UNIVERSITAS
Miguel Hernández

Trabajo Fin de Grado

El Derecho penal (de nuevo) ante los retos de la IA: aproximación a un modelo de criminalización de los deepfakes

Alumno: Carlos Pérez Olivares

Tutor: Fernando Miró-Llinares

Universidad Miguel Hernández

Facultad de Ciencias Sociales y Jurídicas de Elche

Grado en Derecho

Curso académico 2023-2024

«Todos saben que existe, todo el pueblo de Omelas. Algunos han ido a verlo, otros se contentan únicamente con saber que está allí. Todos saben que tiene que estar. Algunos comprenden la razón, otros no, pero ninguno ignora que su felicidad, la belleza de su pueblo, la ternura de sus amigos, la salud de sus hijos, la sabiduría de sus becarios, la habilidad de sus artesanos, incluso la abundancia de sus cosechas o el esplendor de su cielo dependen por completo de la abominable miseria de ese niño».

Úrsula K. Le Guin, *Los que se alejan de Omelas.*

ÍNDICE

I. INTRODUCCIÓN Y OBJETIVOS DEL TRABAJO

La inteligencia artificial (en adelante, IA) ha irrumpido con fuerza en nuestras sociedades en los últimos años y ha recabado en gran medida el interés de la ciudadanía y los medios de comunicación. Tanto es así que el término ‘’inteligencia artificial’’ fue escogido por la Fundación Español Urgente (FundéuRAE), entidad promovida por la Agencia EFE y la Real Academia Española, como la expresión del año 2022 (FundéuRAE, 2022), habiéndose, incluso, atribuido a este tipo de tecnología, que se identifica como la canalizadora de una cuarta revolución industrial, el efecto de estar ‘’difuminando las líneas entre las esferas física, digital y biológica’’ (Schwab, 2020, p.6).

Dentro del ámbito de las herramientas digitales que emplean la IA ha cobrado especial notoriedad el de los modelos generativos, por medio de los cuales se consigue la creación de contenidos multimedia hiperrealistas, denominados comúnmente *deepfakes,* a partir de la introducción de una serie de términos, habiéndose logrado tal hazaña gracias a los recientes avances en las llamadas ‘’redes neuronales profundas’’ (Gil *et al.,* 2023); habiendo, nuevamente, llegado a ser candidata a palabra del año 2023 ‘’ultrafalso’’, término empleado como sustituto en castellano para *deepfake* (FundéuRAE, 2023). Y no es de extrañar la repercusión mediática de esta tecnología, pues en los últimos meses se han venido sucediendo las noticias en que estas herramientas aparecían como medio empleado para la realización de graves conductas. Quizá uno de los que más impacto tuvo en nuestro país fue el caso de los falsos desnudos de Almendralejo, en que se difundieron en septiembre del año 2023 imágenes falsas de más de 30 menores de edad desnudas generadas por medio de un instrumento que emplea este tipo de IA (RTVE, 2023). Otros sucesos recientes que cabe señalar a nivel internacional son el caso del trabajador de una multinacional que realizó una transferencia de 25 millones de dólares tras celebrar una videollamada con quien parecía ser el director financiero de su empresa pero que, en realidad, era una simulación generada con IA (Chen y Magramo, 2024); o el de la difusión de un vídeo pornográfico protagonizado por la Presidenta del Consejo de Ministros de Italia, Giorgia Meloni, creada con esta tecnología (Gozzi, 2024). Debe tenerse en cuenta que la capacidad de generar contenido multimedia sintético e hiperrealista se ha puesto al alcance del público general, produciendo una notable proliferación de los *deepfakes*, habiéndose detectado que desde diciembre de 2018 hasta septiembre de 2019 el número de vídeos ultrafalsificados se había duplicado (Ajder *et al*. 2019).

