



Crimen, conexión y castigo: una objeción a la perspectiva retribucionista de la pena de muerte

Crime, Connection and Punishment: an Objection to the Retributionist View of the Death Penalty

DAYRÓN TERÁN¹

Resumen: Este artículo plantea una objeción a la perspectiva retribucionista de la pena de muerte. Si adoptamos un criterio psicológico acerca de lo que es moralmente relevante en términos de autointerés, nuestras intuiciones sobre la retribución y el merecimiento podrían verse seriamente afectadas. La consideración de que la pena de muerte es la justa retribución a cierta clase de delitos parece presuponer en el ejecutado la existencia de una entidad subyacente a todos los cambios subjetivos diacrónicos. Si no hay tal entidad, o si esta no es lo moralmente relevante, el carácter retributivo de la pena podría verse significativamente reducido.

Palabras clave: pena de muerte, retribución, identidad personal, relación R, responsabilidad.

Abstract: This paper raises an objection to the retributive view of the death penalty. If we take a psychological approach about what morally matters, our intuitions on retribution and merit could be significantly modified. The claim that death penalty is the fair retribution to certain kind of crimes presuppose in the executed the existence of an underlying entity which would remain identic to itself troughth all the diachronic subjective changes. If this entity does not exist, or if its continuity is not what matters, the retributive nature of the death penalty could be considerably reduced.

Key words: death penalty, retribution, personal identity, relation R, responsibility.

Introducción

Consideremos el siguiente experimento mental:

La extraña máquina: un malvado científico condenado a muerte por múltiples asesinatos escapa de prisión pocos días antes de su ejecución. Escondido en un remoto laboratorio, pone en marcha un plan maléfico, para el cual desarrolla una extraña máquina: una cabina que modifica la

Recibido: 01/06/2020. Aceptado: 05/01/2021.

1. Universidad de Santiago de Compostela. Investigador predoctoral contratado. Correo electrónico: dayron.teran.pintos@usc.es. Área de estudio: filosofía moral. Líneas de investigación: futuro a largo plazo, ética de poblaciones, identidad personal, ética animal.

personalidad de las personas que entran en ella. Mediante un complejo sistema de ruptura y reelaboración de conexiones psicológicas, la máquina es capaz de crear, en detrimento de conexiones previas, nuevas redes de contenidos mentales coherentes entre sí. El científico puede eliminar de esta forma los rasgos que desee de la personalidad de quien entra a la máquina. Así, podría por ejemplo, modificar preferencias, editar rasgos de carácter y hasta suprimir recuerdos de actos cometidos. La máquina permite introducir en el individuo nuevas creencias, valores y recuerdos, de tal modo que es capaz de convertir a un peligroso asesino en un ciudadano ejemplar, respetuoso de las leyes y preocupado por los demás. El científico utiliza la máquina cuando la policía golpea a su puerta. Todas sus conexiones psicológicas previas son destruidas. Cuando es detenido no recuerda haber cometido ninguno de sus crímenes, se considera una persona pacífica y parece mostrar unas fuertes creencias acerca de la santidad de toda vida humana. Las autoridades se plantean entonces si la condena a muerte debe aplicarse tal y como estaba programada, o si por el contrario, el cambio de carácter sufrido por el condenado puede requerir algún tipo de revisión de la condena. Finalmente, la ejecución se lleva a cabo, tal y como el científico creyó que ocurriría. Él asumía que moriría al ingresar a la máquina, y que quien saldría sería otra persona. De este modo, su colección de muertes tendría una víctima más: el ejecutado.

Aunque el caso descrito aquí es una mera ficción y no existe ninguna máquina semejante en el mundo, la utilización de este experimento mental nos permite sacar a la luz algunas de las intuiciones más frecuentes en lo relativo a la aplicación de la pena de muerte. A lo largo de estas páginas se presentará una objeción contra la justificación retribucionista de tal pena, apoyándonos para ello en un determinado criterio acerca de lo que importa en función de la clase de entidad que somos. Con este fin, se dividirá la exposición en cuatro partes. La primera de ellas analizará cuál es la justificación retributiva para la aplicación de la pena de muerte. La segunda parte expondrá la problemática en torno a la cuestión de la identidad personal y su relación con lo que comúnmente valoramos como importante en términos prudenciales.² En la tercera parte se mostrará cómo es posible tener un criterio para fines autodirigidos sin apelar a la identidad personal. La cuarta parte concluirá que, si la identidad personal no es lo relevante, entonces la pena de muerte puede no ser una forma de retribución adecuada.

2. En el campo de la ética es común la distinción entre fines autodirigidos y fines morales. Los primeros son también llamados “prudenciales”, pues abarcan lo que es importante para uno mismo. Sobre esto, véase: Richardson (2018), Chignell (2018). Cabe señalar que esta distinción no es imprescindible, y que se puede defender que la moral de un individuo es el conjunto de todos sus fines últimos, sean estos autodirigidos o no. Aquí asumiré de manera exclusivamente metodológica esta distinción.

