

La justificación del castigo, la complejidad de la vida moral y los cambios de paradigma

Leo Zaibert

Union College, New York

lzaibert@gmail.com

DOI: <https://doi.org/10.20318/eunomia.2019.4711>

Le estoy sumamente agradecido tanto a *Eunomía* por organizar este *Book Forum* como a los Profesores Jerónimo Betegón y Daniel Rodríguez Horcajo por participar en el mismo y por sus comentarios sobre mi reciente *Rethinking Punishment* (2018). No sólo son sus comentarios incisivos y profundos sino que son también generosos de una manera especial, que va más allá, me parece, de la civilidad y de los buenos modales. Pues mi libro es en muchos sentidos un libro polémico, en el que intento, como bien dice Betegón, “abrir ventanas que aporten aire fresco” a una discusión que considero más que anquilosada. Y esta es una discusión – la discusión de la justificación del castigo – a la cual mis comentaristas han hecho aportaciones sumamente valiosas (Betegón, 1992 y Rodríguez Horcajo, 2016). De entrada, no puedo menos que reconocer el espíritu con el que me han leído –un espíritu que evidencia un genuino interés por realmente avanzar un debate que a los tres nos parece importante, por no decir importantísimo.

Quizás el punto en el que mis comentaristas y yo coincidimos más plenamente es que no sólo es ésta discusión importante, sino que en cierto sentido mi tesis de que los participantes en ella se encuentran en una extraña posición improductiva está bien encaminada. Considero este acuerdo un logro en sí mismo. Sin embargo, y como ocurre con frecuencia, es cuando uno se enfoca en los desacuerdos que los debates se tornan más interesantes, y es por ello que aquí me enfocaré más en algunos de esos desacuerdos que en aquellos puntos en los que coincidimos. Pero aprovecharé esta oportunidad para aclarar algunos de mis objetivos en este libro a la luz del contexto de mi obra vista de una manera holística. Esta respuesta se divide en tres secciones. Una primera – y muy breve – sección en la que esbozo una genealogía de *Rethinking Punishment*; una segunda sección en la que discuto ciertos detalles del debate tradicional sobre la justificación del castigo que me motivaron a escribir el libro; y una tercera sección en la que intento explicar lo novedoso de mi propuesta. El grueso de mi reacción a Rodríguez Horcajo se encuentra al final de la segunda sección, y el grueso de mi reacción a Betegón se encuentra al final de la tercera sección.

1. Trasfondo

Mi interés en el problema de la justificación del castigo responde a una preocupación que me ha acompañado a lo largo de mi carrera académica, relacionada con nuestras posibles respuestas a lo que percibimos como el mal comportamiento. En cierto sentido mi enfoque es impudicamente filosófico, al menos en el sentido de ser abstracto y general. En el libro discuto el castigo como tal (o la pena como tal: uso ambas expresiones como sinónimos) y no este o aquel castigo, o este o aquel sistema de derecho penal actual. Estoy convencido de que algunas de las conclusiones de mi último libro tienen implicaciones con respecto a sistemas de derecho penal contemporáneos. Pero esto se debe principalmente a que creo que estas conclusiones tendrían implicaciones con respecto a *cualquier* sistema de derecho penal, incluso aquellos que fuesen enteramente justos y funcionen bien. Pero el grueso de mi aproximación ha sido siempre bastante teórico. Betegón y Rodríguez Horcajo no están solos – ni equivocados – al expresar cierta insatisfacción con lo poco que hay en mi libro relativo a respuestas concretas a problemas concretos del derecho penal – una insatisfacción a la que regresaré más adelante – pero trataré de explicar por qué aun concediendo este punto, me decidí por un proyecto tan aparentemente distante de algunos problemas concretos.

Evidentemente, son muchas las respuestas que podemos dar a lo que percibimos como mal comportamiento (dejo de un lado si nuestra apreciación es correcta o no). Entre las respuestas filosóficamente más interesantes encontramos el reproche, el castigo y el perdón. En mi primer libro, *Five Ways Patricia Can Kill Her Husband: A Theory of Intentionality and Blame* (2005), traté de analizar qué sucede exactamente en nuestras mentes cuando reprochamos a alguien. A veces nuestro reproche es totalmente privado – secreto, incluso – y muchas de esas veces por buenas razones. Pero en otras ocasiones comunicamos este reproche. Y en aun otras ocasiones decidimos castigar a las personas que hacen lo que consideramos reprochable. Por supuesto, no tenemos que castigar todo lo que consideramos reprochable, ya que podríamos ignorarlo o perdonarlo, y es de hecho mucho mejor que no lo hagamos siempre, pues la vida sería verdaderamente invivible si la dedicáramos a castigar todo lo que consideramos reprochable. Pero a lo largo de mi carrera me ha parecido obvio que, conceptualmente, sólo podemos genuinamente castigar aquello que consideramos reprochable (y todavía me sorprende que este punto sea polémico). En cualquier caso, si bien el castigo no es la *única* respuesta posible frente a los actos reprochables, es la más peligrosa de esas respuestas. Y así, en mi segundo libro, *Punishment and Retribution* (2006), investigué el castigo en particular.

Los filósofos a menudo distinguen entre investigaciones conceptuales y normativas. En términos generales, las investigaciones conceptuales preguntan acerca de lo que las cosas son y las investigaciones normativas preguntan acerca de lo que convierte a las cosas en obligatorias, justificadas o buenas. Mi plan original para *Punishment and Retribution* era que fuese una investigación fundamentalmente normativa: quería escribir sobre lo que *justifica el* castigo. Sin embargo, a menudo, uno puede *querer* abordar las preguntas normativas solo para verlas pospuestas, a veces considerablemente, por preguntas puramente conceptuales. Eso es exactamente lo que me sucedió en el proceso de escribir aquel libro anterior: me quedé atascado en asuntos conceptuales sorprendentemente difíciles y sorprendentemente contenciosos, y no pude abordar los asuntos justificativos con la profundidad que merecían. De manera ingenua, pensé que al terminar el libro conceptual, sería capaz de escribir rápidamente aquel libro normativo. Pero me tomó más de una década y docenas de publicaciones adicionales, antes de que lograra

articular mi posición normativa sobre el castigo como un todo coherente en mi libro más reciente, *Rethinking Punishment*.

Para mí no pequeña sorpresa, en este último libro me encuentro a mí mismo rechazando algunas presuposiciones fundamentales de la ortodoxia con respecto a la justificación del castigo – incluidas algunas posiciones que yo mismo he sostenido en el pasado. Es por ello que quisiera esbozar un poco el estado de la literatura especializada sobre la justificación del castigo y sobre mis razones para rechazar tanto de lo que tiene que ofrecer, antes de pasar a dibujar, con trazos muy generales, los contornos de lo que propongo en su lugar.

2. El debate clásico sobre la justificación del castigo

Independientemente de cualquier otra cosa que con el castigo se persiga (y hay potencialmente muchas), no me cabe duda que, por definición, el castigo persigue infligir sufrimiento (o dolor, o privación, o alguna consecuencia no deseada, etc.) a un malhechor. Que esto sea parte de lo que castigo es muestra de inmediato por qué el castigo necesita una justificación moral. En principio, es malo causar sufrimiento intencionalmente. Y, en principio, las cosas malas reclaman una justificación moral. Pero apenas afirmamos esto, dos tipos de escepticismo pueden surgir. Primero, el escepticismo de quienes creen que el castigo no es realmente una cuestión de hacer sufrir a los malhechores y, segundo, el escepticismo de los llamados “abolicionistas”, quienes creen que no existe una justificación moral para el castigo (y que por ello debe ser abolido). Estos tipos de escepticismo no deben preocuparnos demasiado.

