el papel del engañado: la creencia de que Q implica no-P (o la creencia de que no-P) se conserva fuera del segundo y la creencia de que P es producida en éste por medio de la operación funcional del primero.

Hay muchas preguntas importantes acerca de la identidad, los elementos, las operaciones y otras características de las divisiones; en particular, necesitamos un criterio adecuado para dividir sistemas mentales dentro de una misma mente. Pears ha propuesto tomar como criterio la no-intervención, o el aislamiento, funcional de algunos elementos mentales. Como él señala, podemos distinguir dos niveles de aislamiento funcional: uno débil, que encontramos en casos de los dos primeros tipos, donde todos los elementos pueden ser conjuntados aunque hay elementos que no interactúan para producir los estados que debiesen; y otro fuerte, que encontramos en casos de los dos últimos tipos, donde no sólo la interacción sino también la conjunción es descartada. Pears sugiere tomar el sentido fuerte como criterio de división. Antes de considerar esta propuesta más detenidamente, haré algunas aclaraciones y expondré la propuesta de Davidson para solucionar la relación anómala que encontramos entre el deseo de creer que P y la formación de la creencia de que P; la relación causal no-racional que se da entre estos dos elementos mentales.

## VI

Para establecer cuándo es que dos sistemas diferentes deben de estar en juego, y cómo puede ayudar el partir la mente para explicar el que el sujeto crea que P, debemos de establecer un criterio para marcar la línea divisoria entre sistemas mentales. Y, para establecer dicho criterio, debemos de considerar, al menos de una manera general y superficial, qué es lo que integra a varios fenómenos mentales dentro de

un mismo sistema mental, qué nos fuerza a segregar fenómenos mentales como elementos de diferentes sistemas pero dentro de una misma mente. También debemos tener en mente cuáles son los puntos que requieren de explicación: (i) el hecho de que algunas de las creencias del sujeto no producen el estado mental correcto de acuerdo con la operación funcional bien integrada de un sistema mental, ni tampoco bloquean la producción de una creencia que va en contra de ellas; y esto es un fracaso deseado, cometido con un propósito atrás; (ii) el hecho de que un deseo produce una creencia burlando a la razón y, al mismo tiempo, usualmente juega un papel causal en la falla mencionada anteriormente.

El objetivo es tratar de acomodar todos los elementos mentales envueltos de una manera tal que se preserve la racionalidad o consistencia dentro de un mismo sistema; al mismo tiempo, queremos dar una explicación inteligible del proceso que lleva a la formación de una creencia irracional motivada. Queremos que las creencias mutuamente contradictorias estén separadas; además, queremos darle sentido a la operación del deseo, el deseo no es una razón para lo que causa, entonces ¿cómo puede el deseo mismo y su operación guardar un lugar dentro de un sistema racional?

Lo que unifica varios fenómenos mentales en un mismo sistema, lo que constituye un sistema unificado de estados mentales, es el tipo de relaciones entre sí en que se encuentran estos fenómenos; el que ellos interactúen entre sí y se combinen para producir otro fenómeno (mental) de acuerdo con los principios que gobiernan el funcionamiento racional de un sistema mental. Cuando asignamos elementos a un sistema mental, tenemos que enfrentarnos a una restricción de racionalidad. Si alguien cree que su carro está descompuesto, entonces debe creer que tiene un carro y debe tener algunas otras creencias sobre carros. Si alguien cree que 4 más 4 suman 8, entonces debemos esperar que dé la respuesta "8" cuando se le pregunta cuánto suman 4 más 4, si asumimos que él quiere dar la respuesta correcta y que no tiene ninguna razón en contra de ello. Si alguien está preocupado porque no puede pagar la renta, él debe creer que no puede pagar la renta y

que si no paga ello puede acarrearle algún problema. Si alguien está descontento de que su suegra venga a visitarlo, él debe creer que no es deseable que su suegra venga a visitarlo. Si alguien cree que es mejor salir de vacaciones que quedarse en casa, esperamos que forme la intención de salir de vacaciones, a menos de que tenga alguna razón para no hacerlo. Si alguien forma la intención de ir a Acapulco por avión, debe tener el deseo de ir a Acapulco y la creencia de que se puede ir a Acapulco por avión. Si alguien toma un analgésico intencionalmente, debe tener el deseo de tomar el analgésico y la creencia de que lo que está tomando es un analgésico. En cada caso, es la conexión lógica entre los contenidos de varios elementos mentales y lo que ellos causan, su funcionar y operar de una manera racional, lo que los integra como elementos de un mismo sistema; y es la carencia de una conexión lógica adecuada, o la aparición de inconsistencias, entre los contenidos de ciertos elementos mentales, lo que crea un problema para integrar los elementos relacionados causalmente pero no-lógica o no-racionalmente, o los elementos que fallan en interactuar como la racionalidad demanda.