1. La justificación retributiva de la pena de muerte

Aunque las razones que comúnmente se esgrimen para justificar el castigo son varias (hacer justicia, prevenir, disuadir, proteger a la sociedad), podríamos afirmar que estas son fundamentalmente de dos tipos: o bien son de carácter retributivo o bien están orientadas a los efectos que la aplicación del castigo tiene.³ Una razón para castigar es retributiva cuando afirma que el criminal debe pagar por lo que hizo, y que las consideraciones relativas a las consecuencias del castigo no pueden tener más peso que la necesidad de hacer justicia. El castigo sería parte de lo que esa persona merece por sus actos, y no retribuirlos de la manera adecuada sería pues un acto de injusticia. Esto no implica que quienes defiendan un enfoque retribucionista no puedan alegar que este, además, produce las mejores consecuencias. Se podría argumentar que al enviar a prisión a un criminal por un cierto delito, no solo se está aplicando una retribución a sus actos, sino que también se está protegiendo a la sociedad del peligro que supone esta persona. Esto es difícilmente cuestionable en el caso de las penas de prisión, puesto que al mantener a alguien que es un peligro público en la cárcel se está protegiendo al resto de individuos de la sociedad, sean o no retributivas las razones que lo mantienen entre rejas. Sin embargo, no está claro que los retribucionistas puedan apelar a las consecuencias en el caso de la pena de muerte, ya que incluso asumiendo que la muerte fuera una retribución justa en ciertos casos, está lejos de ser un hecho probado el que haya un vínculo real entre la pena capital y una menor incidencia de aquellos delitos que están castigados con ella.⁴ De hecho, hay autores que sostienen que las ejecuciones podrían incrementar la criminalidad (Bowers & Pierce, 1980). Además, si defendemos un enfoque retribucionista puro, entonces sería irrelevante que las consecuencias de la pena de muerte fueran manifiestamente peores. Si creemos que la retribución justa a ciertos crímenes es la muerte y consideramos además que hacer justicia es un deber moral, entonces, incluso en el caso de que la pena de muerte no solo no disuadiera, sino que incitara a diez asesinatos por cada ejecución, tendríamos que seguir defendiendo su aplicación, puesto que sería incorrecto no ejecutar a alguien que lo merece; aunque en términos de consecuencias esto fuera, permaneciendo todo lo demás igual, mucho peor. Quienes defienden posiciones normativas basadas en lo justo o en el deber pueden quedar comprometidos con conclusiones como esta. El que de un acto se sigan consecuencias terribles, no nos eximiría de

3. También se puede defender lo que en ocasiones se denomina *doctrina mixta* (García Amado, 2018). Aquí consideraremos que las posiciones mixtas no son de un tercer tipo, sino que son, o bien retribucionistas o bien consecuencialistas. Pueden ser retribucionistas si forman parte de una concepción deontológica que no desvincule totalmente lo correcto de lo valioso, como ocurre, por ejemplo, con el deontologismo del umbral. Sobre esto véase Alexander (2000). Por otra parte, existen posiciones consecuencialistas que pueden considerar que aplicar una retribución justa a los delitos es parte de lo que hace que las cosas sean mejores. Esto es compatible con la idea de que un mundo injusto es un mundo peor.

4. Sobre esto, puede consultarse: Glover (1990); Demetrio Crespo (2013); Albrecht (2016).

tener que llevarlo a cabo si este constituye un deber moral.⁵ Por otra parte, muchas personas consideran que el compromiso con consecuencias nefastas socava necesariamente el núcleo de la justificación retributiva: la idea de que hacer justicia es siempre hacer lo correcto. Si somos consecuencialistas, es decir, si defendemos que debemos hacer aquello que produzca las mejores consecuencias, entonces podemos sostener que en muchos casos lo correcto puede ser actuar injustamente, si con ello hacemos que las cosas sean mejores. Esto implica que alguien podría sostener que la pena de muerte es un castigo justo, y al mismo tiempo creer que produce consecuencias negativas, y que por tanto debería ser rechazada. De igual modo, es posible sostener que, siendo un castigo excesivo o injusto, produce (al menos en ciertos casos) mejores consecuencias, y que por tanto, es legítima su aplicación. Por otra parte, no todos quienes aceptan un enfoque retribucionista del castigo apoyan la pena de muerte. Tanto el rechazo como la aceptación de la pena son compatibles con posiciones retribucionistas y no retribucionistas.

Este artículo expondrá una razón para rechazar el argumento retribucionista en su defensa, el cual podría tener la siguiente forma:

- (1) Debemos hacer siempre lo que es justo.
- (2) Aplicar una retribución adecuada a los delitos es una parte de hacer lo que es justo.
- (3) La retribución adecuada para los delitos x es la pena de muerte.
- (4) Por tanto, debemos retribuir los delitos x con la pena de muerte.

Hay diferentes formas de poner en cuestión la conclusión de este argumento. Se puede negar la premisa 1, esto es, negar que siempre debemos hacer aquello que es justo. Como antes indicamos, si somos consecuencialistas, no tenemos que comprometernos con esta premisa (a no ser que consideremos que lo justo coincide siempre con lo valioso). Por su parte, quienes asumen la premisa 1 como verdadera no tendrán demasiados problemas en aceptar lo que nos dice la premisa 2, la cual parece bastante plausible en conjunción con la anterior. La premisa 3 es la que se presenta más problemática, puesto que incluso aceptando lo que dicen las premisas 1 y 2, puede ser negada. De hecho, en muchas ocasiones se cuestiona esta premisa aceptando las dos previas. Esto se puede ver, por ejemplo, en aquellas posiciones que sostienen que la pena de muerte debe ser rechazada por cruel o inhumana. En estos casos no se cuestiona que debemos hacer lo que es justo o que castigar con justicia sea una parte de hacer aquello que es justo. Más bien se critica la idea de que la pena capital sea un castigo adecuado.⁶ Sin embargo, estas posiciones se enfrentan al problema de que

5. Sobre la asunción de implicaciones controvertidas en una ética del deber, véase: Kant (1987).

6. Existen diferentes objeciones que se pueden plantear a la pena de muerte en términos de adecuación. Una muy seria es la que apela la imposibilidad de matar varias veces a quien ha cometido más de un crimen cuyo castigo es la pena

estos calificativos dependen en gran medida de juicios de valor. No todos llamamos crueles a las mismas cosas. No todos consideramos inhumanas las mismas prácticas. Además, desde una defensa retribucionista de la pena, se podría alegar perfectamente que es justo tratar con crueldad o de manera inhumana a alguien cruel e inhumano. En las próximas páginas, nos centraremos en una objeción bastante fuerte a la premisa 3, y que es de un tipo diferente a las anteriores. Esta proviene del examen acerca del tipo de entidad que somos, y de cómo se relacionan los conceptos de responsabilidad o castigo con esta clase de entidad.