Existe un consenso entre los teóricos del castigo con respecto al hecho de que, como lo expresó H.L.A. Hart (2008), el castigo “debe implicar dolor u otras consecuencias que normalmente se consideran desagradables”¹. En este punto estoy de acuerdo con la ortodoxia de la justificación del castigo. “Castigar” a un malhechor dándole algo que sabemos que quiere, es, en el caso estándar, una tontería. Este “dolor y otras consecuencias que normalmente se consideran desagradables” son lo que intento capturar con el término “sufrimiento” en mi libro. Sin embargo, los escépticos en el primer grupo sugieren que el castigo por el sufrimiento puede ser incidental, y que el castigo como tal no tiene por qué ser *necesariamente* o *conceptualmente* doloroso. Creo que este es un ejercicio de autoengaño, que busca endulzar lo que realmente estamos haciendo cuando castigamos. Si el castigo no buscara infligir sufrimiento, sería francamente difícil entender por qué los humanos habrían con tanto ahínco y por tanto tiempo tratado de encontrar una manera de justificarlo.

Los escépticos en el segundo grupo no niegan que el castigo busque infligir sufrimiento, pero afirman que el castigo (en gran medida precisamente por el hecho mismo de que *no* causa sufrimiento) es tan malo que simplemente no se puede justificar. Encuentro esta posibilidad completamente ilusoria. Esto se debe en parte a que no estoy de acuerdo con otra opinión común en la ortodoxia de la justificación del castigo según la cual el castigo es, paradigmáticamente (o para algunos incluso exclusivamente), el castigo penal infligido por el estado. No niego, por supuesto, que el estado castigue; de hecho algunos estados castigan demasiado y de maneras no solo injustas, sino terriblemente crueles. Pero las personas comunes y corrientes también se castigan entre sí en el contexto de sus relaciones interpersonales y actividades cotidianas. El castigo no-estatal es tan antiguo como la humanidad misma; una vida sin castigo es a duras penas imaginable, y seguramente no luce como una vida de seres humanos. Sin duda el estado debería castigar menos; sin

¹ Todas las traducciones son mías.

duda nosotros como individuos deberíamos castigar menos. Pero un mundo completamente carente de castigo me parece una imposibilidad tanto política como psicológica.

Por lo tanto, estoy de acuerdo con la gran mayoría de los filósofos en el rechazo de los escépticos: el castigo es un acto de infligir sufrimiento y el mismo *puede* a veces ser justificado. Por supuesto, la pregunta crucial es “¿Cómo?”. Con respecto a la manera de contestar esta pregunta difiero de la gran mayoría de los filósofos – al menos en el contexto angloamericano – en cuanto ellos creen que la única forma en que se puede justificar el castigo es asegurándose de que su imposición impida un mayor sufrimiento. El castigo es, de acuerdo a esta muy generalizada opinión, una herramienta, un medicamento o una tecnología cuyo único objetivo es minimizar el sufrimiento en el mundo.

De lejos, la formulación más famosa e influyente de esta visión es la del padre del utilitarismo moderno, Jeremy Bentham, a quien cito: “El castigo es en sí mismo un mal”, que solo puede justificarse si “promete excluir un mal mayor” (1962, p. 83). El único objetivo utilitario en la vida, tanto a nivel personal como político, es maximizar la felicidad / placer / bienestar / utilidad, o minimizar el sufrimiento / dolor / miseria / des-utilidad. (En este contexto, maximizar la felicidad y minimizar el sufrimiento son formas diferentes de lograr un mismo objetivo.)

Aunque objetar al utilitarismo clásico es uno de los pasatiempos favoritos de los filósofos contemporáneos, en versiones aparentemente más sofisticadas que la de Bentham el utilitarismo sigue siendo la doctrina moral dominante en el mundo de habla inglesa. La famosa broma de Nietzsche: “La humanidad no lucha por la felicidad; solo el inglés hace eso” es – ahora con la inclusión de “el estadounidense” – tan pertinente hoy como cuando Nietzsche acuñó la frase (2005, p.157).

En ningún otro contexto la fijación angloamericana con el utilitarismo y, en particular, con su más cruda versión benthamita, es más obvia que en el contexto del castigo.² Críticos penetrantes del utilitarismo *en general* sin darse cuenta sucumben a él cuando se trata de sus puntos de vista específicos sobre el castigo. Esto es, en mi opinión, cierto de filósofos extraordinariamente influyentes como Martha Nussbaum, Derek Parfit, John Rawls, o Tim Scanlon. No puedo ofrecer aquí un diagnóstico completo de sus extrañas inconsistencias, pero en el libro sugiero que la razón principal de las mismas es que estos autores reaccionan de forma exagerada ante los excesos del sistema de justicia penal (norteamericano, en particular) y que por ello se aferran a cierta benevolencia que erróneamente asumen es esencial al utilitarismo.

Dudo mucho que estos autores reconozcan que son utilitaristas con respecto al castigo. Y es importante explicar por qué es ajustado insistir que sí lo son. Todos estos autores terminan defendiendo lo que se conoce como justificaciones “consecuencialistas” del castigo. Entre estas famosas justificaciones se encuentran la prevención, la rehabilitación, la incapacitación, la disuasión y la comunicación. La *estructura* de todas estas justificaciones es esencialmente la misma: cualquier

² De hecho, esta fijación utilitarista trasciende el ámbito de los teóricos del castigo angloamericanos. Considérese el diagnóstico de Juan Antonio García Amado: “A la mayoría de los penalistas españoles se les eriza la piel cuando oyen hablar de justificaciones retributivas [vale decir: no utilitaristas – un punto al que regresaré de inmediato] de la pena. Posiblemente esa especie de prejuicio proviene de la muy potente e influyente dogmática alemana, donde retribución penal es asociada a venganza primitiva o ley del talión” (en <https://almacenederecho.org/sin-retribucion-no-hay-pena-justa/>, visitado en Marzo 7 2019). O sea, no sólo Inglaterra, no sólo los Estados Unidos, sino también España y Alemania, etc.

sufrimiento particular causado por la prevención, la rehabilitación, la incapacitación, la disuasión o la comunicación se justificaría si se lograra reducir el sufrimiento en general.

Si bien existen versiones variadas y sofisticadas, en el fondo todas estas justificaciones se basan, en última instancia (y a veces a pesar de que sus proponentes lo nieguen vehementemente), en el utilitarismo benthamita. Considérense las palabras de Parfit en *On What Matters*, posiblemente el libro de filosofía más célebre de esta generación en el contexto angloamericano: “Aunque creemos que las personas inocentes no merecen ser castigadas, también creemos que las personas culpables no merecen ser castigadas. En nuestra opinión, todo castigo es malo en sí mismo” (2011, p. 651). Si esto no es una reducción al absurdo del concepto mismo de “merecimiento”, no sé qué es. Pero más crucialmente: si esto no es una reafirmación de la posición benthamita, no sé qué es.

Independientemente de las inconsistencias internas en la obra de estos autores, no me queda duda que disminuir el sufrimiento es, en principio, algo bueno, por lo que podría no ser suficientemente aparente qué es lo que es criticable (si es que acaso algo) acerca de las justificaciones consecuencialistas del castigo. Recuérdese que la objeción más famosa al utilitarismo en general está relacionada con el castigo: el utilitarista puede justificar a sabiendas “castigar” a una persona inocente (es decir, acusarla por un crimen que sabemos que no cometió), si hacerlo puede minimizar la cantidad total de sufrimiento en el mundo. En líneas similares, los consecuencialistas tienen dificultades para articular la importancia de la proporcionalidad. Tomemos la disuasión, por ejemplo: si se supone que el castigo debe servir para disuadir la comisión de cierto tipo de delito, ¿por qué no estipular castigos muy severos, desproporcionadamente severos, para disuadir la comisión de aquellos? Digamos: la pena de muerte por conducir a exceso de velocidad.

