

PARPADEOS DE LIBERTAD: LA CAUSA AGENTE Y EL *PRINCIPIO*  
*DE POSIBILIDADES ALTERNATIVAS*

CARLOS G. PATARROYO\*  
Universidad Nacional de Colombia

**Resumen**

En 1969 Harry Frankfurt publica su seminal artículo "Alternate Possibilities and Moral Responsibility". Allí ataca la necesidad del *principio de posibilidades alternativas* (PPA) para la adscripción de responsabilidad moral a los agentes cuando estos realizan una acción. Desde entonces, muchos han sido los intentos de defender el PPA y de mostrar su necesidad para la adscripción de la responsabilidad moral. En este ensayo se examina uno de esos intentos: el que se basa en la causalidad-agencial propuesta por Thomas Reid (1710-1796), y que cuenta con William Rowe como uno de sus defensores dentro de la discusión contemporánea. Esto, con el objetivo de mostrar que, si bien esta defensa podría permitir la aparición de lo que se conoce como un "parpadeo" de libertad, ello no es suficiente para sustentar la atribución de responsabilidad moral.

PALABRAS CLAVE: Harry Frankfurt, parpadeo de libertad, Thomas Reid, responsabilidad moral, racionalidad.

**Abstract**

In 1969 Harry Frankfurt publishes his seminal paper "Alternate Possibilities and Moral Responsibility". His purpose is to attack the idea that the *principle of alternate possibilities* (PAP) is a necessary condition for the adscription of moral responsibility to an agent when he performs an action. Since then, many attempts have been made to defend PAP as a necessary condition for the adscription of moral responsibility. This paper examines

---

\* Doctorando en Filosofía (Universidad Nacional de Colombia).

one of these attempts: that which is based on the causal-agency proposed by Thomas Reid (1710-1796), and which has been defended within the contemporary debate, among others, by William Rowe. All of these in order to show that, although this approach may allow the appearance of what has been called a “flicker” of freedom, it is yet not enough to ground the ascription of moral responsibility.

KEY WORDS: Harry Frankfurt, flicker of freedom, Thomas Reid, moral responsibility, rationality.

## 1. Los contraejemplos de Frankfurt

Tradicionalmente se ha creído que un agente sólo es moralmente responsable de una acción que realiza si está en su poder actuar de otro modo. El conductor que atraviesa una calle cuando la luz está roja es responsable porque ha podido detener su auto y esperar a que la luz se torne verde. Esto corresponde a la vieja intuición que se tiene acerca de lo que significa ser *libre*. Y esa intuición también dice que solo se es moralmente responsable de una acción cuando quien la efectúa era libre en el momento de realizarla. Es gracias a esta intuición que se exculpa a individuos que padecen de enfermedades mentales, a quienes resultan dominados por su propia ira en cierto momento (llamado “locura temporal”) y a las víctimas de coerción. Esta intuición ha sido formalizada de muchas maneras, pero la más conocida de ellas es la que Harry Frankfurt ha propuesto en su escrito “Alternate Possibilities and Moral Responsibility”: “Una persona es moralmente responsable de lo que ha hecho sólo si hubiera podido actuar de otra manera”<sup>1</sup>. A esto se lo ha llamado el *principio de posibilidades alternativas* (en adelante PPA), y el artículo de Frankfurt está encaminado a ofrecer un fuerte ataque contra la idea popular de su necesidad para la adscripción de responsabilidad moral. Lo que desea mostrar Frankfurt es que un agente puede ser moralmente responsable de una acción, aun si no podía haber actuado de otra manera.

Para sustentar esto, acude a una serie de contraejemplos en los que presenta a dos agentes: uno está a punto de realizar una acción y el otro (llamado interventor contra-fáctico) tiene el poder de influir en las acciones del primero. Sin embargo, no desea hacerlo a menos que sea necesario, así que espera a que el primero dé algún signo de qué es lo que va a hacer. Si la acción es la que el interventor contra-fáctico desea, no intervendrá, pero si es

<sup>1</sup> Frankfurt (1969: 829).

la contraria, manipulará al primer agente para que realice cierta acción. Un ejemplo de este estilo sería el siguiente:

Cástor y Pollux se conocen hace años. Sin que Pollux lo sepa, Cástor ha desarrollado un dispositivo que le permite controlar el cerebro de Pollux, de manera que tiene control sobre sus deseos y acciones. Sin embargo, Cástor no desea utilizar este dispositivo a menos que sea necesario. Ambos se encuentran en el auto que Pollux conduce. De repente, en medio de la vía, hay un hombre que ambos reconocen como el ladrón que los dejó en bancarrota hace unos días. Pollux, que tiene el control del automóvil, puede hacer dos cosas: decidir virar para esquivar al hombre o decidir seguir recto para arrollarlo. Cástor desea que Pollux arrolle al hombre, pero decide esperar a que Pollux dé algún signo de qué es lo que va a hacer. Si Pollux no hace nada, será indicación de que desea arrollar al ladrón y entonces Cástor no hará nada. Pero si, por ejemplo, arquea las cejas y grita, será señal de que va a virar para esquivarlo. En esa situación Cástor prenderá el dispositivo manipulando el cerebro de Pollux de manera que este forme la intención de matar al ladrón y la siga con su acción.

Según Frankfurt, en esta situación Pollux sólo tiene una posible vía de acción: arrollar al ladrón. No hay posibilidades alternativas ( $PA$ ), pues ya sea que él mismo decida arrollarlo, por sus propias razones y sin intervención de Cástor, o que en un principio parezca que va a decidir no hacerlo, en ambos casos tanto la formación de la intención como el resultado serán los mismos. Nótese que aquí no se trata de un caso de coerción *posterior* a la formación de una intención, pues se supone que Pollux no ha decidido conscientemente no arrollar al ladrón, sino que hasta ahora sólo ha mostrado un signo de que esa será la decisión que va a tomar. Cástor interviene antes de que forme la intención de no arrollarlo, produciendo con su dispositivo la intención en Pollux de arrollar al ladrón. Sin embargo, el punto fundamental está, para Frankfurt, en el hecho de que en una de las dos situaciones, la primera, se tiende a afirmar intuitivamente que Pollux es moralmente responsable, mientras que en la segunda, en la que Cástor ha intervenido ya, no se atribuye responsabilidad moral a Pollux. Cuando Pollux decide arrollar al ladrón, sin intervención de Cástor ni de su dispositivo, se dice que es moralmente responsable. Cuando, por el contrario, lo ha hecho bajo el efecto del dispositivo de Cástor, se dice que no es moralmente responsable.

Hay una diferencia entre la adscripción o no de responsabilidad moral en ambas situaciones, pero el punto de Frankfurt es que r.o hay  $PA$ , ni siquiera cuando instintivamente se ha atribuido responsabilidad a Pollux. Es por eso

que Frankfurt concluye que el PPA no es necesario para la adscripción de responsabilidad moral:

Pero el principio de posibilidades alternativas es falso. Una persona puede ser moralmente responsable de lo que ha hecho aunque no haya podido actuar de otro modo. La plausibilidad del principio es una ilusión.<sup>2</sup>

Hay, sin embargo, una reacción casi que instintiva ante los contraejemplos de Frankfurt y su conclusión. Lo primero que viene a la cabeza es que Pollux sí ha tenido una PA, pues ha intentado hacer otra cosa, pero ha sido obligado, después de ello, a hacer la acción que Cástor deseaba. En otras palabras, el hecho de que Cástor sólo pueda actuar después de que Pollux ha mostrado un signo que indica que va a hacer otra cosa, parece dar cabida a una PA real. A este pequeño atisbo de libertad es a lo que John Martin Fischer ha llamado un “parpadeo” (*flicker*) de libertad.