## VII

Davidson sugiere, primero, considerar casos de interacción social y entonces ver esas interacciones como una analogía de lo que sucede en una misma mente. Considerando las interacciones sociales, encontramos que las relaciones causales no-lógicas o no-rationales entre fenómenos mentales no presentan ningún problema (y no necesariamente ninguna irracionalidad) cuando la causa y el efecto ocurren en mentes diferentes.<sup>10</sup> Por ejemplo, mi desear que tú creas que voy a comprar un carro nuevo puede causarte el creer esto sin que mi desear tal cosa sea una razón para tu creencia de que voy a comprar un carro nuevo ni tampoco una razón con base en la cual tú llegas a creer eso. Aunque mi deseo es una razón para hacer todo lo que yo hago para hacerte creer que voy a com-

<sup>10</sup> D. Davidson, *op. cit.* p. 300.

prar un carro nuevo, y una razón con base en la cual yo lo hago. Puedo estar engañándote, o puedo no estarlo, el asunto es que aunque mi desear causa tu creencia sin ser una razón para ello, no parece haber ningún problema para explicar lo que sucede aquí. Porque mi deseo y tu creencia pertenecen a mentes diferentes e independientes: en cada mente encontramos una estructura de razones que apoyan lo que sucede, encontramos creencias, deseos, intenciones, etcétera, engarzados apropiadamente. Desde mi punto de vista, es racional para mí hacer lo que creo que puede ayudar a que tú creas que voy a comprar un carro nuevo; desde tu punto de vista, puede parecer que no hay ninguna razón en contra de que tú creas eso y que hay algunas razones para pensar que compraré un carro nuevo. Davidson sugiere que esta idea puede ser aplicada dentro de una misma mente asumiendo que la mente puede ser dividida en estructuras semi-independientes que interactúan de maneras que el principio de racionalidad no puede explicar. De acuerdo con Davidson, si vamos a ser capaces de explicar la relación anómala entre fenómenos mentales que ocurre en formas comunes de irracionalidad, entonces debemos asumir que la mente puede ser dividida en sub-sistemas semi-autónomos, dentro de cada uno de los cuales hay un grado de racionalidad más alto del que encontramos en el todo que constituye la vida mental del sujeto, y donde un sub-sistema puede operar sobre otro en la modalidad de “causación no-lógica o no-racional”. La idea es que si diferentes estructuras de la mente son hasta cierto punto independientes, entonces podemos entender cómo es que pueden existir en una misma mente creencias inconsistentes, o que están en conflicto, y cómo es que ellas pueden interactuar de maneras que conllevan irracionalidad. De esta forma podemos acomodar los elementos relacionados causalmente, aunque no racionalmente, de una manera que preserve la racionalidad relativa a cada sub-sistema. El principio de racionalidad gobierna las interacciones causales de los elementos dentro de un sistema; es sólo en las interacciones causales entre diferentes sistemas que “la razón no juega su papel normativo y racionalizador usual”. Para Davidson la causa, si es una causa mental que no es una razón, debe ser

localizada en un sistema diferente a aquel donde localizamos el efecto; de acuerdo con Davidson, para explicar la ocurrencia de causas mentales que no son razones para los estados mentales que causan, es necesario ver a la mente como teniendo dos o más estructuras semi-autónomas; sólo mediante la división de la mente es posible explicar cómo es que un fenómeno mental puede causar otro fenómeno mental con el cual no guarda una relación racional.<sup>11</sup>

La idea de Davidson parece ser que sólo segregando a la causa y al efecto en diferentes estructuras podemos dar una explicación de la relación anómala en la que se encuentran. Si la causa ha de guardar su carácter de causa mental, entonces debe ser localizada en un sistema de estados y sucesos mentales engarzados, diferente del sistema donde el efecto es producido. La causa, si ha de ser una causa mental, debe ajustarse dentro de un patrón racional y para ello debemos asumir que la creencia es producida en un sistema diferente y hasta cierto punto independiente. Nótese que estas consideraciones deberían aplicarse no sólo al deseo sino también a lo que éste hace para producir la creencia, a la operación misma del deseo. Aún más, es la operación del deseo la que debe quedar fuera del sistema donde la creencia irracional se produce y no el deseo mismo; el deseo es un elemento compartido por ambos sistemas, de no ser así no se entendería para qué se produce la creencia de que P en el sistema donde se produce y por qué su producción ocasiona placer.

Siguiendo a Davidson, encontramos que dos estructuras diferentes de estados mentales deben estar en juego en cada caso de una formación motivada de una creencia. Porque, de acuerdo con el criterio que propone Davidson, es el rompimiento de las relaciones de razón lo que define la línea de una subdivisión,<sup>12</sup> y esto ocurre siempre que el deseo de tener una creencia causa la formación de la creencia deseada y siempre que ese deseo causa la falla funcional de creencias ya formadas que van en detrimento de la creencia deseada. Los con-

<sup>11</sup> *Op. cit.*, p. 303.