Si el ejemplo luce exagerado, entonces cambiémoslo por un caso real: secuestrar a los hijos de inmigrantes ilegales, una práctica que el actual gobierno norteamericano ha tratado explícitamente de justificar basándose en su supuesto poder disuasivo. Pero es importante enfatizar que el primer ejemplo no es realmente inverosímil, ya que no existe un *principio* que le impida al consecuencialista respaldar la pena de muerte por conducir a exceso de velocidad, si dicho castigo minimizara el sufrimiento general. A algunos nos resulta abominable castigar a los inmigrantes ilegales separándolos de sus hijos (potencialmente para siempre), y no encontraríamos este estado de cosas menos abominable incluso si resultara que al infligir un castigo tan desproporcionado como este estaríamos en efecto disminuyendo el sufrimiento general en el mundo.

No es exagerado decir que los consecuencialistas se saltan a la torera precisamente aquello que es distintivo del castigo y su justificación: la *justicia*. Esto es confirmado por el hecho de que si los consecuencialistas pudieran lograr cualquiera de sus (usualmente loables) objetivos de reducción del sufrimiento sin causar ningún sufrimiento, preferirían hacerlo.

Los pensadores anti-consecuencialistas creen que lo que justifica el castigo es que el mismo es merecido. Es justo, y valioso, que la estudiante que ha escrito un excelente artículo obtenga una calificación excelente: se la *merece*. Y así como aquello es justo y valioso, también lo es que quien hace algo malo reciba el castigo que merece. Y es precisamente en torno a la noción del merecimiento que gira la principal alternativa al consecuencialismo: el retribucionismo. Los retribucionistas no niegan que algunas de las posibles consecuencias del castigo en las que se centran los consecuencialistas (rehabilitación, disuasión, reconciliación, etc.) sean valiosas, o

incluso muy valiosas. Pero insisten en que la raíz de lo que justifica al castigo (y que evita los chivos expiatorios, el castigo desproporcionado y otros excesos utilitarios) no es ninguna de esas cosas buenas, sino el mero hecho de que es merecido. Dar a las personas lo que merecen es, en sí mismo, justo y valioso. Si resulta que el castigo merecido en este o aquel caso *también* rehabilita o disuade, estos resultados serían una suerte de guindas en la torta o, como los ha calificado Michael Moore, “superávits felices” (Moore, 1992).

Los consecuencialistas, por supuesto, plantean una serie de objeciones contra el retribucionismo que, como es habitual, pueden dividirse en normativas y conceptuales. Los críticos normativos afirman que el retribucionismo es bárbaro, ya que, en su opinión, puede justificar castigos verdaderamente horribles. Los críticos conceptuales afirman que el retribucionismo es una cuestión de “pensamiento mágico”, ya que afirman que no hay forma de discernir qué es exactamente lo que la gente merece. Me parece que Rodríguez Horcajo comparte algunas de estas críticas conceptuales con respecto al retribucionismo que percibe en mí. (En contraste, Betegón comparte mucha de mi simpatía, o al menos respeto, por el retributivismo) Antes de ahondar en estas críticas conceptuales, me gustaría decir algo sobre las críticas normativas.

Es cierto que el término “retribucionismo” y sus afines han sido apropiados por políticos, particularmente por aquellos que libran esa usualmente absurda e injusta (al menos en el contexto de los Estados Unidos) “guerra contra el crimen”, por lo que tal vez uno podría encontrar una explicación del atractivo de esta objeción allí. Pero hay muy poca conexión entre el retribucionismo político y el retribucionismo filosófico. Entre quienes atacan al retribucionismo filosófico propiamente dicho, la objeción es casi siempre el resultado de confundir el retribucionismo con otras posiciones, sobre todo con la famosa ley del talión. Esta confusión es un error generalizado, y se puede remontar al hecho de que tal vez el retribucionista más famoso de todos los tiempos, Immanuel Kant, conspicuamente cometió este error. La razón principal por la que esta objeción fracasa es que ningún retribucionista contemporáneo recomienda violar a los violadores o malversar a los malversadores, etc. El retribucionista no está comprometido con ningún programa de castigo en particular, sino sólo con honrar la importancia del merecimiento en la justificación del castigo. De hecho, un retribucionista puede estar, y con frecuencia está, radicalmente opuesto al tipo de punitivismo excesivo con el que a menudo (pero indebidamente) se le asocia.

Los críticos del sistema disfuncional de justicia penal estadounidense, por ejemplo, cuando atacan al retribucionismo pierden de vista cual habría de ser su verdadero objetivo. Al socavar la importancia del merecimiento, se privan de un poderoso argumento *en contra* del sistema incuestionablemente disfuncional de la justicia penal en los Estados Unidos: que impone castigos inmerecidos.

Finalmente, estos críticos objetan que para el retribucionista el hecho de que un cierto castigo sea merecido significa que *debe* ser infligido. Es cierto que esta es una posición retribucionista común y, respecto a ella, los críticos tienen razón. Desafortunadamente para estos críticos, esta no es la única posición a la que el retribucionista tiene acceso. Pero aun para estos críticos, esta no es una objeción que puedan hacer de buena fe, como explico en mi libro y como resumiré en breve.

En respuesta a la crítica conceptual del retribucionismo, debería ser suficiente señalar que los supuestos misterios del merecimiento tienden a ser exagerados. Con frecuencia sí que sabemos lo que las personas merecen, y sin mayor problema aplicamos este conocimiento cada vez que calificamos un artículo de un alumno. El hecho innegable de que hay casos difíciles no niega la existencia de casos

perfectamente claros. No veo nada mágico, ni siquiera misterioso, sobre esto. Además, incluso en los casos en que los juicios de merecimiento muy precisos puedan ser difíciles, a veces los *rankings* son suficientes. Por ejemplo, es fácil ver que, *mutatis mutandis*, la violación y el asesinato merecen castigos más severos que el exceso de velocidad o el fumar alguna droga ilegal, incluso si, tal vez, no pudiéramos asignar cantidades cardinales exactas de merecimiento a cada uno de estos actos. Empleamos estas distinciones cotidianamente y negarlas me parece pura deshonestidad intelectual.

Desarrollando de la teoría de los todos orgánicos de G. E. Moore y el particularismo moral de Jonathan Dancy, sugiero el contextualismo. Los contextualistas reconocen que las cosas pueden tener un valor o desvalor *inicial* (en el sentido de un valor *asumido de entrada*), pero que puede cambiar según el contexto. El sufrimiento es, en principio (en este sentido *inicial, asumido de entrada*), algo malo, pero no es malo, o no es tan malo, cuando, por ejemplo, es merecido. Esto constituiría “pensamiento mágico” sólo si reconocer que el contexto es importante fuese pensamiento mágico. No hay nada de pensamiento mágico en afirmar que no está en lo absoluto claro que el deseo de Joaquín Sabina de “que los que matan se mueran de miedo” constituya una defensa de la pena de muerte: el contexto importa. Desde *El libro de Job* hasta Leibniz, y desde Nietzsche a Weber, la importancia del contexto rara vez se ha puesto en duda. La pregunta acerca de por qué el sufrimiento existe siempre ha estado más ligada al sufrimiento *inmerecido* que al sufrimiento como tal (Zaibert, 2018, pp. 4-12). Pensadores y textos de todas las tendencias y persuasiones han negado, a lo largo de la historia, que la valencia moral del sufrimiento sea invariante, y es demasiado improbable suponer que todos estos autores y fuentes han sido presas del pensamiento mágico.

Pero ahí es donde se encuentra el debate y donde se ha mantenido durante siglos: encontrando el retribucionismo desagradable, los consecuencialistas se aferran a su posición; y viceversa. Y este es el punto muerto que he tratado de resolver en mi último libro. Al comienzo del libro me he referido a este punto muerto como “*a kind of stalemate*”. Curiosamente, Rodríguez Horcajo se ha referido a esta situación con otro símil ajedrecístico: como una situación de “tablas”. Pero hay una diferencia, sutil, entre nuestras apreciaciones: Rodríguez Horcajo considera que ninguna de la justificaciones de la pena “puede ser declarada vencedora”, mientras que desde mi punto de vista el énfasis ha de ponerse en lo estéril de continuar el debate en los términos tradicionales –aun si alguna postura pudiese resultar “ganadora” en algún sentido. De hecho, con frecuencia me he referido a la situación como una de *Zugswang*, un término que pertenece a la teoría de los juegos de manera más general y no sólo al ajedrez, y que creo quizás captura de manera más nítida lo *lamentable* del estado de cosas, o, como Betegón describe mi diagnóstico: una situación de “estancamiento intelectual”, que ha “añadido más calor que luz al debate”. La diferencia sutil es que en el contexto del ajedrez, a veces las tablas pueden ser muy meritorias, y no lamentables en lo más mínimo.