## 2. La estrategia de los parpadeos de libertad

Los casos tipo-Frankfurt parecen, a primera vista, no involucrar posibilidades alternativas. Pero después de una inspección más cercana se puede ver que, pese a que no involucran posibilidades alternativas del tipo normal, sí pueden involucrar algunas posibilidades alternativas. Esto es, pese a que el interventor contra-fáctico elimina la mayoría de las posibilidades alternativas, posiblemente no elimina todas las posibilidades: aún en los casos tipo-Frankfurt parece haber un “parpadeo de libertad”.<sup>3</sup>

Cuatro son las maneras en las que se puede defender que hay un parpadeo de libertad en los contraejemplos de Frankfurt<sup>4</sup>. La primera de ellas dice que el interventor contra-fáctico sólo puede actuar cuando Pollux ha mostrado algún signo relevante de qué es lo que va a hacer. En el ejemplo antes visto, este signo es arquear las cejas y gritar. Sólo entonces Cástor hará su intervención y forzará a Pollux a arrollar al ladrón. Se puede decir que Pollux no tenía otra opción más que la de arrollar al ladrón (y por ello Fischer en la cita dice que no hay PAS del tipo normal), pero eso no descarta el que

---

<sup>2</sup> Frankfurt (1969: 829).

<sup>3</sup> Fischer (1994: 134).

<sup>4</sup> Para una descripción detallada de ellas, cfr. Fischer (1994: 134-147).

Pollux sí ha tenido una PA menor: la de arquear las cejas y gritar. Las dos cadenas de eventos no son iguales. En una no hay cambio en el comportamiento de Pollux y por ello Cástor no interviene, pero en la otra sí hay un pequeño cambio que, si bien no afecta el resultado final (el ladrón sigue siendo arrollado por el auto), sí constituye una PA. Llamaremos a esta primera defensa la *muestra de un signo relevante*.<sup>5</sup>

La segunda defensa puede ser llamada *defensa del principio de individuación de eventos*. A esta defensa subyace la idea de que cuando un agente realiza una acción, causa la ocurrencia de cierto tipo de evento. Según el principio de individuación de eventos, los antecedentes causales que de hecho dieron lugar a un evento específico son esenciales a este; esto quiere decir que si el evento en cuestión es *e*, cualquier evento que tenga antecedentes causales diferentes a los de *e* no podrá ser idéntico a *e*. Aplicado esto al ejemplo anterior, se tienen dos situaciones: por un lado aquella en la que Pollux arrolla al ladrón sin intervención de Cástor; por el otro, aquella en la que Pollux arrolla al ladrón gracias a la intervención de Cástor. Si bien el resultado de ambas parece ser el mismo (el ladrón es arrollado), los antecedentes causales son distintos, pues en uno interviene Cástor y en el otro, no. Por el principio de individuación de eventos antes visto, se tiene entonces que los dos eventos finales no pueden ser idénticos. El parpadeo de libertad está aquí en el hecho de que, si bien no estaba en poder de Pollux causar el evento de no arrollar al ladrón, sí ha estado en su poder causar un evento distinto con el mismo resultado, puesto que Cástor sólo interviene si Pollux muestra algún signo de que no va a arrollar al ladrón. El hecho de que esté en poder de Pollux causar un evento diferente es una muestra de que tiene una PA y se le juzgará moralmente responsable en el caso en el que él mismo decide arrollar al ladrón, puesto que esta alternativa estaba disponible en el momento de su acción.

La tercera defensa puede ser llamada *defensa de la especificidad*. Se basa en el análisis cuidadoso de qué es aquello de lo que Pollux es moralmente

<sup>5</sup> Los contraejemplos tipo-Frankfurt que se construyen con base en este "signo relevante" han sido fuente de una acalorada discusión. Algunos autores (cfr. Widerker (2002) y Ginet (2003)) han elaborado un argumento conocido como *la defensa del dilema*, según el cual si el signo relevante precede invariablemente a la acción de la que es signo, ello implicará que la acción está causalmente determinada por algún proceso sobre el cual el agente no tiene control alguno y, por ello, no sería jamás libre, es decir, jamás tiene posibilidades alternativas (aún en ausencia del interventor contra-fáctico). Pese a lo interesante de esta discusión, no es la intención de este artículo defender a los contraejemplos de Frankfurt de este ataque, sino mostrar que la causa agente no sirve para defender el principio de posibilidades alternativas. Quien desee conocer algunas respuestas al ataque de Widerker y Ginet puede remitirse a Mele y Robb (2003) y Pereboom (2007).

responsable. ¿Acaso es responsable de arrollar al ladrón? ¿Acaso de decidir arrollar al ladrón? Quienes acogen esta defensa dicen que se juzga moralmente responsable a Pollux de algo mucho más específico que esto: “decidir por sí mismo arrollar al ladrón”. Ya que en las dos secuencias de eventos se puede decir que claramente hay una en la que Pollux no decide por sí mismo (pues lo hace gracias a la influencia del dispositivo de Cástor), se puede ver que hay una PA disponible para Pollux, pues puede decidir por sí mismo, o no decidir por sí mismo.

La cuarta defensa es una reformulación de la segunda, con la diferencia de que esta rescata la teoría de la causalidad-agencial propuesta por Thomas Reid. La idea detrás de ella es que, si bien en ambas secuencias de eventos hay una volición antes de la acción, la diferencia (el parpadeo) está en que las dos voliciones no se producen de la misma manera. Una de ellas es producida por el aparato de Cástor, mientras que la otra es causada por el agente mismo, de una manera muy particular: mediante un proceso conocido como causa-agente, que no es reducible a la causalidad de eventos. Según la teoría de Reid, un agente sólo puede ser moralmente responsable de lo que él mismo causa agencialmente, y no lo es de lo que es causado eventualmente. Así, Pollux tiene PAS, pues puede ser la causa-agente de su volición, o puede no serlo. Ya que esta será la defensa en la que se centrará el presente artículo, y ya que ella parece involucrar elementos de las tres defensas anteriores, a continuación se explicará en detalle en qué consiste la causalidad-agencial de Reid y cómo ella puede ser utilizada como defensa del PPA, para luego exponer una crítica que sirve para refutar a esta defensa.

### 3. Reid y la causa-agente

Pese a presentar una fuerte crítica a la filosofía empirista y a sus consecuencias escépticas, Reid es heredero de muchas de las propuestas del empirismo inglés. Para el tema que se trata en este artículo, es de especial importancia la herencia que recibe de Locke acerca de los *poderes* que ostentan los cuerpos. Se ha de recordar que para Locke hay dos clases de poderes en los cuerpos: por un lado está el poder activo, que se presenta cuando un cuerpo es “capaz de efectuar algún cambio”,<sup>6</sup> mientras que el pasivo se presenta cuando un cuerpo es “capaz de sufrirlo”.<sup>7</sup> El ejemplo más conocido con el que Locke ilustra esta diferencia es el de la cera y el sol. El sol tiene el poder activo de producir un cambio en la cera, i.e., derretirla. A la vez que la

<sup>6</sup> Locke (1992: 215).