Todo ello ha hecho surgir en todo el mundo cierta preocupación respecto a esta tecnología y las amenazas que de su mal empleo pueden derivarse para con nuestras sociedades. En este sentido se ha entendido que las herramientas de IA generativa podrían llevar consigo, al margen de las oportunidades a nivel creativo y de entretenimiento, aunque también siendo polémico; una gran amenaza a múltiples niveles, destacándose una notable preocupación por su posible uso como herramienta al servicio de la violencia contra las mujeres (Simó, 2023) o su identificación como una amenaza para la democracia (Citron y Chesney, 2019; Lavanda, 2022; Pawelec, 2022). El interés por el asunto en el ámbito académico se ha plasmado, como muestran Gil et al (2023), en un aumento significativo entre los años 2018 y 2021 de las publicaciones sobre *deepfakes*, lo cual no deja de ser lógico si se tiene en cuenta que el propio término surgió en noviembre del año 2017 cuando un usuario de *Reddit* lo acuñó al crear un espacio en esta red dedicado a la creación de vídeos pornográficos en los que se colocaban por medio de inteligencia artificial las caras de distintas celebridades (Ajder *et al*. 2019).

Sin embargo, la revisión sistemática mencionada sí arroja datos de gran interés en lo relativo al área de conocimiento en que se inscriben los 311 artículos científicos

analizados, pues un 60,3% se integrarían en el ámbito de las Ciencias de la Computación y de la Ingeniería en general; mientras que tan solo un 9,4% de las publicaciones encontradas por dicho estudio pertenecen al ámbito de las ciencias sociales, siendo los temas que principalmente se afrontan en dichas publicaciones esencialmente relacionados con cuestiones técnicas informáticas relativas a esta tecnología, como *deep learning*, *face recognition* o *convolutional neural networks* (Gil et al., 2023). Esta observación muestra cómo, pese a que muchas de las cuestiones y retos que plantean los *deepfakes* son de una naturaleza social, jurídica y económica; no están siendo afrontadas de forma expresa y generalizada desde áreas de conocimiento como el Derecho en general y, por cuando aquí nos ocupa, por el Derecho penal en particular; o al menos estas cuestiones no han despertado tanto interés académico como los asuntos más estrictamente ingenieriles.

Aunque es muy probable que esta tendencia finalice pronto y estas problemáticas sean abordadas por la doctrina penal de forma más prolífica no deja de ser especialmente llamativo cómo, si bien sí se ha despertado en cierta medida el interés sobre las implicaciones penales de la IA en general, especialmente enfocados en las preguntas sobre la atribución de responsabilidad penal por los resultados lesivos producidos por sistemas de IA; los análisis y publicaciones que más particularmente se centran en las ultrafalsificaciones han aparecido de una forma más discreta y desde tres perspectivas principales: En primer lugar, la del proceso penal y la Administración de Justicia Penal, como los de Malacarne (2023), Blázquez (2023), o Durães, Freitas y Novais (2024), por mencionar algunos. Por otro lado, encontraríamos la línea de tratar de acomodar ciertas conductas relacionadas con estas herramientas en un tipo penal existente o en dejar patente la imposibilidad de hacerlo y la necesidad de inclusión de nuevos preceptos penales o de actualización de los mismos, lo que podríamos incluir en un análisis más cercano a la Parte Especial del Derecho penal, pudiéndose encontrar un análisis ciertamente precoz y embrionario sobre este aspecto en la Circular 2/2015 de la Fiscalía General del Estado, donde se cuestiona el encaje en de la que denomina ‘’pornografía virtual’’ en el tipo de la pornografía infantil; dedicándose a esta perspectiva también unas páginas del trabajo de Simó (2023). En último lugar, también se ha propuesto la doctrina elaborar un estudio comparativo de las distintas propuestas e iniciativas legislativas a nivel mundial relativas al fenómeno de los *deepfakes*, como las aportaciones de Rodrigo (2020) y Mania (2024)