2. La identidad personal

En muchas ocasiones los debates éticos y jurídicos acerca del merecimiento se centran en el tradicional problema de la libertad de voluntad y el determinismo.⁷ Sin embargo, no menos importante a la hora de dilucidar cuestiones retributivas es el análisis de la identidad personal a través del tiempo. Podemos entender el concepto de identidad en dos sentidos diferentes: por una parte, podemos hablar de identidad cualitativa y, por otra, de identidad numérica (Parfit, 2004, 375). Decimos que hay identidad cualitativa entre dos objetos cuando estos comparten todos aquellos aspectos que se consideran relevantes en su comparación. Derek Parfit pone como ejemplo de este tipo de identidad el caso de dos bolas blancas de billar. Ambas son iguales, y en la medida en que las utilizo para el mismo fin, son indistinguibles la una de la otra. Por otra parte, aunque estas dos bolas sean cualitativamente idénticas, no son numéricamente idénticas, ya que no son una y la misma bola (Parfit, 2004, 375). Si existieran infinitos universos paralelos, con infinitas versiones de mí mismo, tal y como soy en este universo, yo sería cualitativamente idéntico a todas estas versiones, pero no sería numéricamente idéntico a ellas. Esta distinción es relevante, y puede tener implicaciones metafísicas y morales muy significativas. Cuando pensamos en la identidad personal, nuestro sentido común tiende a identificarla con una forma de identidad numérica: seamos lo que seamos, somos una y la misma cosa durante toda nuestra vida. Así, no podría haber nadie más en el multiverso que también sea “yo”, ni nada que no sea numéricamente idéntico a mí podría ocupar “mi futuro”. Esto implicaría que, en términos de autointerés, ningún perjuicio a mi “yo” numérico podría ser compensado por el beneficio de un “yo” cualitativo. Sin embargo, no está claro que nuestra identidad personal responda al concepto de identidad numérica, ni que nuestras relaciones prudenciales se agoten en esta. Como sostiene Parfit, la pérdida de identidad cualitativa bien podría

de muerte. A esto se podría responder, quizás, que existe un umbral a partir del cual las faltas dejan de contar. De este modo, cualquier cantidad de transgresiones más allá del umbral sería irrelevante. Esto, sin embargo, resulta sumamente contraintuitivo e implausible. Para un listado de los problemas que presenta la idea de que la pena es el “justo castigo” para ciertos delitos, véase: Amnistía Internacional (1999).

7. Véase, por ejemplo: Demetrio Crespo. E. (2013).

conllevar la destrucción de la identidad numérica (Parfit, 2004, 375). Esto no parece ocurrir en el caso contrario, ya que si de pronto este universo fuera suprimido y mi identidad numérica destruida, mi identidad cualitativa “sobreviviría” en todos los demás universos en los que exista una versión mía.

Cuando nos preguntamos acerca de la identidad personal, lo que de algún modo nos estamos preguntando es qué clase de entidad somos y bajo qué condiciones permanecemos o somos destruidos. Existen, a grandes rasgos, tres tipos de posiciones en lo relativo a esta cuestión.⁸ En primer lugar, hay quienes creen, como Descartes, que somos entidades que existen separadamente de un cuerpo (Parfit, 2004, 386). Esta posición puede tomar varias formas, algunas de las cuales son compatibles con la idea de que existimos previamente a nuestro organismo y existiremos después de él. Según un enfoque de este tipo, podríamos perfectamente ser almas inmortales encarnadas. El problema fundamental a la hora de defender afirmaciones como esta es la falta de evidencia que las corrobore. Otra posibilidad es sostener que somos entidades físicas. Según una posición de este tipo, podríamos ser nuestro organismo o una parte de él, como nuestro cerebro (Parfit, 2004, 376; Persson, 2005, 239). De este modo, comenzaríamos a existir en algún momento de nuestro periodo de gestación y moriríamos, bien cuando nuestro organismo deje de cumplir sus funciones vitales, o en el caso de que creamos que somos nuestro cerebro, cuando el funcionamiento de este cese. Si defendemos una posición como esta, creemos que de ningún modo podríamos sobrevivir a nuestro cuerpo o a nuestro cerebro. A esto se puede objetar que no está claro que la identidad numérica de un organismo se conserve durante toda su vida. Si es posible que determinados cambios cualitativos muy significativos destruyan la identidad numérica, entonces tiene sentido preguntarnos si los cambios que sufrimos a lo largo de nuestras vidas no serán acaso de este tipo. Se podría responder, sin embargo, que todos estos cambios han de sucederle a una entidad física subyacente, que es la que conservaría la identidad numérica. La cuestión es entonces cuál es esa entidad. Dado que casi todas las células de nuestro cuerpo son reemplazadas en el transcurso de nuestra vida, entonces ¿somos solo aquellas partes de nuestro cuerpo que no sufren tales reemplazos? Esto parece bastante implausible. Además, mucho de lo que consideramos importante en términos prudenciales no coincide con la identidad numérica de nuestro cuerpo o de ciertas partes de él. Siguiendo a Parfit (2004, 373), imaginemos que existe una máquina capaz de producir réplicas exactas (física y psicológicamente) de nosotros. Durante el proceso de replicación, toda nuestra información es copiada y nuestros cuerpos son destruidos y replicados. Cuando la réplica toma conciencia, su último recuerdo es nuestro recuerdo de ingresar a la máquina. Para nuestra réplica, lo que ocurre

8. Existen en realidad múltiples posibilidades de articular criterios de identidad personal. Sin embargo, y siguiendo a Parfit (2004), considero que estas posiciones son, fundamentalmente, de tres tipos.

tras el proceso no es muy diferente a despertar después de una anestesia. Consideremos así el siguiente caso:

Última semana: tengo una enfermedad incurable que me matará en una semana. Cuando ya he perdido toda esperanza, alguien me ofrece la posibilidad de utilizar la máquina de replicación. Sé que si uso la máquina mi réplica no tendrá la enfermedad, y su esperanza de vida será de cincuenta años. Sin embargo, si quiero replicarme, tengo que hacerlo inmediatamente.