<sup>7</sup> Locke (1992: 215).

cera tiene el poder pasivo de ser afectada por el sol. La cera no puede producir este cambio en sí misma, sino que solo puede sufrirlo al ser afectada por otra cosa, en este caso, una fuente de calor.

Reid hace dos modificaciones a la teoría de Locke. La primera de ellas es descartar el poder pasivo:

Pese a que él divide el poder en “activo” y “pasivo”, no creo que el “poder pasivo” del señor Locke sea un poder en absoluto. Con este nombre él se refiere a la posibilidad de ser cambiado. Llamar a esto “poder” parece ser una mala aplicación de la palabra. No recuerdo haber encontrado el nombre “poder pasivo” en ningún otro buen autor. El señor Locke parece haber sido desafortunado al inventarlo, y debería ser permitido que fuera desechado de nuestro lenguaje.<sup>8</sup>

La segunda es especificar que el poder activo debe ser originado *en* el agente y no fuera de él. El ejemplo perfecto para ilustrar esto es la bola de billar. Si bien la bola de billar tiene el poder de producir un cambio en otra con la que colisiona, si este poder ha sido causado, a su vez, por el taco de billar que la ha golpeado, entonces se dirá que la bola no tiene un poder activo. Es cierto que en un caso así Locke también diría, con Reid, que la bola no ha tenido un poder activo. Pero la modificación de Reid a la teoría de Locke está en que para él poseer un poder activo para producir algo implica también la capacidad de poder no producirlo, y esto es lo que parece faltar en Locke:

El poder de producir un efecto implica el poder de no producirlo. No podemos concebir cómo un ser pueda tener el poder de hacer x, pero no tener el poder de no hacer x.<sup>9</sup>

¿Cómo puede ser posible que algo tenga un poder activo, si ello implica tener la capacidad de no utilizarlo también? Visto de esta manera el sol no tendría un poder activo, pues pese a que puede derretir la cera, y ese poder viene de sí, no tiene el poder de *no* derretir la cera. Pero esto no es un contraejemplo o una deficiencia de la teoría de Reid. Por el contrario, él mismo acepta esto como una consecuencia acertada de sus ideas. Para él hay dos tipos de lenguaje, el filosófico, que es preciso y acertado, y el popular, que está plagado de imprecisiones y términos mal empleados:

---

<sup>8</sup> Reid (1983a: 519).

<sup>9</sup> Reid (1983a: 65).

Pese a que los filósofos, para lograr ser comprendidos, deben hablar el lenguaje de los vulgares —como cuando dicen—“el sol se levanta y se pone, y pasa a través de todos los signos del zodiaco”—, frecuentemente piensan de manera muy diferente a la de los vulgares.<sup>10</sup>

La importancia de esta diferencia está en que en el lenguaje popular se habla de los objetos como si estos realizaran acciones (el sol “se levanta”, en la cita), pero, para Reid, los seres humanos no comprenden realmente cuál es la causa eficiente que produce los cambios en los cuerpos. No importa cuán profundas sean las investigaciones, no está en la capacidad de los hombres entender cómo es que funcionan las cuestiones de la naturaleza, y es por ello que, para evitar las complicaciones en el habla popular, el filósofo se refiere a los cambios de los objetos como si ellos mismos fueran la causa de sus propias modificaciones, cuando en realidad no se sabe si ellos son sólo títeres manipulados por otras causas que se desconocen:

En el teatro de la naturaleza vemos innumerables efectos que requieren un agente dotado de poder activo, pero el agente está fuera de escena. Si es la Causa Suprema, o una causa o causas subordinadas; y si las causas subordinadas son empleadas por el Todopoderoso, cuál sea su naturaleza, su número y sus diferentes oficios son cosas ocultas al ojo humano, por razones sabias sin lugar a dudas.

Es solo en las acciones humanas, que pueden ser imputadas para alabanza y pena, que es necesario para nosotros conocer quién es el agente; y en esto la naturaleza nos ha dado toda la luz que es necesaria para nuestra conducta.<sup>11</sup>

Se puede ver que para Reid es solo en el habla popular que los objetos inanimados pueden tener un poder activo, pero en la jerga filosófica solo se atribuye poder activo a los agentes. Puede ser que los objetos inanimados sean causas secundarias, como la bola de billar que choca con otra, detrás de los cuales haya un agente que ha producido, por sí mismo, el inicio del cambio.

Adicionalmente, Reid es bastante enfático en decir que el poder activo sólo puede estar en posesión de quien tiene *voluntad*. Esto no es difícil de entender si se recuerda que parte de este poder es *no* utilizarlo en ocasiones, cosa que ningún objeto inanimado puede hacer:

Me parece, entonces, más probable que solo los seres que poseen algún grado de entendimiento y voluntad, pueden poseer poder activo; y los

---

<sup>10</sup> Reid (1983b: 310).

<sup>11</sup> Reid (1983b: 313).

seres inanimados deben ser meramente pasivos y no tener actividad real. Nada de lo que percibimos fuera de nosotros nos suministra ningún fundamento fuerte para adscribir poder activo a un ser inanimado; y todo lo que podemos descubrir en nuestra propia constitución nos lleva a pensar que el poder activo no puede ser ejercido sin voluntad e inteligencia.<sup>12</sup>

No se trata solamente de que el poder activo sea exclusivo de los seres que poseen voluntad e inteligencia, la idea fundamental de Reid es que un agente sólo es responsable de algo si lo ha hecho como producto del ejercicio de su voluntad, es decir, si él mismo ha sido la causa de su acción:

El poder *humano* sólo puede ser ejercido por la voluntad, y somos incapaces de concebir algún poder activo que sea ejercido sin la voluntad. Cada hombre sabe infaliblemente que lo que ha hecho por su voluntad consciente e intencional, le ha de ser imputado, como agente o causa; y lo que sea que ha hecho sin su voluntad e intención no le puede ser imputado correctamente.<sup>13</sup>

Una vez se han visto todos estos elementos de la teoría de Reid, se puede establecer la máxima de la imputación de responsabilidad moral de la siguiente manera: “un agente es moralmente responsable de lo que ha hecho, sólo si él es causa-agente de ello”, donde para ser causa-agente se han de cumplir tres condiciones:<sup>14</sup>

- a. El agente es la única causa de sus acciones o intenciones.
- b. Su poder de causalidad puede ser ejercido en dos direcciones: o bien hacer algo, o bien abstenerse de hacerlo.
- c. Su causar una acción o elección libre es la causación de un evento u ocurrencia por un agente, que no puede ser explicada como la causación de un evento por otra serie de eventos.

La primera condición se establece para impedir que un agente sea como una bola de billar, es decir, que sus acciones sean el producto de fuerzas externas a él y fuera de su control. La segunda postula el requerimiento de

---

<sup>12</sup> Reid (1983b: 308).

<sup>13</sup> Reid (1983b: 307).