Dejando las sutilezas acerca de diferentes tipos de estancamiento un tanto a un lado, el hecho de que Rodríguez Horcajo esté de acuerdo en que el debate sobre la justificación del castigo se encuentre estancado me resulta algo extraño. Al fin y al cabo, pareciera que para él las justificaciones retribucionistas de la pena son inaceptables. Y si esto es en efecto así, resulta difícil entender cómo puede Rodríguez Horcajo considerar que el debate ha quedado en “tablas” – pues uno de los contrincantes (el retribucionismo) ha, en su opinión, perdido ya.

Es menester matizar: Rodríguez Horcajo podría quizás aceptar algo como el concepto de retribución en su aproximación a la justificación de la pena, pero como él

mismo reconoce, ello sería sólo de una manera indirecta. Sería simplemente consecuencia de admitir que la idea de retribución es parte de nuestra psicología, incluso de nuestra psicología ingenua – pero en ningún caso sería cuestión de admitir el retribucionismo en un sentido robusto, predicado a su vez en el valor real y directo de dar a cada quien lo que merece. Correctamente, Rodríguez Horcajo reconoce que este no es *mi* sentido de la retribución. Pero lo que no me queda claro es si reconoce Rodríguez Horcajo también que este tampoco es el sentido de retribución habitualmente invocado por los retribucionistas. Rodríguez Horcajo rechaza el retribucionismo en sentido robusto, y en *este* sentido su posición central es que el retribucionismo ha sido derrotado, no alcanzado “tablas”.

Ahora bien, la mayoría de las razones que en su recensión Rodríguez Horcajo esgrime para defender esta posición son, precisamente, algunas de las específicas críticas al retribucionismo que discuto en el libro (y que he esbozado en los párrafos anteriores): que constituye una especie de alquimia moral, que es una forma de pensamiento mágico, y que quizás su imposibilidad radique en el mero hecho de que el sufrimiento nunca puede ser merecido. Para mayor ahondamiento a lo ya dicho, agregó lo siguiente.

La supuesta (y supuestamente problemática) alquimia moral del retribucionismo exhibe una suerte de doble vida. El tipo de razonamiento que subyace al retribucionismo es absolutamente familiar e indispensable en nuestras vidas: la noción del merecimiento juega papeles evidentes en un inmenso número de aspectos de nuestra cotidianidad, desde nuestra apreciación del arte, hasta los deportes y la gente misma en todo tipo de contextos. Es sólo cuando el merecimiento es movilizado a los específicos fines de justificar castigos que sus críticos se acuerdan de esta supuestamente problemática alquimia. No conozco de objeciones al hecho de que, por ejemplo, Leonard Cohen haya recibido el premio Príncipe de Asturias en Literatura basadas en que la idea mera del merecimiento sea misteriosa o de que sea el fruto de pensamiento mágico. Por supuesto, habrá quienes quizás piensen que Cohen no merecía este premio – pero, por definición, estas personas no estarían cuestionando la coherencia de la noción misma del merecimiento, sino más bien invocándola. Esto también explica, creo, lo cuesta arriba de la idea, que Rodríguez Horcajo parece acariciar, de que el sufrimiento (en el sentido explicado anteriormente) simplemente no puede ser merecido. En el libro alego que la manera en la que Parfit defiende esta posición es problemática, por una variedad de razones. Mencionaré aquí solo una: Parfit no es un escéptico con respecto al merecimiento en general, sino sólo cuando lo que supuestamente se merece es el sufrimiento. Y esto hace que la carga de la prueba argumentativa recaiga en él. No estoy seguro de si el escepticismo de Rodríguez Horcajo es acomodaticio como el de Parfit. Si así fuere, pues la carga de la prueba también recaería sobre él; y si no, su problema sería otro: pues de ser su escepticismo *global*, abarcando a la idea mera de merecer en sí, gran parte de la discusión moral parecería amenazada por posturas nihilistas, o al menos radicalmente reductivas.

Rodríguez Horcajo afirma: “me parece dudoso, en lo referente a la pena, que un Estado moderno (aconfesional y respetuoso con la autonomía de los ciudadanos) pueda erigirse como el encargado de realizar el valor moral que se esconde tras la imposición del castigo merecido y de su perdón”. Este es un punto importante, y con respecto al cual estoy más de acuerdo con Rodríguez Horcajo de lo que podría parecer a simple vista. No cabe duda que desde la perspectiva de la filosofía política, la pregunta acerca de cuáles son las legítimas atribuciones del estado es absolutamente crucial. Pero en mi libro hay relativamente poco sobre filosofía política. Quisiera subrayar, en cualquier caso, que mi posición explícita es que el hecho de que el castigo sea merecido, o incluso de que infligir un castigo merecido sea en sí

mismo valioso, no implica que exista una obligación de imponerlo (por parte del estado o de cualquier otro ente). Es decir, mis posiciones con respecto al castigo no me obligan a defender un estado (u otro ente) más o menos interventor, o más o menos respetuoso de la autonomía individual u otros principios liberales y laicos.

Pero tampoco niego el hecho de que un estado que sea *totalmente* indiferente a la idea de justicia me resulte francamente deplorable. Y la idea de la justicia en el contexto del castigo está íntimamente ligada a la idea del merecimiento. No es insignificante que el término para el área que nos ocupa, en inglés, sea el de “retributive justice” o “justicia retributiva”, ámbito este diferente al de, por ejemplo, la “justicia distributiva”. Discutir aspectos relativos a la “justicia retributiva” no implica ser un retribucionista (pues un consecuencialista también está discutiendo “justicia retributiva”) – pero la proximidad lingüística revela *algo* acerca de la centralidad del merecimiento con respecto a este tipo de justicia.³

3. Un cambio de paradigma

A pesar de las posiciones anti-retribucionistas de Rodríguez Horcajo, y dado que en muchos casos las acciones (y los castigos) que tanto él como yo recomendaríamos serían parecidos, él afirma que nuestras diferencias “son meras discusiones teóricas con poca incidencia real”. Sin embargo, de inmediato agrega “aunque entiendo que para Zaibert esto no es ni mucho menos así de simple”. Encuentro significativo que existan, en efecto, tantos puntos de coincidencia entre las *políticas* específicas que probablemente Rodríguez Horcajo y yo – un crítico del retribucionismo y un defensor del mismo – recomendaríamos. Esta especie de convergencia, en el ámbito de lo deontológico, sugiere la existencia de otros valores e ideales que pueden ser compartidos por pensadores que en otros aspectos estén en desacuerdo. Pero quisiera explicar las razones por las cuales, en efecto, esto no me parece tan simple (independientemente de mi enfoque “impudicamente filosófico” mencionado al comienzo). Esta explicación me lleva necesariamente a esbozar por qué pienso que mi aproximación al castigo constituye – o al menos pretende constituir – una suerte de cambio de paradigma.

Durante gran parte de mi carrera he criticado al utilitarismo en general, y a las justificaciones consecuencialistas del castigo en particular. No cabe duda que *Rethinking Punishment* contiene renovadas críticas de este tenor. Podría parecer entonces que mi estrategia estaría clara: debería defender el retribucionismo (como he hecho en el pasado). El problema es que mi relación con el retribucionismo se ha tornado complicada. Si bien sigo creyendo que el hecho de que un cierto castigo sea merecido deber ser *parte* de la historia de su justificación (y que un estado o persona que no tomase en cuenta esta parte sería criticable, como acabamos de ver), también me he desilusionado del retribucionismo, pues he llegado a la conclusión de que a pesar de sus aparentes diferencias, el consecuencialismo y el retribucionismo comparten bastante.