De este modo, lo que debo decidir aquí es si es mejor para mí vivir una semana más como mi organismo y luego morir, o por el contrario, realizar una réplica, física y psicológicamente idéntica a mí, que vivirá mucho tiempo y que podrá continuar realizando todos mis proyectos. Tiene mucho sentido considerar que, permaneciendo todo lo demás igual, lo primero no puede ser mejor para mí. Sin embargo, utilizar la máquina implicaría, según el criterio físico de la identidad, morir antes. Cuando elijo la opción de replicarme, parece que lo hago desde la convicción de que así sobrevivo de algún modo. Pero si creo que soy un organismo, la destrucción de mi cuerpo no puede ser un modo de supervivencia.⁹

3. Lo que importa

Cuando elegimos la replicación, parece que asumimos que no somos ni cuerpos ni almas. Podemos aceptar entonces una tercera posición, y considerar que somos flujos continuos de contenidos de conciencia. Esto es lo que defiende el criterio psicológico de la identidad (Parfit, 2004, 383). Un enfoque de este tipo, aun aceptando que la vida mental tenga una base física, puede afirmar que lo relevante para determinar nuestra permanencia a lo largo del tiempo es la *continuidad* de nuestro flujo de conciencia así como el grado de *conexión* entre los diferentes contenidos mentales que lo conforman. Esto es lo que Parfit denominó *relación-R* (2004, 383). Por *conexión* se entiende el grado de relación entre diferentes contenidos psicológicos, mientras que la *continuidad* haría referencia a la superposición de cadenas de contenidos psicológicos *conectados* entre sí (Parfit, 2004, 382). Son modos típicos de *conexión* los recuerdos de experiencias, la relación entre una preferencia y la motivación de llevar a cabo los actos que la satisfacen, o el mantenimiento de las creencias a lo largo del tiempo. Así, presumiblemente yo tendré un alto grado de *conexión* conmigo mismo mañana, y mañana además la tendré conmigo pasado mañana, de tal modo que se crea una de cadena de superposición de *conexiones* que puede durar toda mi vida. Esto significa que puedo ser *continuo* conmigo mismo dentro de treinta años, a pesar de que haya pocas *conexiones* directas entre yo ahora y yo entonces. Si pensamos en los cambios psicológicos que tienen lugar al cabo de

9. Ciertos criterios de tipo psicológico coincidirán con el criterio físico en que realizar la réplica sería para mí tan malo como morir. Sobre esto, véase: McMahan (2002, 57). Contrariamente a esto, Parfit afirmará que “mi relación con mi Réplica contiene casi todo lo que importa” (Parfit, 2004, 509)

un día en la vida de un individuo corriente, vemos que no se trata de modificaciones muy significativas. Al despertar, una persona suele tener las mismas creencias que el día anterior, parece recordar las mismas cosas y puede tener la misma determinación a llevar a cabo sus objetivos previos. No ocurre lo mismo cuando analizamos la relación entre estos elementos en momentos que están muy separados en el tiempo. Alguien puede tener toda una serie de creencias en un momento dado, y creer lo contrario varios años después. Se puede querer que sucedan ciertas cosas a los veinte años y no desearlo a los cuarenta. Así, aunque exista *continuidad* psicológica, la *conexión* entre las diferentes partes de la vida mental es gradual, pudiendo darse incluso en una forma muy débil entre momentos muy distantes en el tiempo. Parfit postula que existe una *conexión fuerte* entre dos momentos dados si las conexiones llegan como mínimo al 50% de las que llegarían con respecto al transcurso de un día cualquiera (2004, 382). Según esta asunción, con toda probabilidad yo estaré *fuertemente conectado* a mí mismo dentro de tres días, pero no a mí mismo dentro de cuarenta años, si es que aún sigo con vida.

Quienes aceptan la *relación-R* como criterio de identidad personal pueden divergir acerca de si esta debe tener la clase correcta de causa, o si cualquier tipo de causa es válida. Quienes acepten lo primero, creerán que, en el caso anterior, yo “no soy” mi réplica, puesto que aunque entre ella y yo haya *continuidad* y *conexión fuerte*, la causa de estas no es la “correcta”, ya que no tenemos el mismo cerebro. Por otra parte, quienes consideren que la identidad personal viene dada por la *relación-R* con cualquier tipo de causa, pueden defender que yo “soy” mi réplica (Parfit, 2004, 468).

Aunque en principio puede parecer que la *relación-R*, en alguna de sus formas, es un criterio plausible de identidad personal, lo cierto es que presenta problemas muy importantes. Si la *continuidad* garantizara la identidad personal, sería imposible que yo fuera *psicológicamente continuo* con alguien que no es “yo”. Parfit utiliza el experimento mental de los recuerdos implantados como argumento en contra de esta idea. Si de algún modo, yo empezara a recordar ciertas cosas que en realidad nunca viví, pero sí otra persona, podría decir que hay continuidad entre ella y yo. Sin embargo, no parece que esto me permita concluir que yo soy esa persona (Parfit, 2004, 402). Otro argumento en contra de aceptar la continuidad como criterio de identidad lo proporcionan los casos de división, en los que la vida mental de un individuo toma una forma ramificada (Parfit, 2004, 454). Volviendo al caso de la máquina de replicación, podemos imaginar que debido a un error en el proceso, la máquina crea dos réplicas y no una. Ambas serían psicológicamente continuas conmigo, sin embargo, si la identidad personal es una forma de identidad numérica, no podemos afirmar que yo “soy” las dos réplicas.