<sup>14</sup> Estas tres características son expresadas de muchas maneras por los seguidores y críticos de la teoría de Reid, sin embargo, y pese a la diferencia de formulación, todas comparten el mismo contenido que aquí se recoge. Para ver algunas de las formulaciones más populares, cfr. Kane (1995: 118) y Rowe (2003: 222).

que no se trata solo del poder de ejercer un cambio (como el sol) sino también de la capacidad de controlar ese poder, y, de esta manera, utilizarlo en ciertas ocasiones y en otras no (es por eso que solo seres con voluntad pueden tener el poder activo). La última condición muestra que la causalidad de los agentes en la teoría de Reid, si bien es tratada como un tipo de causalidad, no es igual ni puede ser reducida a la causalidad presente en los objetos inanimados. Se trata de un tipo de causalidad especial. Lo que la hace especial es su poder de iniciar una cadena de eventos sin ella misma depender de una cadena anterior que la determine<sup>15</sup> (ahí radica, para Reid, la libertad del agente).

Una vez que se conoce la generalidad de la teoría de la causa-agente de Reid, se puede ver cómo es que esta entra a formar parte de una de las defensas del PPA en contra de los contraejemplos de Frankfurt. William Rowe es hoy en día el defensor más asiduo de la reivindicación de la olvidada teoría de Reid,<sup>16</sup> y es él quien propone que con ella se puede enfrentar el reto propuesto por los contraejemplos de Frankfurt.

#### 4. Rowe, la causa-agente y los contraejemplos de Frankfurt

Sería fácil pensar, en un principio, que la teoría de Reid no tiene cabida frente a los contraejemplos de Frankfurt. Si el sujeto, para ser moralmente responsable de una acción, debe ser causa-agente de ella, y si para ser causa-

---

<sup>15</sup> Varias son las maneras en las que se ha intentado defender la plausibilidad de esta diferencia entre los tipos de causalidad. Un ejemplo de ello está en O'Connor, quien defiende que se debe concebir la causalidad de eventos tal como se la concibe según la mecánica clásica, y la causa-agente debe ser concebida como un género de causalidad distinto, gobernado por una propiedad emergente del cerebro humano (la conciencia). Esta misma línea de defensa es utilizada por Rowe. Clarke, utiliza una estrategia diferente, propone concebir la causalidad de eventos como meramente probabilística, de manera que la indeterminación característica de la causa-agente no sea incompatible con la concepción general de la causalidad. Ya que no es el objetivo de este artículo el de defender la plausibilidad general de la teoría de la causa-agente, sino mostrar que, aun si se acepta su plausibilidad, su uso para la defensa del PPA no resulta satisfactoria, no se incurrirá en discusiones como las que O'Connor, Rowe y Clarke mantienen al respecto. La crítica que aquí se presentará no se basa ni depende de ninguna de las maneras específicas de defender esta diferencia. Para quien esté interesado en esta discusión, cfr. O'Connor (2003: 257-265), Rowe (2005: 243-247) y Clarke (1993: 285-296).

<sup>16</sup> Marjorie Naylor ya había desarrollado una estrategia similar para defender este tipo de libertad (cfr. Naylor, 1984). Su propuesta no se diferencia mucho de la de Rowe; sin embargo, en este escrito se utilizará a Rowe como el interlocutor principal, en vista de que es a él (y a sus argumentos concretos) a los que responden autores como Fischer. Para la defensa de Rowe, no solo de la causalidad-agencial, sino de los demás aspectos de la filosofía moral de Reid y su vigencia en la actualidad, cfr.: Rowe (1991).

agente de una acción una de las características es poder no efectuarla, entonces los sujetos no son causa-agente en los casos que Frankfurt postula. Es cierto que Pollux tenía el poder de arrollar al ladrón con su auto, pero no es cierto que hubiera podido *no* arrollarlo. Pareciera entonces que Pollux no puede referirse en cuanto al uso de su poder, incumpliendo con la segunda condición de las causas-agente.

Si Pollux no es una causa-agente en el ejemplo que aquí se trata, no puede hacerse una defensa del PPA, pues la idea de Rowe es mostrar que el sujeto es una causa-agente, y que si es moralmente responsable en una de las dos situaciones, es porque tenía disponible una PA, a saber, la de no ejercer su poder para realizar la acción.

Sin embargo, Rowe dice que se ha de examinar con detenimiento qué es lo que está en poder del agente para hacer y no hacer, y encuentra dos posibles situaciones:<sup>17</sup>

- (i) Está en su poder *no desear* arrollar al ladrón.
- (ii) Esta en su poder *no causar su deseo* de arrollar al ladrón.

Las dos situaciones son muy diferentes, pues en el ejemplo que se ha venido tratando (i) es falsa respecto de Pollux. No está en su poder no desear arrollar al ladrón, pues ya sea que él mismo lo haga sin intervención de Cástor, o que Cástor intervenga para que su dispositivo forme en él el deseo de arrollar al ladrón, en ambos casos terminará por formarse el deseo. Pero, para Rowe, (ii) es verdadera respecto de Pollux, pues este, al mostrar un signo de que va a virar y evitar arrollar al ladrón, produce un evento en el que Cástor ha de intervenir, utilizando su dispositivo y formando la intención de arrollar al ladrón en Pollux. Pero esta formación de la intención ya no es debido a una causa-agente, pues se trata de la formación hecha por algo externo al agente (el dispositivo de Cástor, violando así la primera condición de las causas-agente). Según Rowe, al mostrar un signo de lo que va a hacer, Pollux es una causa-agente, del hecho que le impide ser una causa-agente de su deseo de arrollar al ladrón. Esto es complicado, pero es posible desglosarlo para que se entienda mejor. En (i) Pollux no es causa-agente, pues no está en su poder *no desear* arrollar al ladrón. En (ii), sin embargo, Pollux sí es una causa-agente, pues sí está en su poder *ser la causa de su deseo* de arrollar al ladrón, o no serlo. Para explicar esta diferencia, Rowe acude al ejemplo de un tren, donde un agente tiene el poder de desviarlo o no. Estos son los tres escenarios que Rowe plantea:

<sup>17</sup> Cfr. Rowe (2003: 226).

*Escenario 1:* El tren se aproxima por una vía. En cierto punto hay una bifurcación que lo llevará por la vía alterna 2. Al final de esa vía hay un perro amarrado, al que el tren atropellará si va por esa vía. El agente tiene en su poder una palanca que controla el cambio de las vías. Si el agente no hala la palanca, el tren sigue por la vía 1 y no atropella al perro. Si la hala, el tren cruza hacia la vía 2 y mata al animal. "Aquí parece claro que el hecho de que el tren atropelle o no al perro depende enteramente de usted [el agente]. Y si usted hala la palanca, desviando el tren a la vía 2, usted es causalmente responsable de que el tren atropelle al perro".<sup>18</sup>

*Escenario 2:* El tren se aproxima por la vía 1 y al final de la vía 2 está amarrado el perro. El agente tiene en su manos la palanca que desvía al tren hacia la vía 2. Sin embargo, la vía 1, unos metros después de la bifurcación sobre la que tiene control el agente, se une a la vía 2. Así, no importa qué haga el agente ni por qué vía pase el tren en el momento de la bifurcación, terminará por atropellar al perro.

¿De qué es usted [el agente] causalmente responsable en el escenario II? En particular, ¿es usted causalmente responsable de que el tren atropelle al perro? No. Pues aquí, antes de que usted haga cualquier cosa, ya es causalmente inevitable (dadas las circunstancias) que el tren atropellará al perro. A lo sumo, usted es causalmente responsable de que el tren proceda por la vía 2 y de que atropelle al perro *por la vía 2*.<sup>19</sup>

*Escenario 3:* El tren se aproxima por la vía 1, pero esta vez hay dos bifurcaciones que pueden desviar al tren hacia la vía 2. En la primera está el agente que nos interesa y en la otra está el interventor contrafáctico. Este interventor sólo halará su palanca si nota que el agente que está en la primera bifurcación no hala la suya y permite que el tren siga por la vía 1.