Parece sorprendente que, sin embargo, entre estas aportaciones doctrinales no se encuentren a penas intentos de abordar esta cuestión tratando de dar respuesta a las problemáticas jurídico-penales derivadas de los modelos generativos de IA desde el análisis de si la naturaleza de las amenazas que estas generan suponen en realidad una novedad que desborda el marco establecido por los actuales sistemas penales y produce una necesidad de su actualización o si, por el contrario, nos encontramos ante problemáticas que pueden afrontarse satisfactoriamente desde el modelo actual. Pueden encontrarse intentos en este sentido como el realizado por Paulo Rodrigues, que se cuestiona sobre la legitimidad del castigo de los *deepfakes* pornográficas, preguntándose cuál sería el bien jurídico protegido, la relevancia jurídico-penal de estas conductas y las características del riesgo que generan (Rodrigues, 2023); y hasta cierto punto, aunque respecto a la ciberdelincuencia en la era digital y no sólo en relación con la IA y *deepfakes*, el de José R. Agustina, que plantea que este nuevo contexto digitalizado puede implicar necesarios cambios en la forma de concebir el delito y en ciertas categorías dogmáticas jurídico-penales (Agustina, 2021).

Entendemos que las respuestas a esta cuestión centradas exclusivamente en un examen de las descripciones típicas existentes en el Código Penal para tratar de afirmar

que las ultrafalsificaciones pueden ser o no castigadas serán en última instancia respuestas incompletas y que no consiguen dar cuenta satisfactoriamente de la verdadera profundidad que el asunto presenta, lo cual no implica necesariamente que dichas respuestas no sean acertadas en su conclusión final y a nivel práctico. Con esto queremos decir que el mero contraste de unas conductas con una entidad y naturaleza novedosa y disruptiva, como son las que nos ocupan aquí, con un tipo delictivo previsto por el legislador con anterioridad no basta para poder defender con suficiente solvencia y solidez el castigo de dicha conducta o la necesidad de establecer nuevos delitos que así lo permitan. Para ello es necesario que este examen se realice partiendo de un modelo de criminalización en que se fijen los criterios que en un Estado social y democrático de Derecho deben informar la decisión sobre la criminalización o no de determinadas conductas; así como analizar si la naturaleza de estos actos novedosos se corresponde con la del principio que fundamenta la criminalización del delito tipificado que se pretende aplicar. De lo contrario corremos el riesgo de subsumir en un tipo preexistente una nueva realidad, que no pudo ser tenida en cuenta por el legislador, cuyo castigo puede no responder en puridad al fundamento que justifica la criminalización de la conducta típica descrita en el Código Penal; o bien de establecer nuevos delitos para poder castigar estos hechos desatendiendo a los pilares más básicos en que entendemos que debe sostenerse un sistema penal liberal propio de un Estado social y democrático de Derecho como el nuestro. En definitiva, la preocupación que subyace en esta reflexión es si, al contrario de en lo relativo a la criminalización de las conductas ofensivas, en que podría darse un ‘’déficit aparente de legitimidad para la criminalización’’ (Miró-Llinares, 2015, p.46); en esta cuestión puede estar generándose un ‘’superávit aparente de legitimidad para la criminalización’’ de los actos relacionados con este tipo de herramientas de IA que lleve a un castigo injustificado de los mismos al aplicar para ello unos tipos penales existentes que aluden a una conducta de una naturaleza diferente o bien a la creación de nuevas figuras delictivas ilegítimas.

Es cierto que en la práctica puede ser improbable que esto ocurra y, en la mayoría de las ocasiones, la reconducción de estos nuevos fenómenos a figuras delictivas preexistentes no suponga una criminalización inadmisible ya que la conducta entraña en sí mismo unos riesgos que se alinean, en definitiva, con el fundamento que se entendió suficiente para castigar en abstracto una conducta. Por otro lado, sí parece más justificado, en un contexto de expansión del Derecho penal y cada vez más marcadamente punitivista como el actual (Silva, 2001), sospechar que surjan pulsiones tendentes a la creación de nuevos delitos que no siempre sean respetuosos con los requisitos y criterios que se identifican como necesarios para afirmar que la criminalización es legítima.