Por su parte, la *conexión fuerte* no se da entre periodos muy distantes en el tiempo, por lo que postularla como criterio de identidad podría llevarnos a conclusiones paradójicas. Imaginemos

que yo estoy *fuertemente conectado* a un yo-2, y que a su vez yo-2 está *fuertemente conectado* a yo-3, pero que debido a la distancia en el tiempo, yo no estoy *fuertemente conectado* a yo-3. Si la *conexión fuerte* fuera el criterio de identidad personal, me vería forzado a afirmar que yo soy la misma persona que yo-2, que yo-2 es la misma persona que yo-3, pero que yo y yo-3 no somos la misma persona. Un criterio de identidad no transitivo no parece plausible (Parfit, 2004, 382).

Así pues, nuestras intuiciones sobre la identidad personal parecen decirnos que, de existir, esta tiene que ser una forma de identidad numérica, ya que aceptar un criterio meramente cualitativo nos llevaría a admitir, por una parte, que podríamos ser más de una entidad al mismo tiempo, y por otra, que nuestra existencia es mucho más breve de lo que solemos creer. Esto se debe a que podríamos ser sincrónicamente idénticos en sentido cualitativo a otras entidades (universos paralelos), pero no podemos ser diacrónicamente idénticos a nosotros mismos en sentido cualitativo, puesto que nuestro grado de *conexión fuerte* no abarca periodos muy amplios de tiempo y nuestras estructuras físicas están cambiando constantemente. Al mismo tiempo, si no tenemos un criterio cualitativo de permanencia, no parece que pueda haber un criterio numérico sólido, ya que como indicamos anteriormente, la destrucción de la identidad cualitativa parece conllevar la destrucción de la identidad numérica.

De las posiciones anteriormente expuestas, únicamente las de tipo cartesiano serían capaces de conjugar la identidad con aquello que importa. Sin embargo no tenemos ninguna evidencia de que seamos entidades que existen separadas de un cuerpo. El criterio físico, por su parte, es un criterio de la identidad poco plausible y no parece captar nuestras intuiciones acerca de lo que importa. Finalmente, el criterio psicológico, aunque no puede ser un criterio de identidad personal, es un criterio plausible de lo que importa.¹⁰ Parece que la razón por la que me preocupa mi futuro no es que seré el mismo cuerpo, ni porque mi alma permanecerá inmutable en él, detrás de todos los cambios. Lo que parece que me importa cuando acepto replicarme es el mantenimiento de mis creencias, deseos, recuerdos, de mi percepción de mí mismo, la realización de mis proyectos, y tiene sentido considerar que todo esto permanece en la replicación, del mismo modo en que permanece de la noche a la mañana cuando despierto. Algunos autores matizarán esta idea, defendiendo que la *conexión* y la *continuidad psicológica* son lo que importa, siempre y cuando tengan la clase correcta de causa, esto es, el mismo cerebro.¹¹ Parfit, por su parte, afirma que lo importante es la *relación-R* con cualquier clase de causa (Parfit, 2004, 506). Cuando se defiende que la causa es relevante, parece que se está presuponiendo la identidad personal, y que se asume

10. Para un argumento en contra de esta posición, véase: Williams (1970)

11. McMahan (2002, 67), por ejemplo, sostiene que la base de nuestras relaciones prudenciales sería la *continuidad* de “la misma” conciencia (*sameness of consciousness*). De este modo, la causa de la *relación-R* se vuelve fundamental.

que, de algún modo, somos esencialmente nuestro cerebro. Consideremos la siguiente variación del caso anterior:

Última década: me diagnostican una extraña e incurable enfermedad, y me dicen que me quedan diez años de vida. Los médicos me ofrecen entonces la posibilidad de recibir un novedoso tratamiento de replicación celular. Sé que si recibo el tratamiento la enfermedad desaparecerá. El proceso de replicación es gradual, y consiste en ir sustituyendo cada determinado tiempo todas las células de ciertas partes de mi cuerpo por células replicadas. Al final del proceso, todas las células de mi cuerpo serían réplicas, sin embargo, durante las distintas fases, células replicadas conformarían mi cuerpo junto con células originales.

Este experimento mental, inspirado en uno muy similar de Parfit (2004, 802), plantea un problema semejante al del barco de Teseo. Aquí, incluso considerando que lo que importa es la *relación-R* con la clase correcta de causa, parece muy difícil negar que es mejor aceptar el tratamiento. Sin embargo, si esto es mejor, parece que es precisamente porque la causa de la *relación-R* es irrelevante. Al aceptar el tratamiento asumimos que podemos tener vínculos prudenciales con una entidad cuyo cerebro ha sido enteramente replicado. Y parece poco plausible defender que existe una diferencia relevante que justificaría aceptar la replicación en *última década* y no en *última semana*.

Así pues, si no apelamos a la existencia de una entidad que exista separadamente, como un *ego* cartesiano, podemos concluir que lo que importa no es la identidad personal, sino la *relación-R* con cualquier clase de causa (Parfit, 2004, 495).¹² No es el hecho de si soy o no “una y la misma persona” en diferentes momentos lo que me debe preocupar, sino en qué grado estoy conectado con mis distintos “yoes”. La *continuidad* y el grado de *conexión psicológica* son lo que tengo que valorar a la hora de considerar en qué medida me pueden afectar determinados eventos en el tiempo. Asumiendo esto, tiene sentido sostener que también son estos factores los relevantes para evaluar mi grado de responsabilidad actual con respecto a los distintos actos cometidos a lo largo de mi vida.