¿De qué es usted [el agente] causalmente responsable en el escenario 3? Mis intuiciones me dicen que en el escenario 3, a diferencia del escenario 2, usted es causalmente responsable de que el tren atropelle al perro. Pues en el escenario 3 depende de usted que sea usted mismo la causa de que el tren atropelle al perro, o que un evento posterior ocurra (que la segunda palanca sea halada) y cause que el tren atropelle al perro.<sup>20</sup>

El primer escenario corresponde a la intuición tradicional acerca de la libertad y el PPA. Hay varios cursos de acción disponibles al agente y cada

<sup>18</sup> Rowe (2006: 305).

<sup>19</sup> Rowe (2006: 306).

<sup>20</sup> Rowe (2006: 307).

uno de ellos tiene un resultado diferente, del cual el agente será responsable. El escenario 2 involucra una suerte de fatalismo, en donde no importa qué haga el agente, el resultado será el mismo. El escenario 3, en cambio, presenta una situación en la que, si bien el resultado será el mismo (el perro es atropellado), la causa de ello es distinta, pues el perro puede ser atropellado porque el agente haló la palanca o porque el interventor la haló. Cuando el agente no hala su palanca, *causa* una situación en la que propicia que otra causa posterior actúe.

Traducido esto al ejemplo que se ha venido utilizando de Cástor y Pollux, se puede decir que el resultado será el mismo, pues el ladrón será atropellado, pero Pollux puede ser la causa-agente en el sentido en que puede propiciar una situación en la que otra causa posterior será la responsable de su deseo de arrollar al ladrón. Es por eso que se puede decir, tal como ya se hizo unas líneas más arriba, que Pollux no es causa-agente de su deseo de arrollar al ladrón, pero sí es causa-agente de que sea él o no la causa de su deseo de arrollar al ladrón.

Si esto es correcto, como defiende Rowe, Pollux es moralmente responsable de arrollar al ladrón aún cuando Cástor no interviene, porque ha tenido la *PA* de ser la causa de su deseo de arrollarlo. Se ha de recordar que en el ejemplo Pollux sólo es responsable en una de las dos situaciones, a saber, aquella en la que arrolla al ladrón sin que intervenga Cástor. Lo que Frankfurt defiende es que es moralmente responsable pese a que no ha tenido *PAS*, lo cual demostraría la futilidad del *PPA* para la adscripción de responsabilidad moral. Sin embargo, la estrategia de Rowe es mostrar que Pollux sí ha sido causa-agente de una de las situaciones involucradas en el ejemplo. Si es una causa agente es porque ha tenido *PAS*, ya que, según la segunda condición de las causas-agente, se ha de poder utilizar el poder o (esta *o* señala que hay una *PA*) abstenerse de utilizarlo. Pollux ha logrado ponerse en una situación en la que puede utilizar su poder para generar dos situaciones, o bien una en la que él es la causa de su intención (utiliza su poder), o bien otra en la que la causa es externa y él ya no se consideraría causa-agente de su intención (se abstiene de utilizarlo y entonces otra causa posterior interviene). Así, habría un parpadeo de libertad en el que el agente tiene una *PA* y es por ello que se lo considera responsable de su acción. Si se lo considera responsable de su acción gracias a que ha tenido una *PA*, entonces se logra demostrar que el *PPA* no ha sido refutado y que sigue siendo fundamental para la adscripción de responsabilidad moral. Por supuesto, esta propuesta no ha pasado desapercibida y ha recibido fuertes ataques, en especial de John Martin Fischer.

## 5. Fischer vs. Rowe y el argumento de la insuficiencia de *oomph*!

Rowe ofrece su defensa del PPA en 1991, con la publicación de su libro *Thomas Reid on Freedom and Morality*, a la que responde Fischer en su libro *The Metaphysics of Free Will* (1994). Rowe responde a Fischer en su artículo "Alternate Possibilities and Reid's Theory of Agent-causation" (2003), al que Fischer responde ese mismo año con su escrito "Responsibility and Agent Causation". El último golpe en la pelea lo ha dado Rowe cuando publica, en 2006, su artículo "Free Will, Moral Responsibility and the Problem of *oomph*". Los argumentos van y vienen, pero ¿cuál es el punto en disputa?

La crítica de Fischer a Rowe parte de admitir que en los contraejemplos de Frankfurt sí hay unas pequeñas PAS, y es por eso que es él mismo quien las bautiza "parpadeos de libertad": "es difícil ver cómo un ejemplo del tipo Frankfurt pueda ser construido de manera que no tuviera en absoluto tal parpadeo".<sup>21</sup> Sin embargo, dice que este parpadeo no es lo suficientemente *robusto* como para poder ser base de la responsabilidad moral:

aun si hay un tipo de parpadeo de libertad aquí, no parece ser capaz de jugar el rol requerido en fundamentar adscripciones de responsabilidad moral –no parece ser lo suficientemente robusto–.<sup>22</sup>

¿Por qué considera Fischer que la PA propuesta por Rowe no es lo suficientemente robusta? Básicamente porque ella no implica una intención. Se ha de recordar que según el ejemplo, Cástor no reacciona cuando Pollux actúa intencionalmente, ni siquiera lo hace cuando éste ha formado una intención definida de no arrollar al ladrón; actúa cuando Pollux *da un signo* de estar a punto de formar la intención de virar el auto y no arrollar al ladrón. La PA propuesta por Rowe se presenta antes de este proceso, precisamente se trata de la posibilidad de mostrar o no mostrar *el signo*. Pero el signo es anterior a la intención y es por eso que Fischer dice que no se puede pensar que alguien como Pollux actúe libremente, pues no forma la intención de refrenar su deseo de arrollar al ladrón. Bajo esta interpretación, un agente es libre cuando tiene el poder de actuar o no actuar, pero ambas acciones han de ser realizadas intencionalmente. En otras palabras, Fischer parece decir que la segunda condición de las causas-agente debe ser entendida de la siguiente manera: "el poder de causalidad de un agente puede ser ejercido en dos direcciones: o bien hacer algo *intencionalmente*, o bien abstenerse de hacerlo

<sup>21</sup> Fischer (1994: 136).

<sup>22</sup> Fischer (1994: 143).