Las *estructuras* del consecuencialismo y el retribucionismo son mucho más similares de lo que los teóricos del castigo – yo mismo incluido – han solido reconocer. Sea lo que sea lo que produce la justificación del castigo (la disminución del sufrimiento si eres un consecuencialista, o la imposición del sufrimiento merecido si eres retribucionista), una vez que esto se logra, habríamos terminado, y todos los problemas habrían sido supuestamente resueltos. Desde la perspectiva consecuencialista, si castigar a una persona disminuye el sufrimiento en el mundo,

³ “Retributivo” y “retribucionista” se expresarían en inglés como “retributive”, lo cual intensifica la proximidad lingüística a la que aludo.

entonces alguien tiene la obligación de infligir este castigo, y esta persona estaría justificada en infligirlo. Desde la perspectiva retribucionista, si alguien merece un cierto castigo, entonces alguien tiene la obligación de infligir este castigo, y ella estaría justificada en infligirlo. Todo muy cuadrulado.

Bernard Williams, el autor más influyente en mi último libro, llamó a la visión de la moralidad “*obligations-out, obligations-in*”, una visión que él demostró ser demasiado mecanicista para su propio bien y, de hecho, más bien perteneciente al mundo de la contaduría que al de la moral, pues está bastante desconectada de la complejidad del mundo moral como lo experimentamos realmente (Williams, 2006, p. 181). El hecho de que algo sea *valioso*, bien porque es merecido o bien porque reduce el sufrimiento, no es suficiente para generar la *obligación* de crearlo o promoverlo. Esta fue precisamente la crítica consecuencialista del retribucionismo que concedí en la sección anterior. El problema para el consecuencialista es que él mismo comete este error: la única diferencia entre él y el retribucionista a quien ataca es que él encuentra un valor generador de obligaciones sólo en la disminución del sufrimiento, mientras que el retribucionista también lo encuentra en la imposición de sufrimiento merecido.

Obsesionados con la pregunta de *cómo actuar*, las teorías del castigo de todas las tendencias han olvidado la cuestión de *lo que es valioso o bueno*. Esta última pregunta pertenece a un área importante de la ética conocida como axiología (o teoría del valor), y es por eso que en el libro recomiendo un *giro axiológico* en la teoría del castigo, y por qué la versión de retribucionismo que continúo defendiendo afirma – por encima de todo – el *valor* de infligir un sufrimiento merecido, sin afirmar que este valor necesariamente genere obligaciones. Los valores, después de todo, a menudo entran en conflicto entre sí, y no hay razón para suponer que el valor del sufrimiento merecido siempre prevalecerá. Y esta es una razón por la cual pienso que es importante recalcar que aun si al nivel deóntico dos posiciones pueden ser coincidentes – como podrían serlo la Rodríguez Horcajo y la mía – las diferencias axiológicas pueden ser importantes e iluminadoras.

Si bien en la medida en que no genera obligaciones, mi retribucionismo axiológico es más modesto que el típico, generador de obligaciones, también es más complicado. Mi retribucionismo no ofrece fórmulas o pautas acerca de cuándo se justifica el castigo y cuándo no. Algunos – Betegón incluido, cuya posición discutiré enseguida – ven esto como una potencial debilidad de mi posición, pero yo lo veo como una fortaleza. Porque creo que la noción misma de “justificación” en juego en la teoría del castigo tradicional es en sí misma demasiado simple y demasiado aplanadora de la experiencia moral. Esta es una de las áreas en las que *Rethinking Punishment* importa aportaciones recientes dentro de la teoría ética con el objeto de enriquecer la discusión del castigo.

A grandes rasgos, justificar algo es tornarlo justo, o tornarlo bueno. (Lo que es asumido como bueno *ab initio* no necesita de una justificación.) Esto significa que una vez que se justifica una acción, ya no está mal. Considérese un caso estándar de justificación en la ley, la legítima defensa: uno repele una amenaza no provocada, ilegítima, e inminente a su vida (que uno no tenía otra forma de evitar, etc.) matando al agresor. Dadas estas y otras condiciones, matar al agresor puede estar “justificado” o, en otras palabras, considerado como una acción “no mala”.

Pero considérese el perspicaz comentario de Hans Welzel: “hay una gran diferencia entre matar a un ser humano en defensa propia (algo que no está mal), y matar (por alguna razón) un mosquito” (Welzel, 1961, p. 24). La muerte de un ser humano, incluso cuando está justificada, es un asunto mucho más serio, complejo y,

de hecho, *trágico* que la muerte de un mosquito. El hecho de que la muerte de un ser humano *pueda* justificarse, al igual que el hecho de que el castigo de un ser humano *pueda* justificarse, no debe cegarnos en cuanto a la rica complejidad de estas situaciones. Y, sin embargo, exactamente esta ceguera es puesta de manifiesto en la teoría del castigo contemporánea: el castigo se considera “no malo” por algunas de las buenas consecuencias que genera, o por el hecho de que es merecido, y el resto del contexto en el que esto ocurre es obviado.

En general, soy partidario de simplificar. Sin embargo, dentro del contexto de la justificación del castigo, ciertas simplificaciones han devenido en minimalismo, y este minimalismo ha demostrado ser altamente problemático, ya que, después de todo, la vida moral es mucho más complicada de lo que éste puede reconocer. Las justificaciones tradicionales del castigo han intentado ponerle una suerte de camisa de fuerza a la vida moral, y esto es parte de la explicación del estancamiento improductivo en el que nos encontramos. Por lo tanto, si una cierta explicación de la justificación resulta ser más compleja que sus rivales, como lo es la mía, creo que este es un punto a su favor.

Como primera aproximación a la complejidad que tengo en mente, y que los teóricos del castigo de todas las tendencias han sistemáticamente ignorado, considérese el famoso caso hipotético que Williams sugirió (1976, PP. 115-135). Un camionero completamente inocente, es decir, actuando sin culpa alguna, atropella a un bebé que de alguna improbable manera gateó hasta la carretera. Williams sugiere que queremos que este conductor se sienta *culpable*, incluso si sabemos (y él sabe) que no ha hecho nada malo, e incluso cuando eventualmente nos esforzaríamos por ayudarlo a dejar de sentir esa culpa. Como muchos han comentado, nos resultaría escalofriante si este camionero – verdaderamente “inocente” – al darse cuenta (justamente, en cierto sentido) de su inocencia, simplemente se alejase de la horrible escena del accidente silbando una melodía feliz. Y, sin embargo, la mayoría de los filósofos morales contemporáneos se resisten a la sugerencia de Williams, alegando que la idea de “sentirse culpable sin haber hecho algo malo” es simplemente absurda. En contraste, considero que la visión de Williams no sólo no es “absurda”, sino que es *profundamente* correcta.

Ya debería estar claro lo que quiero hacer con las ideas de Welzel y Williams: un castigador *decente* debería sentirse *culpable*, incluso cuando esté justificado para castigar. Si queremos que el conductor de Williams que no ha hecho nada malo se sienta culpable, o el que mata en legítima defensa de Welzel (cuya acción está justificada) se sienta culpable, ¿por qué no querríamos que el castigador justificado también se sienta culpable? Consciente de que el castigo es una cuestión de hacer que la gente sufra, y que, sin embargo, a veces estamos justificados en infligirlo, el castigador, mucho más que el camionero y quien mata en defensa propia, enfrenta un dilema moral complejo.