Por ello, la rigurosidad que debe caracterizar la actividad doctrinal exige, consideramos, dejar patente y explicitar el encaje sistemático que pueden tener los *deepfakes* en el modelo de criminalización por el que optemos y mostrar cómo el castigo de estas conductas desde el mismo está justificado. Esto conlleva la necesidad de determinar en primer lugar qué características de una conducta pueden justificar su castigo por medio del Derecho penal para, posteriormente, comprobar si este nuevo fenómeno auspiciado por la irrupción de la IA responde a alguna de estas razones, y sólo entonces pasar a examinar si pueden incluirse sus distintas manifestaciones en el ámbito de los tipos penales existentes en la parte especial del Código Penal y, si no es posible, dirimir la cuestión sobre si es aconsejable o no introducir nuevos delitos.

Si bien sobre esta reflexión se volverá más adelante, creemos que con lo aquí expuesto se justifica suficientemente el interés por abordar la relación entre *deepfakes* y Derecho penal con vocación de sistematicidad, que parta de un análisis propio de la Parte

General como es el de cuestionarnos qué tipos de conductas pueden ser legítimamente castigadas por el Estado pero que no por ello dé la espalda a las particularidades específicas de cada delito concreto con que se puedan relacionar estas ultrafalsificaciones, y los conflictos y problemas que el castigo de cada conducta en particular puede conllevar.

En línea con este planteamiento, este trabajo se propone como objetivos los de analizar si es posible el castigo legítimo desde nuestro sistema de Derecho penal de ciertas conductas relacionadas con el uso de modelos generativos de inteligencia artificial, fijar en qué condiciones lo sería y explorar qué tipo penal, en su caso, podría emplearse para el castigo de algunas de ellas. Para conseguirlo trataremos de ofrecer un modelo de criminalización general que establezca los criterios que habiliten el legítimo castigo por el Estado de determinadas conductas de forma coherente con los principios básicos y elementales que deben informar el sistema penal de un Estado social y democrático de Derecho; y todo ello tomarlo como base que nos permita examinar, si desde dicho prisma teórico, pueden castigarse ciertas conductas relacionadas con el fenómeno *deepfake* y bajo qué condiciones podría darse el castigo de algunas de ellas.

II. APROXIMACIÓN AL PROBLEMA DE LOS MODELOS GENERATIVOS DE INTELIGENCIA ARTIFICIAL Y CÓMO AFRONTARLOS

1. Concepto de IA y *deepfake*: De los cepillos de dientes cantantes a la pornografía falsa hiperrealista.

Tratar de establecer una definición de inteligencia artificial no es sencillo ni han sido pacíficamente aceptados los conceptos que se han ofrecido. Aunque no es objeto de este trabajo tratar de indagar pormenorizadamente en la literatura relativa a la definición de la IA, podemos encontrar en el Reglamento Europeo sobre Inteligencia Artificial, habiéndose aprobado el día 13 de marzo de 2024 por el Parlamento Europeo su última versión a la fecha de escribir estas líneas (Resolución legislativa del Parlamento Europeo P9_TA (2024)0138), la primera definición legal de la misma, siendo este instrumento de gran interés para este trabajo puesto que, además de pronunciare sobre el concepto de IA, también trata de ofrecer una noción de lo que debe entenderse por *deepfake*.

Así, el artículo 3 del Reglamento define «sistema de IA» ‘’ un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales’’ (art. 3.1 Resolución Legislativa del Parlamento Europeo P9_TA(2024)0138). Esta definición sintetiza la concepción de los sistemas de IA dada por la OCDE, que establece que ‘’an AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment’’ (OECD/LEGAL/0449) y la definición que había venido siendo manejada en la Propuesta de Reglamento de la Comisión Europea, en cuyo artículo tres se dispone que por sistema de IA se entenderá ‘’el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o

decisiones que influyan en los entornos con los que interactúa'' (art. 3.1 Propuesta de la Comisión Europea COM/2021/206) y que ha sido actualizada en la nueva versión del Reglamento.