*intencionalmente*”, donde este “intencionalmente” se ha de entender como “formando una intención previa”. Ya que la PA que propone Rowe es *previa* a la formación de la intención de arrollar al peatón, o la de no arrollarlo, Fischer dice que se trata de un evento a-intencional y que por eso no se puede considerar que cumpla con la segunda condición de las causas-agente. Si no cumple con esta condición, no se trata de una causa-agente y de allí se deriva que tampoco se trata de un caso que logre defender la pertinencia del PPA para la adscripción de responsabilidad moral. Es a esto a lo que se le ha llamado, en el seno de este debate —y basado un poco en la camaradería de los participantes—, la objeción de la insuficiencia de *oomph*. *Oomph* es una onomatopeya norteamericana, que trata de imitar el sonido que los hombres producen cuando una mujer muy atractiva pasa delante de ellos (como una especie de gruñido acompañado de una exhalación). La idea es que las PA propuestas por Rowe no son lo suficientemente *atractivas*:

En la secuencia alternativa, Jones [Pollux en el ejemplo que se ha venido tratando] *no* forma una intención de refrenarse de causar la volición en cuestión, y proceder entonces a llevar a cabo esta intención de la manera apropiada. Nuevamente, se puede conceder que Jones tiene el poder de no causar la volición de votar por Clinton [arrollar al ladrón]. Pero al no causar esta volición estaría, por supuesto, actuado de manera *no libre*, porque gracias a la intervención de Black [Cástor], no sería verdad que Jones libremente se refrena de causar la volición de votar por Clinton [arrollar al ladrón].<sup>23</sup>

La respuesta de Rowe a esta crítica está dada en dos partes: en primer lugar, acusa a Fischer de confundir los dos niveles en los que se desarrolla el ejemplo. Es cierto que Pollux no es libre de decidir si arrolla o no al ladrón, pero sí es libre de ser causa o no de la intención de arrollar al ladrón. En una secuencia es él quien es causa de su intención, y en la otra produce un evento en el que otra causa posterior es la que ocasiona su intención de arrollar al ladrón. El error de Fischer está, para Rowe, en creer que Cástor actúa también en la situación en la que Pollux forma o no el signo que señala que va a actuar de cierta manera. Pero, como claramente lo muestra el ejemplo, el actuar de Cástor sólo es posterior a la aparición del signo de Pollux. La segunda respuesta tiene que ver con la presencia de una intención en cada una de las posibles PAS que Rowe presenta. Es cierto que hay una intención de arrollar al ladrón, dada por dos vías: la formada por Pollux libremente, y la que es inducida gracias al dispositivo de Cástor. La pregunta es: se está

<sup>23</sup> Fischer (1994: 143).

hablando de la formación de una intención (la de arrollar o no al ladrón); ¿debe haber una intención previa cuyo contenido sea el de formar la intención de arrollar o no al ladrón? Para Rowe la respuesta es un *no* rotundo. Aceptar tal cosa tendría la nefasta consecuencia de que la formación de toda intención tendría que tener una intención previa, cuya formación requeriría de una intención previa, y así *ad infinitum*.

Cuando miramos cuidadosamente lo que Fischer dice, se hace claro que él usa la expresión "Oomph" de una manera en la que, a menos que la alternativa incluya una *volición* o una *intención* de hacer algo más (o de no hacer nada), entonces se sigue, *por definición* que no hay ninguna alternativa presente con suficiente "oomph" para fundamentar la responsabilidad moral. Y si esto es así, Fischer gana el debate. Pues en la teoría de Reid la alternativa a que alguien sea la causa-agente de la volición de actuar, es simplemente su *no causar-agencialmente* esa volición de actuar. Y, claramente, no causar-agencialmente una volición de actuar no es lo mismo que causar-agencialmente una volición de no actuar... Así que, por supuesto, la teoría de Reid de la causa-agente no tiene suficiente "oomph" para satisfacer a Fischer. Pues Fischer ha definido "oomph" de tal manera que es lógicamente imposible que la teoría de Reid de la causa-agente pueda satisfacer la definición de Fischer. Lo que Fischer no ha mostrado es que la teoría de la libertad de Reid es inadecuada para fundamentar adscripciones de responsabilidad moral... No hay nada bueno o malo en usar la expresión "oomph" de esta manera. Pero lo que Fischer debe hacer es construir un *argumento* serio [para la conclusión de que la PA propuesta por Rowe no es lo suficientemente robusta].<sup>24</sup>

Para Rowe, Fischer no es justo con la teoría de Reid, pues parece ser que él considera que las alternativas de un agente son: o formar la intención de hacer algo o formar la intención de no hacerlo. Rowe defiende que no es esto lo que Reid propone, pues según la teoría de la causa-agente, las opciones de un agente son: o causar la intención de hacer algo o no causarla, pero esta última opción no implica causar la intención contraria (la de no actuar). Ya que Fischer impone su modelo a la teoría de Reid, es claro que esta última no encaja en él y de ahí viene la crítica de Fischer. Lo que Fischer no ha hecho, según Rowe, es construir un argumento fuerte, que haga justicia a la teoría de Reid, y que aun así sirva para mostrar que la PA existente ahí no es lo suficientemente robusta para soportar la adscripción de responsabilidad moral. Esto es precisamente lo que se intentará hacer en la siguiente parte de este escrito.

<sup>24</sup> Rowe (2006: 312).

## 6. La respuesta a Rowe

La estrategia que se utilizará aquí para responder a Rowe comenzará por mostrar que el análisis que él hace de la teoría de Reid se centra demasiado en la voluntad del agente, pero por ello deja de lado la *racionalidad* de este. Una vez que se haya mostrado esto, se añadirá esta racionalidad a la propuesta de Rowe para mostrar que esta ha de caer inevitablemente en uno de dos extremos insatisfactorios: o bien tendría que dejar de lado la racionalidad del agente para que pueda intentar responder a los contraejemplos de Frankfurt, pero en ese caso sería el mismo Rowe (y no Fischer) quien no sería justo con la teoría de Reid; o bien le responde a los contraejemplos asumiendo que el agente posee una capacidad racional muy superior a la que los contraejemplos permiten, con lo cual Rowe no sería justo con los contraejemplos de Frankfurt. Hay que comenzar, entonces, por rescatar la racionalidad que Rowe trata tan ligeramente.

Ya se ha visto, más atrás, una cita que muestra que para Reid sólo poseen poder activo los seres que tienen voluntad y racionalidad. La cita dice:

Me parece, entonces, más probable que sólo los seres que poseen algún grado de entendimiento y voluntad pueden poseer poder activo; y los seres inanimados deben ser meramente pasivos y no tener actividad real. Nada de lo que percibimos fuera de nosotros nos suministra ningún fundamento fuerte para adscribir poder activo a un ser inanimado; y todo lo que podemos descubrir en nuestra propia constitución, nos lleva a pensar que el poder activo no puede ser ejercido sin voluntad e inteligencia.<sup>25</sup>

El requisito de la inteligencia, o el entendimiento, es postulado por Reid porque para él una causa-agente ha de ser libre, donde este “libre” se ha de entender como no determinada por nada anterior. Una determinación anterior es la que las pasiones o los instintos pueden producir en el hombre. Respecto de ellos, Reid considera que puede haber tres situaciones en las acciones humanas: la primera es aquella en la que son los deseos, los instintos, los que llevan inevitablemente a la acción (como ocurriría, por ejemplo, con los animales). Esta acción sería irresistible y por ello el agente no sería libre, o causa-agente de su acción, ya que violaría la segunda condición de las causas-agente (poder utilizar el poder o poder refrenarse de utilizarlo). En esta situación el agente no es moralmente responsable de lo que ha hecho

<sup>25</sup> Reid (1983b: 308).

porque, como ya se ha visto, sólo es responsable de aquello de lo que él es causa-agente.