La historia de las ideas es a veces sorprendente, y el caso de los dilemas morales es un buen ejemplo. Hasta hace muy poco, los filósofos morales tendían a prestar mínima atención a los dilemas morales. Escribiendo en el 2006, Alasdair MacIntyre describió la situación elocuentemente: “si uno publicara dos volúmenes [sobre dilemas morales], el primero contentivo de toda la literatura filosófica precedente que trata este tema, ampliamente interpretada, desde Platón a W.D. Ross a través de Gregorio, Tomás de Aquino, Kant, Hegel, Mill, Sidgwick y Bradley, y el segundo contentivo de las publicaciones de los últimos treinta años, el segundo volumen sería, por mucho, el más grueso” (MacIntyre, 2006, p. 85).

Si los filósofos morales en general han prestado poca atención a los dilemas morales, los teóricos del castigo en particular casi no les han prestado atención alguna. Por supuesto, algunos teóricos del castigo se han preocupado por algunos dilemas. ¿Deberíamos gastar recursos en el castigo cuando podríamos gastarlos en las escuelas (lo cual, por cierto, ha demostrado ser más efectivo para reducir el crimen que cualquier plan de castigo)? O, en el contexto de la justicia transicional, ¿deberíamos castigar a todos los miembros de un régimen tiránico recientemente depuesto, incluso si ello amenaza la viabilidad misma del régimen naciente y vulnerable? Que este tipo de dilemas son importantes está fuera de toda duda. Pero ellos no son *teóricamente* interesantes. Una afirmación central que adelanto en el libro es que el castigo es *teóricamente* dilemático. Independientemente de cualquier consideración política o de asignación de recursos, el castigo *necesariamente* confronta al valor del sufrimiento merecido y al valor de su remisión misericordiosa (el valor de la justicia y el valor del perdón).

Mi tesis es que el castigo y el perdón deben estudiarse juntos porque el *valor* de uno no puede entenderse completamente sin entender el *valor* del otro: son axiológicamente inseparables. Es más, mi tesis es que el debate *central* de la justificación del castigo no es el que opone el valor de la retribución al valor de ciertas consecuencias favorables (como se ha creído por siglos, y como Rodríguez Horcajo recoge en su título), sino el que opone el valor de la justicia (generado al castigar merecidamente) al valor del perdón (generado al no castigar a quien merece castigo). Por lo tanto, el estancamiento en la literatura especializada sobre el castigo no se superará hasta que los teóricos del castigo comiencen a prestar más atención al perdón.

Los teóricos del perdón están familiarizados con la naturaleza dilemática, en realidad *paradójica*, del fenómeno que estudian; son perfectamente conscientes del hecho de que cuando perdonan y, por lo tanto, generan el valor del perdón, no generan el valor de la justicia. Además, están profundamente interesados en explorar las complejas respuestas emocionales que generan los complejos fenómenos morales. Si los teóricos del castigo siguieran el ejemplo de los teóricos del perdón, verían que cuando castigan y, por lo tanto, generan el valor de la justicia, dejan de generar el valor del perdón y que esto debería generar una reacción emocional compleja.

Desafortunadamente, no es sólo la noción de justificación presupuesta por los teóricos contemporáneos del castigo que resulta demasiado simplona, sino que el mismo *agente* que presuponen es demasiado *simplón* o caricaturesco. Williams definió una cierta condición intelectual como “simplona” (“*simpleminded*”) cuando posee “muy pocos pensamientos y sentimientos para hacerle justicia al mundo como realmente es”. En mi último libro, sostengo que todos, o virtualmente todos, los teóricos contemporáneos del castigo son, en este sentido, “simplones”.

Considérese el caso de Benjamin Murelstein, el rabino vienés que los nazis pusieron a cargo de *Theresienstadt*. Es difícil imaginar las horribles y trágicas decisiones que Murelstein – un prisionero de los nazis, al fin de cuentas – tuvo que enfrentar. Estas no solo implicaban castigar a sus compañeros de prisión, sino abusar de ellos, y enviarlos a su muerte segura en cámaras de gas. En su última película, *Le dernier des injustes*, el gran Claude Lanzmann increpa a Murelstein: “al escucharlo hablar de *Theresienstadt*, uno no tiene la impresión de que era un lugar donde reinaba la desgracia, un lugar de sufrimiento donde miles de personas murieron, y que era una escala en el camino a Auschwitz para miles más. Cualquiera pensaría que no siente usted nada mientras habla de *Theresienstadt*”. Sin la menor vacilación,

Murmelstein respondió: “si durante una operación un cirujano comienza a llorar por su paciente, lo mata. No llegas muy lejos llorando o vacilando”.

Murmelstein es una versión *real* del camionero de Williams y de quien actúa en legítima defensa de Welzel, y esta es una de las razones por las que su apatía es mucho más escalofriante que la de los personajes en esos cuidadosamente contruidos casos hipotéticos. Uno hubiese querido que Murmelstein se sintiera culpable, incluso si tal vez estuviera justificado actuar como lo hizo, e incluso si hubiera quizás sido sensato para él bloquear de alguna manera esta culpa *en el momento en que actuaba*. El hecho de que décadas después de enfrentarse a esas trágicas decisiones, Murmelstein pudiese reflexionar con tanta frialdad sobre ellas revela una notable ceguera – o incluso abyección – moral.

Es por esta razón que he intentado abrir un espacio teórico dentro de mi justificación de castigo para la consideración de la complejidad axiológica y emocional que engendran los dilemas morales – entre los cuales, repito, incluyo al castigo. No dudo que haya casos en los que los actos que castigamos sean tan atroces que generar el inmenso valor de la justicia en esos casos empequeñezca el valor no generado pero minúsculo del perdón. Pero esto no es típico, y ciertamente no es en lo absoluto típico en la justicia penal. En casos típicos, el dilema moral que presumo debería dar lugar a estados emocionales que con muchos más matices que simplemente (o simplonamente) sentirse “bien” por haber hecho justicia. Los teóricos del castigo tienden a asumir que una vez que el castigo se justifica, no queda nada moralmente significativo, ni en el ámbito de los valores ni en el ámbito de nuestras emociones. Esto es, en su opinión, exactamente lo que significa *justificar* el castigo. Desde mi punto de vista, este enfoque empobrecido y simplón falsea la realidad moral, y al hacerlo, frustra el desarrollo de la teoría del castigo.

El punto que quiero subrayar es que no importa cómo el dilema moral que el castigo genera sea “resuelto”, o, en otras palabras, cuál de estos dos valores enfrentados entre sí termina “derrotando” al otro. El valor “derrotado” no se desvanece, sino que *permanece* después de la “derrota”. La noción técnica de un residuo moral es uno de los elementos más importantes en mi aproximación a la justificación del castigo. Dado que el castigo necesariamente implica no generar algún valor, y dado que esta no generación constituye un residuo moral, los castigadores simplemente no pueden evitar ensuciarse las manos.

“Manos sucias” no es solo una metáfora. Es un tipo específico de dilema moral en torno al cual se ha desarrollado recientemente un área de investigación muy fértil. El nombre “manos sucias” se deriva de la obra homónima de Jean Paul Sartre. Una situación de manos sucias se produce cuando no podemos hacer el bien sin hacer el mal: la idea no es que tengamos que hacer algo malo para obtener algo bueno, sino que uno y el mismo acto es *simultánea e ineludiblemente* bueno y malo.

Los dilemas de las manos sucias representan las decisiones trágicas que a veces enfrentamos en la política y en la moral, y que afectan nuestra vida emocional de maneras muy complicadas. Las desagradables consecuencias emocionales que generan algunas de nuestras acciones no son meros *costos* para ingresar a los libros de contabilidad, como muchos (si no la mayoría) de los filósofos morales asumen, y como la mayoría (si no todos) los teóricos del castigo asumen. Más bien, estos residuos morales son *manchas* (o *cicatrices*), que obstinadamente se nos adhieren, que afectan nuestras autoconcepciones y nuestras autobiografías, y que no pueden ser fácilmente lavadas. Además, estas manchas son omnipresentes: es una vana e ingenua esperanza pensar que podemos vivir nuestra vida sin polución moral. No importa cuánto intentemos, o cuán seriamente nos apeguemos a este o aquel sistema

ético, simplemente no podemos evitar contaminarnos. Pero mucho menos lo pueden evitar los castigadores, ya que ellos están en el negocio de hacer sufrir a las personas, y, nuevamente, no puedes estar en ese negocio sin ensuciarte las manos.