En definitiva, estas definiciones aluden a una noción de sistemas de IA que presenta dos elementos principales y esenciales ya señalados como tales en el Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea (2020), a saber, ''los «datos» y los «algoritmos»'', pero añadiendo como punto central adicional su orientación a la consecución de ciertos objetivos dados. Y es que, en efecto, el concepto IA no haría sino referencia a ''la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano'' (Rouhiainen, 2018), dándose este aprendizaje y toma de decisiones para la consecución de los objetivos que persigue por medio de ''operaciones aritmético-lógicas, de lectura/escritura de registros y de control de flujo secuencial'' (Morales, 2021, p.184).

En consecuencia, para los fines del presente trabajo, podemos optar por una conceptualización de los sistemas de IA en sentido amplio que recoja los distintos rasgos característicos destacados anteriormente que nos lleve a entenderlos como una suerte de dispositivos o herramientas informáticas que, a través del empleo de algoritmos, persiguen, de manera más o menos autónoma, la consecución de objetivos determinados por los humanos a partir de un conjunto de datos que se le aportan.

Al margen de las distintas clases de IA existentes en la actualidad, con sus propios modelos de funcionamiento, resulta más interesante para nuestro propósito el examen de uno de los resultados ofrecidos por el desarrollo de esta tecnología: la obtención de sistemas de IA que logran generar contenidos multimedia ficticios realistas solicitados por el usuario. Esto son los conocidos como modelos generativos de IA, que se basan en el empleo de redes generativas adversariales, GAN's por sus siglas en inglés, que se componen de dos redes neuronales que trabajan conjuntamente para la creación de las imágenes, vídeos o sonidos, a saber, una que crea los contenidos falsos, constituida por un codificador y un decodificador; y otra que valora el realismo y precisión de dichos contenidos generados (Rana et al 2024). Este funcionamiento se puede plasmar metafóricamente con una dinámica dialéctica en que ''el modelo generativo se correspondería con un equipo de falsificadores que intentan producir dinero falso, mientras que el modelo discriminativo se personificaría en los cuerpos policiales'' (Simó, 2023).

Con este proceso se logran imágenes, vídeos y sonidos que consiguen representar hechos ficticios con, por lo general, un alto grado de precisión y realismo, pudiendo llegar en ocasiones a generar la idea en quien las observa de que dichos sucesos, palabras o imágenes son reales. Como es esperable, esta utilidad ha sido empleada de modo benigno con fines comerciales o de mero entretenimiento. Un ejemplo de ello sería el conocido anuncio de la cerveza Cruzcampo (Canal Cruzcampo TV, 2021), en que, gracias a este tipo de tecnología, se puede ver y escuchar a la cantante Lola Flores hablar sobre sus raíces andaluzas. Se ha llegado, incluso, al punto absurdo de lograr realizar versiones de conocidas canciones a partir del sonido de un cepillo de dientes (Canal Nikolay Baliev, 2023)[1].

Pese a ello, y tal y como ya se ha apuntado en la introducción, este tipo de herramientas no sólo ha sido empleado con este tipo de finalidades lícitas, sino que han

[1] Un amplio repertorio de ejemplos audiovisuales en YouTube de este tipo de usos de IA generativa puede encontrarse en el trabajo de Cerdán-Martínez et al. (2020).

sido muy variados los ejemplos de su mal uso. En el año 2019 se estimó que el 96% de los vídeos *deepfake* en internet presentaban contenido pornográfico y que esta tipología de pornografía sólo representaba gráficamente a mujeres (Ajdar et al 2019), con lo que, a pesar de que es posible que su actual desarrollo y la proliferación de las herramientas que permiten el acceso a este tipo de tecnología al gran público en los últimos años haya producido que los *deepfakes* no se traten de contenidos casi exclusivamente pornográficos; parece lógico plantear, como hace Elisa Simó (2023), la hipótesis de que las *deepfakes* puedan convertirse en un facilitador de la violencia de género. Pero la potencial problemática de los *deepfakes* no acaba ahí. Se ha suscitado el debate sobre el impacto epistémico que podrían llegar a producir, el grado de amenaza para el discernimiento entre lo real y lo falso que entrañan, y el impulso hacia la posverdad que supondrían, lo que podría poner en apuros a los sistemas democráticos (Rini, 2020; Harris, 2021; Habgood-Coote, 2023; Pawelec, 2022). De igual modo, la generación de este tipo de imágenes puede suponer un medio excelente para la comisión de delitos económicos, pudiendo incidir muy especialmente en la proliferación de estafas donde se logre un engaño bastante en la apariencia de realidad sobre ciertos hechos, circunstancias o características de un determinado objeto o producto que lleve a las víctimas a realizar actos de disposición patrimonial en perjuicio propio desde el convencimiento de la realidad de aquello que han visto o escuchado.