Las acciones variadas de los animales inferiores son desencadenadas por sus instintos, sus apetitos y sus emociones; pero parece que ellos necesariamente *tienen* que seguir el impulso más fuerte del momento, y no tienen capacidad de auto-control. Así que nosotros no los culpamos de nada de lo que hacen, y no tenemos razón para pensar que ellos se culpan a sí mismos.<sup>26</sup>

La segunda situación es aquella en la que hay una combinación de pasiones con razón. En ella, dirá Reid, se considera al agente responsable *solamente en parte*. Se puede entender esto de la siguiente manera: algunas pasiones muy fuertes (mas no determinantes), llevarán a que el sujeto sólo considere ciertas alternativas como posibles vías de acción. Por ejemplo, en el caso de ganar inesperadamente una cierta cantidad de dinero, un sujeto tiene al menos tres opciones: donar la totalidad del dinero a una institución de caridad, donar parte y gastar parte en comprar algo para sí mismo, o no donar nada y gastarlo todo en darse gusto. Sin embargo, un sujeto para el que la avaricia sea un impulso muy fuerte, no tendrá disponible la primera opción y creará que sólo tiene que decidir entre las dos últimas, donde cada una involucra alguna ganancia para él. Se le responsabiliza de la decisión que toma entre esas dos alternativas, pero no se le responsabiliza de las alternativas que se le presentaron en el momento de tomar la decisión.

En las acciones que proceden del apetito o la pasión, somos en parte pasivos y solo en parte activos. Por lo tanto, ellas [las acciones] son imputadas parcialmente a la pasión; y si se supone que esta es irresistible, no se las imputamos en absoluto al hombre.<sup>27</sup>

Finalmente, hay una situación en la que las pasiones no tienen influjo en el sujeto y este puede decidir basado solamente en su razón.

Sin embargo, en ocasiones hay una calma en la mente, de los vendavales de la pasión o del apetito, y al hombre se le permite seguir su camino, en el viaje de la vida, sin los impulsos que estos proporcionan. Entonces, él sopesa calmadamente los bienes y los males, que están muy lejanos como para excitar alguna pasión. Él juzga qué es mejor dentro de ese todo, sin sentir ningún favoritismo que lo arrastre hacia

<sup>26</sup> Reid (1983b: 127).

<sup>27</sup> Reid (1983a: 534).

un lado. Él juzga por sí mismo, tal como lo haría por otro en su situación; y la determinación es totalmente imputable al hombre, y en ningún grado a su pasión.<sup>28</sup>

Estas tres situaciones muestran la importancia y la necesidad de la razón en la teoría de la causa-agente de Reid.<sup>29</sup> Solo en las dos últimas situaciones, en donde la razón interviene, es que se puede considerar al agente moralmente responsable de lo que ha hecho. Se puede decir, entonces, que el agente es responsable cuando es una causa-agente de su volición para hacer algo (o cuando se refrena de causar tal volición), y, a la vez, utiliza su poder para causar la volición o para refrenarse de causarla, *porque puede aducir una razón para ello*. De lo contrario, se trataría de un ejercicio arbitrario, a-racional del poder, en el que el agente, independientemente de su razón, en ocasiones usa su poder y en ocasiones no lo usa.

Una vez que se ha rescatado la importancia de la razón para la teoría de la causa-agente de Reid, se puede regresar al uso que hace Rowe de esta teoría como defensa del PPA. Se ha de recordar que, para Rowe, Pollux puede ser causa-agente (y por ello libre, con PAS), pues logra causar el surgimiento de un evento en el que él ya no es la causa de su intención de arrollar al ladrón. El parpadeo de libertad está en que él puede ser la causa de su intención de arrollar al ladrón, o puede hacer algo para que alguien más lo sea. Fischer critica a Rowe al decir que este “hacer algo” para que alguien más sea la causa de la intención de Pollux es algo que no se hace intencionalmente (pues es previo a la formación de la intención de arrollar al ladrón), y que, al no ser intencional, no puede ser base de la adscripción de la responsabilidad moral. Rowe se defiende, a su vez, diciendo que, según la teoría de Reid, un sujeto puede ser causa-agente cuando utiliza su poder para causar la intención de arrollar al ladrón, o cuando se refrena de usarlo para causar esta intención; pero que esto último no implica (como parece defender Fischer) utilizar el poder para formar la intención de *no* arrollar al ladrón. Precisamente porque el poder no se utiliza, es que no se puede decir que refrenarse de usarlo deba ser entendido como *utilizarlo* para formar la intención contraria. Rowe acusa a Fischer de no lograr hacer justicia a la teoría de Reid y, por

<sup>28</sup> Reid (1983a: 534).

<sup>29</sup> Sería injusto decir que Rowe no se ha dado cuenta de la importancia de la razón en Reid. Por el contrario, sí lo hace, incluso, las dos últimas citas, que muestran las dos situaciones en las que la razón interviene, son también citadas por él en uno de sus artículos (cfr. Rowe, 2003: 224). La crítica aquí presente apunta a que, si bien en su exposición de la teoría de Reid él muestra el papel de la razón, en el momento de aplicar esa teoría a los contraejemplos de Frankfurt, la razón queda marginada.

ello, de no refutar su defensa de las PPA correctamente. Sin embargo, si hay una manera de refutar esta teoría correctamente si se coloca a la razón en su lugar justo en la teoría de Reid.

Al rescatar a la razón se ha podido ver que el agente es causa-agente cuando utiliza su poder o se refrena de utilizarlo, de acuerdo con una razón para ello. En otras palabras, el sujeto es un ser cuya capacidad racional le permite sopesar distintas razones para formar o no una intención. Así, una razón acompañaría al uso del poder activo en la formación de la intención, y otra tendría que acompañarlo en el refrenamiento de su uso (si no se desea que este refrenamiento sea del todo irracional). ¿Qué es lo que hace Pollux en el ejemplo que se ha venido tratando? Pollux tiene, según Rowe, dos opciones: o bien ser causa de su intención de arrollar al ladrón, o bien refrenarse de ello y no ser la causa de su intención; cada una de estas opciones debe venir acompañada de una razón. Es aquí donde se forman una serie de escenarios en los que, sea cual sea la opción que Rowe escoja, no le servirá para defender el PPA sin violentar o bien la teoría de Reid, o bien los contraejemplos de Frankfurt.

Pollux se refrena de ser causa de su intención. La pregunta es: ¿cuál es la razón que puede tener para ello? y, más importante aún, ¿se puede tener esa razón sin formar una intención que la acompañe? Se ha de tener presente que Pollux controla el auto. Que, hasta donde él sabe, está en su total control arrollar o no arrollar al ladrón. Adicionalmente, sabe también que si no hace nada (no gira el volante, no presiona los frenos, etc.), arrollará al ladrón. Supóngase entonces que Pollux tiene la razón  $x$  para refrenarse de formar cualquier intención al respecto. El centro de atención debe estar en la cercanía de esta razón con las consecuencias y su posible conexión con una intención. Si Pollux es un ser racional (como lo exige la teoría de Reid), ha de saber las consecuencias inmediatas de no formar una intención en absoluto (el auto seguirá su rumbo y arrollará al ladrón). Esto es un caso muy diferente de aquel en el que, por ejemplo, un hombre arroja una bolsa de plástico por el drenaje, sin saber que la consecuencia lejana de ello será que un delfín se ahogue con ella y muera. En el caso de Pollux la consecuencia es inmediata, es cercana y es evidente. No se requiere de un gran cálculo ni de una inteligencia extrema para poder verla. Es muy difícil, si no imposible, concebir cómo Pollux puede optar por ser indiferente ante este caso, a sabiendas de las consecuencias, sin tener alguna intención (fuerte, débil, chica, grande, pero una intención al fin y al cabo) de arrollar al ladrón. A esto se lo podría criticar diciendo que es posible desear las consecuencias de algo, sin que haya una intención que acompañe a este proceso. Por ejemplo, se puede desear la paz del mundo, sin que ello implique una intención. Sin embargo,