Esto no quiere decir que todos seamos, indiscriminadamente, malhechores. Porque evidentemente hay errores y errores; hay manchas y manchas. Los actores muy malos ni siquiera pueden ensuciarse las manos en el sentido requerido, ya que estaban sucios desde el principio: Marmelstein podía ensuciarse las manos; sus captores nazis no podían, ya que estaban sucios desde el principio. Y esto tampoco quiere decir que nunca debemos ensuciarnos las manos; sin duda, algunas veces debemos hacerlo. Pero sí quiere decir que reconocer que nos *estamos* ensuciando las manos es un signo de madurez intelectual y de decencia – y una forma de evitar la apatía y la complacencia.

Es por esto que el último capítulo de mi libro está construido alrededor de la obra maestra de Herman Melville, *Billy Budd, Sailor* (1986). Esta es una novela muy compleja en la que contemplamos cómo, para citar a Melville, en “el malabarismo de las circunstancias, la inocencia y la culpabilidad intercambian lugares”. Billy Budd, el “criminal” en la historia, es inocente en un sentido profundo e importante, mientras que Claggart, su víctima, es culpable en un sentido profundo e importante. El Capitán Vere, posiblemente el personaje principal de la novela, tiene la difícil tarea de castigar a Billy, lo cual hace, no sin expresar la desde entonces famosa frase: “¡el ángel de Dios debe ser ahorcado!”. Después de castigar a Billy, el capitán Vere quedó emocionalmente devastado; Melville lo describe como alguien quien, después de informar a Billy del veredicto, “ya no vivía”. Esta devastación no fue el resultado de dudar si estaba *justificado* en castigar a Billy, sino de comprender, a cabalidad, el trágico dilema moral que generó el castigo y del cual no pudo escapar. Obviamente, no todos los casos de castigo son tan trágicos como el de Billy Budd. Pero todos ellos son dilemáticos, y casi todos, en diferente medida, son trágicos.

Así es que sugiero que la justificación del castigo es mucho más complicada de lo que mis oponentes asumen. Rechazo tanto la comprensión mecanicista de la “justificación” como la postulación de agentes simplones. Propongo la inseparabilidad axiológica del castigo y el perdón, y sugiero que el castigo – aun cuando esté justificado – es un caso de manos sucias que genera residuos morales. Como ha sugerido Stephen Garvey, una idea central de mi libro es que “es imposible castigar, no importa cuán justamente, sin perder *nuestra* inocencia moral”. Es fundamentalmente por estas razones que considero que mi libro supone un cambio de paradigma, una nueva manera de aproximarnos al trillado pero todavía escabroso y no resuelto tema de la justificación del castigo.

Como adelanté desde el comienzo, Betegón y yo coincidimos en mucho, y no sólo en nuestra valoración general de que el retribucionismo merece ser tomado mucho más en serio de lo que normalmente se hace. Pero lo que me interesa enfatizar es que Betegón considera que muchos de los aportes en *Rethinking Punishment* (incluso aquellos que son anti-retribucionistas) son al menos “sugerentes e intelectualmente provocadoras”. En su párrafo final, Betegón reconoce que mis esfuerzos en *Rethinking Punishment* “suponen una auténtica renovación” del debate que nos ocupa, y que mis estrategias “conceden plena credibilidad” al título del libro. De nuevo, no puedo más que agradecer (y celebrar) esta evaluación.

A pesar de que Betegón y yo estemos en cierto sentido en una misma onda (por decirlo de alguna manera), es útil detenernos en algunos de los aspectos que le resultan más discutibles de mi libro. Primero que nada Betegón pausa antes de aceptar del todo mi distinción entre lo axiológico y lo deóntico. Aclaro: no es que

Betegón dude de la existencia de esta distinción, sino que sugiere que no he explicado suficientemente las interrelaciones entre estos dos ámbitos. De hecho, la preocupación de Betegón en este sentido es doble: primeramente, desde un punto de vista teórico, Betegón considera que no he ofrecido suficiente guía acerca de las relaciones entre estas diferentes dimensiones; segundo, y desde un punto de vista más práctico, Betegón considera que no he dicho suficiente acerca de las consecuencias concretas que mi replanteamiento del castigo acarrearía.

Betegón sugiere que mi posición es presentada “en clave axiológica”. Sin duda esto es cierto; pero hasta qué punto ello significa un menosprecio por lo deóntico está, sin embargo, abierto a discusión. Si *Rethinking Punishment* es en efecto presentado en clave axiológica, también lo es en clave pluralista: definiendo, después de todo, un “pluralismo axiológico”. Antes de adentrarme en lo que considero es la preocupación fundamental de Betegón quisiera aclarar un punto acerca del tipo de pluralismo que definiendo en el libro. Pues aunque Betegón, me parece, es afecto a este pluralismo, comenta – con un cierto aire de escepticismo – acerca de lo mucho que abarco con el término “monista”, pues la mayoría de las justificaciones del castigo son, en mi opinión, monistas (y, por tanto, no pluralistas). Lo más importante de subrayar aquí es que definiendo un tipo específico de pluralismo, al que he llamado “pluralismo apropiado”, cuya característica fundamental es que reconoce como esenciales dos valores en particular: el valor de la justicia (generado al castigar merecidamente) y el valor del perdón (generado al no infligir el castigo merecido). No es que (como una pregunta de Rodríguez Horcajo pudiese sugerir) mi pluralismo niegue la existencia o la importancia de otros valores, o los relegue exageradamente a un segundo plano. En lo absoluto: al fin y al cabo reconozco explícitamente un sin fin de cosas que suelen ser valiosas y que una justificación del castigo podría tomar en cuenta: la disuasión, la reconciliación, el aprendizaje moral, la reinserción y la paz social, etc. Lo que es cierto es que son específicamente estos dos – la justicia y el perdón – los que distinguen a mi pluralismo de otros pluralismos, y lo que me permite excluir a tantas otras justificaciones del ámbito de *mi* pluralismo. (La importancia que históricamente han tenido estos dos valores en particular – aun en contextos diferentes al del castigo – no me parece en lo absoluto desdeñable.)

Hay quienes podrían pensar que hay algo sospechoso, circular, o, como dice Betegón, que evidencia “una considerable carga estipulativa” en mi definición de este “pluralismo apropiado”: como si, *deus ex machina*, el único “pluralismo apropiado” fuese el mío. Por ello ofrezco aquí una analogía (ficticia) que me permite elucidar mis razones para hacer esta distinción entre diferentes tipos de pluralismo de una forma no circular, o al menos no *problemáticamente* circular. Supongamos que por muchas décadas ha habido debates entre nutricionistas: algunos proponían que a los fines de perder peso es crucial disminuir calorías consumidas, otros que la clave es comer más frutas y vegetales, otros que la clave es disminuir el azúcar, otros que lo central es no cenar, o ayunar intermitentemente, etc. Supongamos que algunos de estos nutricionistas combinaban algunas de estas sugerencias, afirmando que la manera de perder peso es disminuir calorías y comer más vegetales, o eliminar la cenas y el azúcar, etc. No cabe duda que estos últimos nutricionistas serían, en cierto sentido, pluralistas. Pero imaginemos que surge un nutricionista que sugiere que, independientemente de lo que se ingiera, la pérdida de peso requiere de ejercicio aeróbico. Esta sugerencia es otra cosa, pues involucra un tipo de consideración completamente diferente a lo anterior: introduce una nueva *dimensión* en la discusión. Este nutricionista sería, en mis términos, un nutricionista “pluralistamente apropiado”. Y es esta la explicación de mi pluralismo “apropiado” y de su diferencia con los anteriores: concederle al perdón un papel tan importante dentro de la justificación del

castigo como evidentemente le concedo es ciertamente poco común.⁴ Y a pesar de que reconozco que existe una gran diversidad de elementos que una aproximación pluralista a la justificación del castigo podría (o, en última instancia, debería) tomar en cuenta, son estos dos los constituyen el meollo del dilema *teórico* que el castigo necesariamente genera.