4. Continuidad, conexión y castigo

Si asumimos lo que dice el criterio psicológico acerca de lo que importa, para determinar si el científico del ejemplo inicial tiene algún motivo para temer a la pena de muerte, no debemos preguntarnos si será él a quien ejecuten, sino si entre sus estados de conciencia antes y después de utilizar la máquina se da la *relación-R*. La respuesta a esta pregunta es negativa. Él no estará *R-*

12. En los casos prácticos del mundo real, no encontraremos diferencias relevantes entre aceptar una versión de *R* con la clase correcta de causa o con cualquier causa. Ambas posiciones defenderán lo mismo. Esto se debe a que no existen máquinas de replicación ni artefactos semejantes.

relacionado con la persona que salga de la máquina puesto que todas las conexiones psicológicas se han roto. Por tanto, parece plausible sostener que no tiene nada que temer con respecto al día de la ejecución. Esto, claro está, no implica que el científico no sea afectado por la muerte. Según el criterio psicológico, la muerte ocurre cuando se produce un cese irreversible del flujo de contenidos de conciencia (Horta, 2010, 114), que es precisamente lo que ocurre cuando él entra a la máquina. Aunque la persona que sale es plenamente consciente, no hay *continuidad* entre sus estados de conciencia y los del científico en el momento anterior al uso de la máquina. Podríamos afirmar más bien que se trata de dos flujos de conciencia diferentes y discontinuos, aunque la causa de ambos sea la misma: el mismo cerebro. Si la causa no determina la *relación-R* y esta es lo importante, es irrelevante que el científico y el ejecutado compartan el mismo organismo, esto es, la misma causa, ya que no estarían *R-relacionados*.

Si consideramos que lo que importa es la *relación-R*, podemos pensar, como sostiene McMahan (2002, 170), que nuestros conceptos de interés y afectación deben ser relativos al tiempo. Imaginemos que gano la lotería. Sin embargo, de camino a cobrar el premio sufro un terrible accidente que produce que se destruyan casi todas mis conexiones psicológicas. Como resultado de esto, ya no estoy fuertemente conectado conmigo mismo el día antes de ganar el premio. McMahan (2002, 106) sostiene que el interés en vivir de un individuo depende de la cantidad de valor que una vida contiene así como del grado de unidad psicológica entre los distintos momentos de esa vida. Así, si estimo que mi bienestar en caso de cobrar el premio y no tener el accidente fuera, digamos, de 80, y el grado de conexión psicológica entre “yo antes de ganar la lotería” y “yo después de cobrar el premio” es de tan solo un 10% debido al accidente; entonces mi bienestar esperado con respecto al momento de cobrar el premio será el producto del valor que asigno al beneficio (80) por el grado de conexión que tendré conmigo mismo en el momento de recibirlo (0,10). De este modo, mi beneficio esperado en este caso sería de tan solo 8. Así, vemos que aunque parece que debería estar feliz por el hecho de ser millonario, en realidad no tengo demasiado por lo que alegrarme. Sería mucho mejor para mí si mi perspectiva de bienestar fuera, digamos de 40, y entonces no ganara la lotería, no tuviera el accidente y esto me hiciera seguir conectado al 98% conmigo la mañana siguiente. En ese caso, mi bienestar esperado sería de 39,2. De este modo, parece claro que bajo ciertas circunstancias, un determinado beneficio podría no beneficiarnos en el sentido esperado. De manera semejante, se puede argumentar que al ser alguien castigado por un crimen, puede no estar siendo perjudicado en el sentido esperado.

Si el grado de *conexión psicológica* del científico con el individuo que sale de la máquina es igual a cero, entonces ¿quién es perjudicado por la ejecución? Es posible que no exista una respuesta a esta pregunta en términos de identidad personal. Sin embargo, si atendemos a la

relación-R, en este caso no podemos estar castigando de ninguna manera al científico. Esto no implica que no estemos perjudicando a nadie, ya que alguien es ejecutado. Algunas de las personas que defienden que la pena de muerte es justa podrían sostener que alguien debe pagar por los delitos que han sido cometidos, que estos no pueden quedar impunes, y que la ausencia de *relación-R* no puede ser más importante que la necesidad de retribución. Sin embargo, según la idea intuitiva que tenemos de la retribución, parece que para que un castigo sea una retribución debe cumplirse la siguiente condición:

Para ser una retribución, un castigo debe aplicarse sobre la/s persona/s adecuada/s. Esto significa que, dado un cierto delito d , que requiere un determinado castigo c , c solo funciona como retribución de d si se aplica sobre la/s persona/s responsable/s de d .

De esta forma, vemos que la aplicación de castigo no es condición suficiente para que haya retribución. La cuestión en nuestro experimento mental será entonces, si la relación entre el científico y el ejecutado es de tal modo que haría de la ejecución un acto no retributivo. Si esto fuera así, la pena de muerte en este caso no estaría funcionando como retribución. Si la *continuidad psicológica* y la *conexión* determinan la manera en que somos afectados por los eventos a lo largo de nuestra vida, puede tener sentido afirmar que también depende de estos conceptos la responsabilidad de los actos. Si la responsabilidad viniera dada por la identidad numérica de nuestro organismo, podríamos imaginar un caso en el que el científico, en lugar de destruir su conciencia y crear una nueva, simplemente intercambia su cuerpo con el de otra persona inocente. En este caso, el responsable de los crímenes sería el cuerpo del científico, que albergaría una conciencia totalmente diferente. Esto implicaría que la conciencia del científico, ahora otro cuerpo, sería inocente de todos los crímenes. Esto no parece aceptable. La responsabilidad tiene que ser parte de lo que importa, y por tanto, tiene que estar vinculada al hecho de estar *R-relacionados*. Esto puede verse claramente en casos en los que estamos orgullosos de nuestra responsabilidad en un hecho. Imaginemos que intercambio mi cuerpo con el de un amigo durante un día, y que ese día descubro la clave para una teoría física del todo. Al otro día, al transferir nuevamente mis contenidos de conciencia a mi cuerpo habitual se anuncia que se le concederá el *Nobel de Física* a mi amigo por el descubrimiento. En este caso, parece que tendría razones para afirmar que el responsable de tal descubrimiento soy yo y no mi amigo. Si esto es así, estaría considerando que la responsabilidad depende de la *relación-R* y no de la identidad numérica de un cuerpo. De esto se sigue que, si el científico de nuestro experimento mental es responsable de los crímenes, y no está *R-relacionado* con la persona ejecutada, esta no puede ser responsable de los crímenes. La pena de muerte en este caso no puede ser una retribución.