esta no sería una crítica acertada, pues en ese ejemplo la paz del mundo es una consecuencia lejana y, seguramente, independiente de las acciones de quien la desea. Con facilidad alguien puede pensar que, haga lo que haga, sus acciones no tendrán repercusión para la paz mundial y que, por ello, le es mejor abstenerse de hacer cualquier cosa y esperar a que los líderes políticos del mundo logren la consecuencia que tanto quiere. Pero este ejemplo no es análogo al de Pollux, pues en este último caso la consecuencia es inmediata y *depende* únicamente (hasta donde él sabe) de sus acciones. ¿Qué pasaría si alguien creyera fervientemente que la paz del mundo depende de sus acciones y solo de sus acciones? Si hace algo, la paz se obtiene; si no lo hace, la paz jamás llega. ¿Podría este sujeto, en este caso, no formar ninguna intención y seguir queriendo las consecuencias? La respuesta sería un rotundo *no*. Lo mismo ocurre en el caso de Pollux. Si Pollux es un ser racional, y está convencido de que las consecuencias inmediatas del evento dependen únicamente de sí mismo, no puede ser neutral (en cuanto a intención se refiere) frente a él.

Hay, frente a esta situación, dos posibles escenarios en los que Rowe podría intentar defenderse. Por un lado, podría decir que Pollux se limita a abstenerse de formar la intención, ignorante de las consecuencias. Por ello, sería neutral (en cuanto a intención se refiere) ante la situación. Sin embargo, aquí Rowe sería injusto con la teoría de Reid, pues ¿qué tipo de racionalidad posee Pollux en un caso así? Seguramente una razón tan obtusa no es lo que tenía en mente Reid cuando postulaba como condición necesaria de las causas-agente el ser racionales. De nuevo, no se trata de un caso en donde las consecuencias sean lejanas y difíciles de prever. Todo lo contrario, se trata de un caso en donde son cercanas, inmediatas y evidentes. Un sujeto que actúe sin verlas será o bien poseedor de algo que difícilmente podría ser llamado “razón”, o bien un sujeto que no la ha utilizado en el momento de refrenarse de formar la intención. En ninguno de los dos casos se trata de una causa-agente.

Por el otro lado, Rowe podría decir que Pollux se refrena de formar su intención porque sabe que, finalmente, otro evento ocurrirá que cambiará la situación (ya sea que el auto mismo falle y vire, evitando arrollar al ladrón, o que Cástor intervenga con su dispositivo exculpando entonces a Pollux). Sin embargo, aquí Rowe no estaría siendo justo con los contraejemplos de Frankfurt, pues estos se basan en la idea de que el agente que está a punto de decidir si realiza o no cierta acción es ignorante de la capacidad del agente contra-fáctico de manipular su cerebro, o de cualquier otro evento natural que ocurra y que cambie la situación. Hasta donde sabe Pollux, la situación depende de él y solamente de él. Decir que deja de formar una intención

porque sabe que la situación cambiará es otorgarle un conocimiento que no tiene y que los contraejemplos de Frankfurt no admiten.

Cualquiera que sea la opción que tome Rowe, termina por cometer el mismo error que él le atribuye a Fischer, a saber, no hacer justicia a la teoría. En el caso de Rowe, en una situación no hace justicia a la teoría de Reid, y en la otra no la hace a los contraejemplos de Frankfurt.

Con este argumento se espera haber mostrado que la teoría de la causa-agente solo puede ser utilizada para defender el PPA, si violenta alguna de las dos teorías involucradas en el debate: la de Reid o la de Frankfurt. Y es posible pensar que una estrategia similar se puede utilizar para responder a los defensores de los demás parpadeos de la libertad. Si bien ya no se basaría en los detalles de la causa-agente, sí se basaría en la racionalidad del individuo para mostrar que los parpadeos de libertad que involucran dichas defensas o bien otorgan demasiado conocimiento y racionalidad al sujeto, o bien exigen que este pueda actuar casi sin ella. Pero esto es tema de una investigación diferente. Baste por ahora con dejar sugerida la posible vía para enfrentar las críticas.

*Recibido en julio de 2008; aceptado en noviembre de 2008.*

## BIBLIOGRAFÍA

- CLARKE, R. (1993), "Toward a Credible Agent-Causal Account of Free Will", en Watson, Gary (ed.), *Free Will (Second Edition)*, New York, Oxford University Press, 285-298.
- FISCHER, J.M. (1994), *The Metaphysics of Free Will*, Malden, Blackwell.
- , (2003), "Responsibility and Agent-causation", en Widerker, David & Michael McKenna (eds.), 235-250.
- FRANKFURT, H. (1969), "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, Vol. 66, No. 23, 829-839.
- , (1995), *The Importance of What we Care About*, New York, Cambridge University Press.
- GINET, C. (2003), "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing", en Widerker, David & Michael McKenna (eds.), 75-90.
- KANE, R. (1995), "Two Kinds of Incompatibilism", en O'Connor, Timothy (ed.), *Agents, Causes and Events*, New York, Oxford University Press, 115-150.
- LOCKE, J. (1992), *Ensayo sobre el entendimiento humano*, México, Fondo de Cultura Económica, trad. Edmundo O'Gorman.
- MELE, A. & ROBB, D. (2003), "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases", en Widerker, David & Michael McKenna (eds.), 127-138.
- NYLOR, M. (1984), "Frankfurt on the Principle of Alternate Possibilities", *Philosophical Studies*, No. 46, 249-258.
- O'CONNOR, T. (2003), "Agent Causation", en Watson, Gary (ed.), *Free Will (Second Edition)*, New York, Oxford University Press, 257-284.
- PEREBOOM, D. (2005), "Defending Hard Incompatibilism", *Midwest Studies in Philosophy*, Vol. xxix, 228-247.
- , (2007), "Hard Incompatibilism", en Kane, Robert (et. al.), *Four Views on Free Will*, Malden, Blackwell.
- REID, T. (1983a), *Inquiry and Essays*, Indianapolis, Hackett.

- , (1983b), *The Works of Thomas Reid*, United Kingdom, Georg Olms Verlag.
- ROWE, W. (1991), *Thomas Reid on Freedom and Morality*, Ithaca, Cornell University Press.
- , (2003), “Alternate Possibilities and Reid’s Theory of Agent-Causation”, en Widerker, David & Michael McKenna (eds.), 219-234.
- , (2004). “Thomas Reid’s Theory of Freedom and Responsibility”, en Cuneo, Terence & René Van Woudenberg (eds.), *The Cambridge Companion to Thomas Reid*, New York, Cambridge University Press, 222-242.
- , (2006), “Free Will, Moral Responsibility, and the Problem of «oomph»”, *The Journal of Ethics*, Vol. 10, No. 3, 295-313.
- WIDERKER, D. (2002), “Blameworthiness and Frankfurt’s Argument Against the Principle of Alternative Possibilities”, en Widerker, David & Michael McKenna (eds.), 53-73.
- WIDERKER, D. & MCKENNA, M. (eds.). (2003), *Moral Responsibility and Alternative Possibilities*, United Kingdom. Ashgate.