Sin embargo, aun si esto explica satisfactoriamente por qué mi versión del pluralismo no es problemáticamente estipulativa, deja sin responder lo que, creo, es la preocupación fundamental de Betegón: que realmente no ofrezco mayor guía con respecto a las consecuencias deónticas de mi posición. Me parece que es útil subdividir esta preocupación, de nuevo, en dos partes. Primera, que muchas de mis afirmaciones pueden “verse afectadas por su excesiva deuda respecto de un conocimiento puramente intuitivo”. Segunda, que “un mayor desarrollo de las implicaciones deónticas a la luz del análisis axiológico ayudaría a entender mejor el significado de muchas de las afirmaciones que se hacen” en mi libro. Considero que estas son preocupaciones importantes – aun cuando me parecen más valiosas cuando uno las analiza por separado. Es decir, no estoy seguro de que aun si Betegón tuviese razón en que quizás me apoyo demasiado en la intuición, resolver el (a mi juicio independiente) problema de las consecuencias deónticas de mi posición resolvería el problema del papel de la intuición, ni viceversa.

El problema del rol de las intuiciones es hasta cierto punto inevitable. En mayor o menor medida toda teoría ética se apoya, en última instancia, en alguna intuición. Lo ideal, por supuesto, es intentar limitar este apoyo al máximo posible, a sabiendas de que nunca podrá ser eliminado del todo. En *Rethinking Punishment* he, en efecto, tratado de limitar la apelación a las intuiciones al máximo posible. La pregunta es si he tenido éxito en ese intento. Lo único que puedo hacer aquí es mencionar los cinco factores que enumero en el libro acerca de las intuiciones (76 y ss.): que las mismas son realmente inevitables y que tanto yo como mis oponentes eventualmente apelamos a ellas; que las intuiciones a las que apelo yo son, en cualquier caso, mucho más fáciles de aceptar y de hecho mucho más frecuentemente aceptadas por la población en general que aquellas a las que apelan mis oponentes; que *entre los teóricos del castigo* en particular las intuiciones opuestas a las mías – intuiciones utilitaristas – son sumamente prevalentes; que la carga de la prueba ha de recaer sobre los que defienden intuiciones opuestas a las mías; y, finalmente, que existe una diferencia importante entre hablar de intuiciones con respecto a algún caso hipotético aislado (como esta en boga en el contexto de la filosofía moral contemporánea) y hablar de intuiciones con respecto a toda una cosmovisión, a un entendimiento de la ética en sí misma.

Como corolario a todo lo anterior, es útil recordar que mi método podría entenderse como fenomenológico, en el sentido de que consiste en observar nuestra vida moral tal y como comúnmente se nos presenta. Muchos de los fenómenos que introduzco en la discusión del castigo – los residuos morales, la complejidad emocional que circunda decisiones morales difíciles, etc. – son fenómenos que hemos experimentado en nuestras vidas cotidianas, a pesar de la manera sistemática en la que los teóricos del castigo los han ignorado.

El problema de las consecuencias deónticas de mi retribucionismo axiológico, es, me parece, aún más serio. En modo alguno niego que aproximaciones axiológicas, como la mía, estén exentas de dar cuenta de las implicaciones que en el ámbito de la acción (estatal o individual) hayan de generarse. Al contrario, reconozco que en este sentido le incumbe al retribucionista axiológico (como lo soy yo) ofrecer una

⁴ Pero no es inédito: por eso le doy crédito a John Tasioulas por haberlo hecho también.

explicación mucho más de lo que le incumbiría al retribucionista deóntico. Al fin y al cabo, este último nos ofrece *ab initio* la explicación de que es lo que debe hacerse: dar a los malhechores lo que se merecen. (Esto no le da una ventaja al retribucionismo deóntico, pues la agenda deóntica a la que están necesariamente encadenados no es ni suficientemente matizada, ni al final atractiva, como he esbozado arriba.)

Tampoco niego de plano que quizás ahondar más en las consecuencias deónticas de mi posición pueda limitar en algo las preocupaciones con respecto al rol de las intuiciones en mi libro – aun cuando, como ya dije, no estoy muy seguro acerca de exactamente cómo esas clarificaciones deónticas redundarían en clarificaciones meta-éticas con respecto a las fuentes últimas del conocimiento moral. En lo que no difiero en modo alguno de la posición de Betegón (ni de la de Rodríguez Horcajo) es que este ámbito deóntico es, en sí mismo, terriblemente importante. El libro que he escrito no contiene lecciones específicamente deónticas, pero no por que acaso dude de su importancia, sino por que dedicarme a ello me hubiera llevado a escribir un libro bastante diferente. A pesar de que no tengo dudas acerca de lo útil que sería explorar las consecuencias deónticas del tipo de retribucionismo axiológico que definiendo, enfocarme en éste, como hago en el libro, me parece particularmente importante, por lo menos en vistas a lo poco que sobre él se ha dicho en el contexto de la justificación del castigo.

No quisiera concluir sin enfatizar que si bien es cierto que, como apunta Betegón, mucho falta por decir con respecto a las consecuencias deónticas de mi teoría, no es menos cierto que *algo* he dicho al respecto, así sea de manera genérica e indirecta. Pues si la tesis que definiendo de ver al castigo (aun cuando esté justificado) como algo trágico, como algo que nos roba de cierta inocencia, es correcta, esto sugeriría una cierta humildad, una cierta circunspección, a la hora de castigar. Y esto a su vez apuntaría al desarrollo de políticas punitivas (tanto a nivel estatal como individual) más prudentes y humanitarias –y esto sin errar el blanco, sin equivocadamente atacar al supuestamente barbárico retribucionismo.

Bibliografía

- BENTHAM, J (1962), "An Introduction to the Principles of Morals and Legislation", en *The works of Jeremy Bentham* (John Bowring, ed.), Vol. 1, Russell & Russell, New York, (1962)
- BETEGÓN, J. (1992), *La justificación del castigo*, Centro de Estudios Constitucionales, Madrid.
- RODRÍGUEZ, D. (2016), *Comportamiento y pena estatal: Disuasión, cooperación y equidad*, Marcial Pons, Madrid.
- HART, H. L. A. (2008), *Punishment and responsibility: Essays in the philosophy of law*, (segunda edición), Oxford University Press, Oxford.
- MACINTYRE, A. (2006), *Ethics and politics: Selected essays*, Vol. 2, Cambridge University Press, Cambridge.
- MELVILLE, H. (1986), *Billy Budd, sailor, and other stories*, Penguin, New York.
- MOORE, M. (1992) 1953, *Placing blame*, Oxford University, Oxford.
- NIETZSCHE, F. (2005), *Twilight of the idols or how to philosophize with a hammer*, en *Friedrich Nietzsche The antichrist, Ecce homo, Twilight of the idols, and other writings*, (Aaron Ridley and Judith Norman, eds.), Cambridge University Press, Cambridge.
- PARFIT, D. (2011), *On what matters*, Vol. II, Oxford University Press, Oxford.
- WELZEL, H. (1961), *Das neue Bild des Strafrechtsystems: Eine Einfuehrung in die finale Handlungslehre*, O. Schwartz, Goettingen.

- WILLIAMS, B. (1976), "Moral Luck", *Proceedings of the Aristotelian Society*, 50, pp. 115-151.
- WILLIAMS, B. (2006), *Ethics and the limits of philosophy*, Routledge, London.
- ZAIBERT, L. (2005), *Five ways Patricia can kill her husband: A Theory of intentionality and blame*, Open Court, Chicago.
- ZAIBERT, L. (2006), *Punishment and retribution*, Ashgate, Aldershot.
- ZAIBERT, L. (2018), *Rethinking Punishment*, Cambridge, Cambridge University Press.