Aunque este experimento mental presente un caso muy extremo e irreal en lo que a la

pérdida de *conexión* se refiere, plantea cuestiones de relevancia fáctica. Los condenados a muerte en el Estado de Derecho pasan, en la mayor parte de los casos, un tiempo considerable en el corredor de la muerte, y en algunas ocasiones esta situación se puede prolongar durante décadas. Si como antes mencionamos, la *conexión fuerte* no se da entre periodos muy distantes en el tiempo, es probable que los ejecutados estén *débilmente conectados* consigo mismos en el momento en que cometieron el crimen por el que fueron condenados. Esto plantea una objeción bastante seria a la pena de muerte cuando esta se funda en un principio puramente retributivo. Si un individuo no mantiene ninguna *conexión* consigo en el momento del crimen, como en el caso del científico, no puede merecer ser ejecutado, y a decir verdad, ni siquiera merece ser enviado a prisión. Sin embargo, estos casos no suceden en el mundo real. Los ejecutados no parece que dejen de estar *R-relacionados* en ningún momento. A pesar de ello, y al menos en muchos casos, no puede decirse que exista *conexión fuerte* en ellos entre el momento del crimen y el de la ejecución. Si el grado de *conexión* es susceptible de reducirse, se puede pensar que nuestro concepto de responsabilidad debería ser sensible a esta reducción. De este modo, tiene sentido considerar que la retribución también debería ser reducible. Sin embargo, la pena de muerte no es susceptible de aplicarse en un menor grado; supone un daño irreversible que puede ser tremendamente desproporcionado. Cuando se condena a una persona a muerte no se sabe si, en el momento de la ejecución, esta persona estará *débilmente conectada* consigo cuando cometió el crimen por el que es ejecutada. Es concebible pensar que la relación que una persona arrepentida de un crimen grave mantiene con su “yo” pasado puede expresar una diferencia de carácter similar a la que se puede observar entre dos personas diferentes. De hecho, se puede interpretar que parte del significado de las reducciones de condena radica en la posibilidad de que una persona cambie considerablemente.

Se podría alegar, sin embargo, que la responsabilidad no puede ser de grado, y por tanto, que no admite reducción. Si la responsabilidad es una relación de todo o nada, entonces se es responsable o no. Así, incluso si la *relación-R* se diera en un grado muy bajo, podría haber responsabilidad total. De este modo, sería posible decir que el ejecutado en el experimento mental del científico es inocente, pero que en las ejecuciones corrientes, los ejecutados son totalmente responsables de sus crímenes, ya que lo único que los eximiría de responsabilidad sería la ausencia total de *relación-R*. Sin embargo, si nuestro criterio moral no apela a la identidad personal sino a la *relación-R*, la responsabilidad no puede descansar más que a efectos prácticos y discursivos (quizás inevitables) en sujetos de experiencias. En otras palabras, si no existen entidades de tipo cartesiano, difícilmente puedan existir sujetos de experiencias en un sentido estricto. Esta idea parece chocar con todas nuestras intuiciones y con nuestro sentido de la realidad. Parecería un absurdo concebir un mundo en el que no hay personas, sino una multiplicidad de experiencias. Sin embargo, esta es

una de las implicaciones del enfoque de Parfit: la posibilidad de una descripción impersonal del mundo (Parfit, 2004, 489). Desde una descripción semejante, la responsabilidad tiene que descansar pues en otro tipo de entidad. Parece plausible, desde esta perspectiva, decir que son los propios contenidos psicológicos que determinan la comisión de un acto los que son responsables de él.¹³ Simplificando mucho la cuestión, imaginemos que una persona asesina a otra motivada por dos creencias: en primer lugar, cree que el derecho a la vida de otras personas es menos importante que la satisfacción de sus deseos. En segundo lugar, cree que para satisfacer un deseo muy fuerte que tiene no puede evitar matar a una persona. Esto podría suceder por ejemplo si alguien mata a otra persona para robarle con el fin eliminar un fuerte síndrome de abstinencia provocado por la adicción a una droga. Es perfectamente posible que, después de varios años en prisión y una vez desaparecida la adicción a dicha droga, estas dos creencias desaparezcan. Podrían incluso aparecer fuertes creencias en forma de valores morales que apuntaran en la otra dirección, hacia la santidad de la vida por ejemplo. Así pues, es posible que la responsabilidad sea de todo o nada, pero no entendida como responsabilidad de sujetos de experiencias, sino de contenidos psicológicos. Si esto es así, no sería descabellado afirmar que cuando tales contenidos desaparecen, la necesidad de aplicarles retribución también. Esto no implica que todo el sistema penal deje de estar justificado. Existen razones consecuencialistas de peso para pensar que un mundo sin castigos penales sería un mundo peor. Por otra parte, se puede pensar que la flexibilidad de las penas de prisión o las reducciones de condena son de algún modo sensibles en la práctica al darse de la *relación-R*, y que por tanto, cumplen algún tipo de función retributiva.

5. Conclusiones

Podemos concluir que, si asumimos un criterio psicológico de lo que importa en términos prudenciales, se plantea una objeción importante a la pena de muerte, entendida como retribución adecuada a la comisión de ciertos delitos. Esta objeción, aunque es poderosa, no da una razón concluyente en contra de la pena de muerte, puesto que no dice nada en contra de una concepción no retribucionista de la pena. Si nuestro criterio para justificarla fuera la disuasión o la protección de la sociedad, el argumento esgrimido aquí podría ser irrelevante. Esto es debido a que los