<sup>12</sup> *Op. cit.*, p. 304.

ceptos operativos que definen las partes son los de causa y razón: cuando hay causas mentales que no son razones para lo que causan, o bien cuando hay razones que no causan aquello para lo que ellas son razones suficientes, debemos asumir una sub-división. Cualquier función que sea llevada a cabo por una estructura dada debe estar apoyada por elementos de esa estructura que constituyen razones para que esa función se lleve a cabo y los cuales causan el proceso en cuestión. Solamente de esa manera podemos dar sentido a la atribución de esa función a esa estructura.

El que creencias mutuamente contradictorias puedan coexistir dentro de una misma mente puede forzarnos a ver la mente como dividida. En estos casos la inconsistencia, o la irracionalidad, es demasiado grande como para conciliar a todos los elementos dentro de un sistema racional. Pero algo menos asombroso a simple vista que el peligro de auto-contradicción parece forzarnos a segregar elementos dentro de una misma mente. Hay algunos problemas para acomodar a todos los elementos mentales de una manera inteligible, pero al final podemos acomodarlos sin apartarnos mucho de la propuesta de Davidson para definir la línea divisoria y adscribir elementos a cada sub-estructura.

Ahora bien, parece discutible que sea sólo mediante el asumir que la mente está dividida que podamos explicar cómo es que las relaciones de razón pueden romperse entre elementos mentales que se encuentran causalmente relacionados, que sólo localizando la causa y su operación en un sistema diferente de aquel donde el efecto es producido sea posible conservar el carácter mental de la causa. También parece discutible que la única manera en la que podamos explicar la falla motivada de ciertos elementos mentales para producir aquello para lo que ellos constituyen una razón suficiente sea mediante el localizar esos elementos en una sub-división de la mente donde se conservan aislados de otros elementos.

Por supuesto que podemos definir la línea divisoria de sistemas mentales en términos del rompimiento de relaciones de razón; pero la cuestión es si una sub-división es realmente ne-

cesaria para explicar el rompimiento de las relaciones de razón en cada caso en el que este rompimiento ocurre.

La ruptura de relaciones lógicas apropiadas entre los contenidos de elementos mentales causalmente relacionados es más bien común y no siempre conlleva irracionalidad. Por ejemplo, simples casos de asociación, como mi recordar que debo escribir una carta causado por mi deseo de comer un pastel, que no cuentan como casos de irracionalidad. Pero, más importante, un motivo no está siempre detrás de cualquier asociación, tampoco hay siempre una explicación de tipo freudiano para cualquier caso simple de asociación. No parece que tengamos que asumir una sub-división de la mente para explicar estos casos. Davidson puede argumentar que si permitimos que gobierne una causalidad no-racional dentro de un mismo sistema, entonces aflojamos el principio de la ocurrencia de relaciones lógicas o racionales entre los contenidos de los fenómenos mentales y de esta manera arriesgamos la base para atribuir racionalidad y con ello cualquier estado o suceso mental. Si para que una relación causal no-racional entre sucesos mentales pueda ocurrir dentro de una misma mente se requiere que se desarrolle un sub-sistema con un tipo de estructura especial, entonces tenemos una restricción para asignar irracionalidad de manera demasiado amplia: para explicar un fenómeno irracional debemos adscribir un fondo de racionalidad, debemos atribuir una constelación de estados y sucesos mentales engarzados que se combinen de acuerdo con los principios de acción racional para causar el fenómeno que ha de ser explicado. Así, cuando queremos explicar un fenómeno mental que, aunque no sea irracional, es causado por otro fenómeno mental que no es una razón para él, debemos de satisfacer ese requerimiento.

Sin embargo, parece que hay dos diferentes tipos de casos simples de asociación: aquellos donde la asociación es motivada y aquellos que son más como casos de fallas intelectuales no-motivadas. Los primeros requieren de una explicación de tipo freudiano y parecen necesitar de una sub-división de la mente, pero los segundos no son el resultado de "fuerzas ajenas", pensamientos e impulsos alienados, sino que son el resultado

de extraños brincos en inferencias: mi desear comer un pastel me llevó a pensar en los deliciosos pasteles que hace fulana y así a recordar que debía escribirle una carta a fulana. En estos casos hay algo como un mal funcionamiento dentro de un solo sistema; algo como tomar (equivocadamente) como relevante algo que no lo es, o como brincar de una premisa a una conclusión que no se sigue del todo de esa premisa solamente. ¿Cuál sería el motivo de dividir la mente en estos casos? La asociación en estos casos parece ser el resultado de los trabajos de un único sistema, es como seguir un patrón de razonamiento que no está del todo bien pero donde todos los elementos parecen encajar debido a una mala apreciación no-motivada. Es central a la propuesta división de la mente que cada sistema sea concebido como una persona en muchos aspectos importantes, esto requiere de una asociación dinámica donde un lado opera sobre el otro, trabajando con base en los principios de la acción intencional, y causa en él un evento mental que, aunque es el objetivo del proceso, no tiene una razón como su causa. Pero una simple asociación no es siempre el objetivo de algún proceso mental; algunas veces la asociación se produce como se producen los errores intelectuales no-motivados. En dichos casos no hay elementos mentales suficientes para formar una sub-división del tipo requerido.