Todos estos fenómenos, tanto inocuos como problemáticos, se han agrupado bajo la denominación de *deepfake*, siendo un término cuya definición no ha suscitado una gran controversia, pareciendo que existe una noción implícita comúnmente aceptada sobre lo que son. Para Pawelec (2022) serían contenidos audiovisuales con caras, cuerpos o voces humanas usualmente creadas utilizando IA. Según Van der Sloot y Wagensveld (2022) un *deepfake* es un contenido de vídeo, audio o de cualquier otro tipo que se ha fabricado total o parcialmente a partir de contenido existente, ya sea vídeo, audio o cualquier otro tipo. Entiende Rodrigues (2023) por *deepfake* un montaje ultrarrealista en el que el rostro de una persona se superpone al cuerpo de otra en un vídeo, pudiéndose además combinar con una manipulación de voz, a través de sistemas de inteligencia artificial, con el fin de inducir una percepción falsa sobre el participante de ese vídeo. García-Ull (2021) afirma que ‘’los deepfakes son vídeos hiperrealistas manipulados digitalmente para representar a personas que dicen y hacen cosas que en realidad nunca dijeron ni sucedieron’’ (p.107). Por otro lado, en un trabajo para el Parlamento Europeo, Boheemen et al (2021) definieron los *deepfakes* como ‘’manipulated or synthetic audio or visual media that seem authentic, and which feature people that appear to say or do something they have never said or done, produced using artificial intelligence techniques, including machine learning and deep learning’’(p.2).

Puede observarse cómo, pese a sus diferencias, estas distintas conceptualizaciones aluden a una serie de elementos que, más o menos, pueden encontrarse en todas ellas: Se tratan principalmente de imágenes o vídeos, aunque también sonidos, excluyéndose en principio del concepto de *deepfake* los productos de bots conversacionales como ChatGPT. En segundo lugar, en estos contenidos se representan a personas, o al menos ciertos rasgos identificativos de la una persona, de forma realista realizando actos o en circunstancias irreales. Por último, estos elementos son generados de digitalmente por medio del uso de sistemas de IA.

Estos distintos elementos han sido igualmente sintetizados en la propuesta de Reglamento europeo de IA que, bajo la denominación en castellano de ultrafalsificación, ofrece una definición de *deepfake* como ‘’un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares u otras

entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos'' (art. 3.60 Resolución Legislativa del Parlamento Europeo P9_TA(2024)0138). En esta definición dada por el legislador europeo encontramos, además de los rasgos esenciales ya señalados, dos elementos innovadores respecto de los caracteres principales de los *deepfakes* señalados anteriormente.

El primero de ellos es que se introducen en su ámbito todo tipo de contenidos multimedia (al igual que los autores citados parecen excluirse los productos de modelos de IA generativos de texto) en tanto en cuanto se asemejen de forma realista a personas, pero también a objetos o lugares, e incluso a ''otras entidades o sucesos reales''; una noción que parece alinearse en mayor medida con la definición de Van der Sloot y Wagensveld (2022), que no acota el término en atención al objeto que estos contenidos representan y, en consecuencia, ofrecen un concepto mucho más amplio que los demás autores, que en casi su totalidad exigen como rasgo esencial la representación de personas. Parece acertada la inclusión en una definición legal de las ultrafalsificaciones de representaciones realistas pero ficticias de objetos y lugares, pues este tipo de imágenes o vídeos podrían tener especial trascendencia en ciertas manifestaciones de delincuencia económica que quedarían fuera del ámbito conceptual de los *deepfakes* si centramos estas exclusivamente en las representaciones de rasgos de la personalidad de los humanos.