13. Esto es defendido, por ejemplo, por Khoury (2013), quien sostiene que lo relevante a la hora de determinar la responsabilidad diacrónica no es la identidad personal sino el grado de conexión psicológica, y más concretamente, la estructura motivacional que posibilita la comisión de un acto. Walker (2020), por su parte, llega a una conclusión semejante al analizar la problemática de la retribución en casos de división como los planteados por Parfit. Walker asume una concepción de la identidad personal basada en la distinción *type/token*. Según esta, si los nazis hubieran creado una copia idéntica de Hitler poco antes de la caída del Tercer Reich, esta copia sería igualmente responsable de todos los crímenes cometidos por el Hitler original. Esto se debe a que, aunque serían diferentes *tokens*, compartirían el mismo *type*. Lo que identificaría aquí al “*type* Hitler” serían sus contenidos psicológicos (Walker, 2020).

enfoques no retribucionistas pueden defender la pena apelando exclusivamente a sus consecuencias. Muchas de las personas que apoyan la pena de muerte creen que esta es capaz de intimidar a individuos que podrían, de otro modo, verse mayormente motivados a cometer crímenes graves.¹⁴

También se podría objetar que nuestro argumento no refuta la justificación retributiva de la pena de muerte, puesto que solo cuestiona aquellos casos en los que la pena se aplica existiendo un grado de *conexión débil*. Esta objeción es válida, sin embargo, no se puede ignorar que la reducción de *conexión psicológica*, de ser un hecho, tiene que darse en los presos que pasan varios años hasta el momento de su ejecución. Pretender eludir este problema agilizando los procesos en los que la pena de muerte es una de las condenas posibles, podría derivar en la práctica en una vulneración de los derechos de los imputados, cuando no directamente en la aplicación de ejecuciones sumarias. En cualquier caso, esto no constituiría necesariamente una objeción a la justificación de la retribución si sus plazos fueran cortos, sino a las consecuencias que un escenario semejante podría implicar.

Todo lo anteriormente expuesto ofrece, no solo una razón para poner en cuestión la premisa 3 del argumento retribucionista en defensa de la pena de muerte, sino que nos obliga a revisar cómo se relaciona nuestra idea de la retribución con el tipo de entidad que somos.

Bibliografía

- Alexander, L. (2000), «Deontology at the treshold», *San Diego Law Review*, 37, pp. 893-912.
- Albrecht, H-J. (2014), «Pena de muerte, efecto disuasorio y formulación de políticas», en L. Arroyo Zapatero, A. Nieto Martín, & W. Schabas (Eds.), M. Muñoz de Morales Romero (Trad.), *Pena de muerte: una pena cruel e inhumana y no especialmente disuasoria* (pp. 55–72). Cuenca: Ediciones de la Universidad de Castilla-La Mancha.
- Amnistía Internacional, (1999), *Error capital: la pena de muerte frente a los derechos humanos*, Madrid: EDAI.
- Bowers, W.J & Pierce, C.L. (1980), «Deterrence or brutalization: what is the effect of executions?», *Crime and Delinquency*, pp. 453-484
- Chignell, A. (2018), «The ethics of beliefs», *Stanford Encyclopedia of Philosophy*. Consultado en línea, el 01/01/2021, en <https://plato.stanford.edu/entries/ethics-belief/#KinNorPruMorEpi>
- Demetrio Crespo, E. (2013), «Compatibilismo humanista», en E. Demetrio Crespo (dir) y M. Maroto Calatayud, (coord), *Neurociencias y derecho penal*, Madrid: Edisofer.
- Demetrio Crespo, E. (2014) «Sobre el presunto efecto intimidante de la pena de muerte». en L. Arroyo Zapatero, A. Nieto Martín, & W. Schabas (Eds.), *Pena de muerte: una pena cruel e inhumana y*

14 Para una defensa de esta visión, véase: Ehrlich. I. (1975). Por otra parte, organizaciones como la ONU (2015) o Amnistía internacional (1999), desde enfoques abolicionistas de la pena, señalan que no hay datos concluyentes que respalden su efecto disuasorio.

no especialmente disuasoria. (73-78) Cuenca: Ediciones de la Universidad de Castilla-La Mancha.

- Ehrlich, I. (1975), «The Deterrent Effect of Capital Punishment: A Question of Life and Death», *American Economic Review*, vol. 65, núm. 3, pp. 398-414.
- García Amado, J. (2018), «Retribución y justificación del castigo penal», en Guillermo Portilla Contreras & Fernando Velásquez (dir), Esther Pomares Cintas & Juan L. Fuentes Osorio (coord), *Un juez para la democracia: libro homenaje a Perfecto Andrés Ibáñez*, Madrid: Dykinson.
- Glover, J. (1990), *Causing death and saving lives*, London: Penguin Books.
- Horta, O. (2010), «Interés en vivir y complejidad psicológica: un criterio transespecífico». *Revista Laguna*, 26, pp. 109-122.
- Kant, I. (1987), «Sobre un presunto derecho a mentir por amor al prójimo», trad. Mario Caimi, en Cuadernos de Ética, Buenos Aires: Universidad de Buenos Aires.
- Khoury, A. C. (2013), «Synchronic and diachronic responsibility», *Springer: Philos Stud*, 165, 735-752.
- McMahan, J. (2002), *The ethics of killing: problems at the margins of life*, New York: Oxford University Press.
- Organización de Naciones Unidas. (2015), «No hay evidencias del efecto disuasorio de la pena de muerte, alerta la ONU». Consultado en línea, el 30/03/2020, en <https://news.un.org/es/story/2015/11/1344091>
- Parfit, D. (2004), *Razones y personas*, Madrid: Antonio Machado Libros.
- Persson, I. (2005), *The retreat of reason*, New York: Oxford University Press.
- Richardson, H. (2018), «Moral reasoning» *Stanford Encyclopedia of Philosophy*. Consultado en línea, el 01/01/2021, en <https://plato.stanford.edu/entries/reasoning-moral/>
- Walker, M. (2020), «Branching is not a bug; it's a feature: personal identity and legal (and moral) responsibility», *Springer: Philosophy and Technology*, 33, pp. 173-190.
- Williams, B. (1970), «The self and the future», *The Philosophical Review*, 79, N°2, pp. 161-180.