Si lo anterior es correcto, el criterio que Davidson propone debiera ser restringido a casos donde una motivación está en juego. La mente parece estar dividida en aquellos casos donde la irracionalidad es motivada.

## VIII

La propuesta de Pears es dividir la mente para separar elementos cuando la irracionalidad es demasiado grande como para acomodar a todos los elementos envueltos dentro de un mismo sistema: cuando hay un peligro de contradicción lógica, a saber, en los casos del tercero y del cuarto tipo. En los dos primeros tipos de casos puede haber una división, pero esto no es necesario realmente y la división aquí sería como un lujo del proceso. La no-intervención de elementos mentales,

entendida como no-conjunción y no-interacción, definiría la línea de una sub-división; la no-interacción de elementos, cuando su conjunción es permitida, podría ser explicada sin necesidad de asumir una división de la mente. Pears pensó que tomar la no-intervención en el sentido fuerte como criterio de división nos daba una restricción para no dividir la mente demasiado y que esta restricción se necesitaba para disipar el peligro de desintegrar la mente, a la persona y con ello perder toda posible atribución mental.

Davidson podría decir que no podemos dividir la mente demasiado, aunque los cuatro casos nos fueren a una división, por razones similares a las que nos llevan a asumir racionalidad, coherencia y veracidad en las creencias. Por otro lado, el criterio de Davidson funciona como una balanza: nos da una restricción para no adscribir irracionalidad a la ligera, porque para adscribir irracionalidad interna motivada hay que construir un sistema con cierta estructura, la adscripción debe hacerse bajo un fondo de racionalidad.

Por otro lado, si todos los elementos mentales envueltos en los dos primeros tipos de casos pueden estar juntos, como miembros de un mismo sistema, no es claro cómo es que la falla funcional se produce: los conceptos operativos dentro de un sistema son aquellos de causa y razón. Tampoco estaría claro cómo podríamos explicar la relación anómala que se da entre el deseo de tener la creencia de que P y ambos: la creencia de que P y la falla funcional de las creencias relevantes originales del sujeto. La explicación sobre la línea de Davidson fuerza a una división de la mente aun en los casos donde ninguna contradicción lógica está envuelta. Quizá haya otra explicación posible, pero no es claro cuál sea; tampoco está claro cómo responder al problema que parece surgir si se permite un funcionamiento regido por causas que no son razones, y donde una motivación está en juego, dentro de uno y el mismo sistema.

Sin embargo, podemos responder a estas cuestiones si aceptamos como criterio para dividir sistemas mentales el criterio débil que propone Pears y no el criterio fuerte.

Debemos notar que, según el criterio positivo que propone

Davidson, a saber, la ocurrencia de una causa que no es una razón para lo que causa, la mente se dividirá siempre que haya una causación irracional positiva. Mientras que según el criterio débil que propone Pears, a saber, la ausencia de causalidad —o de intervención— cuando hay razones suficientes, la mente se dividirá siempre que haya una falla negativa para intervenir y frenar la irracionalidad. (Ambos criterios deberán restringirse a casos motivados, casos donde encontramos suficientes elementos como para formar la estructura requerida.) Cuál de estos criterios sea preferible es una cuestión interesante e importante, pero dejaré esta discusión para un artículo posterior.

## SUMMARY

In this paper I have tried to present what seem to me the main problems to be confronted in explaining the formation of motivated, irrational beliefs, and the most important recent contributions to a solution of such problems.

First, I make some classifications and present some important features of the different cases of motivated irrational belief formation. After expounding the problems and the central points that stand in need of explanation, I consider Davidson's view of the conflict between our standard way of describing and explaining mental phenomena and the idea that such phenomena can be irrational. Then, I present Davidson's suggestion to reconcile an explanation that shows a belief to be irrational with the element of rationality inherent in the description of any propositional attitude. I consider Davidson's and Pears' criteria for drawing a dividing line between mental systems, and argue that Davidson's criterion seems too wide and should be restricted to cases where motivation is involved. I also argue that it is the weaker version of Pears' criterion that must be defended and confronted with Davidson's.

[S.B.]