En un segundo lugar encontramos como elemento innovador la mención a que estas ultrafalsificaciones pueden ''inducir a una persona a pensar erróneamente que son auténticos o verídicos'' (art. 3.60 Resolución Legislativa del Parlamento Europeo P9_TA(2024)0138). Se trata de un elemento innovador porque, a diferencia de los distintos autores citados, que se refieren al realismo de las imágenes, y que el Reglamento incluye igualmente al exigir que estas se ''asemejen'' a personas u objetos; en este caso constituye la inclusión de un criterio normativo adicional referido a la capacidad de engaño que estas representaciones realistas poseen. Esto es, exige que estas representaciones no sean meramente realistas, sino que es necesario que, además, puedan producir el engaño en el espectador, la creencia en la realidad del evento o elemento observado; con lo que parece que se ha tratado de aludir a una dimensión cualitativamente distinta del realismo en sí mismo de la imagen o vídeo, pues se diferencia esta capacidad de inducir al pensamiento de que es real de la semejanza a la persona u objeto representado.

Desde esta perspectiva, podría entenderse que el Reglamento europeo incluye como rasgo esencial del concepto de *deepfake* no solo el ''realismo'', sino también la ''credibilidad'' de lo que se muestra, la aptitud para producir en el público la creencia de que efectivamente ha tenido lugar aquello que se observa, que una persona determinada ha dicho lo que se oye o es tal y como se ha visto. Esta distinción sólo parece entreverse en la definición dada por Rodrigues (2023), que entiende que estas ultrafalsificaciones deben tener el fin de inducir a una falsa percepción de una persona.

Así entendida, la noción de ultrafalsificación dada por el mencionado Reglamento nos llevaría a afirmar que la imagen presente en la Figura 1, en la que aparecen los líderes de los cinco mayores partidos políticos que concurrían a las elecciones generales españolas del 10 de noviembre de 2019 representados en la escena introductoria de la serie estadounidense ''El Equipo A''; no podría considerarse un auténtico *deepfake*.

Figura 1. Captura del Vídeo ‘’El Equipo E, con E de España”, por Canal FaceToFake (2019). Youtube.

Figura 2. Iglesias y Díaz, por United Unknown (2023). United Unknown Guerrilla Visual.

Figura 3. The Pope Drip, por MidJourney (2023). Reddit.

Sin embargo, la Figura 2, en la que aparecen los políticos Pablo Iglesias y Yolanda Díaz caminando juntos, siendo una imagen ficticia que apareció en la portada de uno de los mayores periódicos españoles, sí podría corresponderse plenamente con el concepto comunitario de ultrafalsificación.

En el caso de la Figura 3 la cuestión es más controvertida, pues pese a que por lo general no existirían dudas sobre la falsedad de una imagen del Papa Francisco de estas características, no existe nada que a priori pueda llevar al convencimiento de que la imagen no es real. La situación que muestra podría ser creíble, pese a que pueda generar en un público que conoce y contextualiza la figura de la persona que representa dicha imagen cierta extrañeza y resistencia a asumir que es una fotografía auténtica. En definitiva, el problema que generaría este tipo de imágenes es que para los conocedores

de la persona, lugar u objeto representado producirán una sensación contradictoria que les llevará a pensar con mayor probabilidad que no se trata de una imagen real; mientras que aquellos otros ajenos a la persona representada no dispondrán en principio de conocimientos suficientemente profundos de la persona que permita discernir la veracidad de la representación realizada, por lo que sería un ámbito de dudosa calificación como *deepfake*.

A la luz de los tres ejemplos mostrados puede llegarse a una conclusión básica, y es que aquello que se entiende en un sentido amplio y general como *deepfake* puede diferir sustancialmente con la noción normativa y jurídicamente relevante que de estos deba manejarse. Conforme a la primera acepción, en un sentido general, las tres imágenes referidas serían ultrafalsificaciones. En general, puede entenderse que todo tipo de producciones audiovisuales generadas con IA en que se representan a personas o elementos reales en situaciones o de formas ficticias con cierta apariencia de veracidad, o al menos un determinado grado de realismo, se corresponde con una noción de ultrafalsificación *lato sensu*. Pese a ello, en el ámbito jurídico se deberá trabajar por la obtención de un concepto de *deepfake* más restringido, especialmente en sectores normativos que restrinjan o impongan obligaciones a los creadores o prestadores de servicios relacionados con la IA y en un sentido máximo cuando se trate del Derecho penal; limitando así la actuación del Ordenamiento Jurídico a los casos en que realmente este tipo de producciones audiovisuales puedan presentar cierta relevancia jurídica. Este intento pasa necesariamente por la inclusión de ciertos elementos, criterios y requisitos de naturaleza normativa, que perfilen un concepto de *deepfake* en un sentido jurídicamente estricto en el que, entendemos, el realismo, la credibilidad y la dificultad de detección con medios accesibles, sencillos y rápidos, deben jugar un papel esencial como núcleo de la problemática de las ultrafalsificaciones.

Es por ello que parece positivo el intento del legislador europeo al introducir en la definición legal de *deepfake* la nota de que estos deben poder inducir al engaño como criterio adicional al mero realismo de la imagen, vídeo o sonido. Sin embargo, que avanzar en este sentido sea positivo no supone en absoluto que esta formulación dada a nivel comunitario esté exenta de posibles precisiones críticas. Concretamente, que la literalidad del la llamada Ley IA se refiera a que el *deepfake* ‘’puede inducir a una persona a pensar erróneamente que son auténticos o verídicos’’ (art. 3.60 Resolución Legislativa del Parlamento Europeo P9_TA(2024)0138) produce dudas acerca de si este criterio supone en realidad una barrera conceptual adicional o si, en realidad, al no incluir un baremo normativamente objetivador, lo que genera es la perversión de esta idea de proporcionalidad y de limitación de la aplicación de la norma a situaciones en que verdaderamente el acto sea jurídicamente relevante.

Con esto pretendemos señalar que prescindir de cualquier referencia a un elemento de valoración normativo del grado de engaño que estas imágenes pueden producir, sin hacer referencia explícita a un observador medio, o sin tener en cuenta las circunstancias y la credibilidad media suelen recibir contenidos en el medio en que esta imagen se ha difundido, sin establecer una mención similar al ‘’engaño bastante para producir error en otro’’ (art. 248 CP) que emplea el legislador penal español en materia de estafas; se corre el riesgo de que la exigencia de que la producción pueda inducir a otros a pensar erróneamente sobre la veracidad de lo que observan opere en un sentido inverso al deseado, expandiendo el concepto a ciertos contenidos audiovisuales que no parecerían reales para la mayoría, pero que puede entenderse que podrían llegar a inducir en alguna persona un error por el mero hecho de ser, hasta cierto punto, realistas; cuando precisamente lo que entendemos que se perseguía era excluir las producciones que siendo

meramente realistas son abierta y claramente ficticias. Por continuar con los ejemplos dados, consideramos que esta definición podría tender a incluir como *deepfake* la imagen de la Figura 3 por poder producir en abstracto un riesgo de que alguien la confunda con una fotografía real, pudiendo entrar entonces en el ámbito de aplicación de las normas reguladoras de este tipo de producciones. Y si bien es posible que en determinados sectores jurídicos esto no sea problemático e incluso pueda ser deseable, cuando se trate del ámbito sancionador administrativo o penal ello resultaría inadmisible.

En este sentido, quizá la sobreinclusión en el concepto de ultrafalsificación en el ámbito del Reglamento europeo de IA (Resolución Legislativa del Parlamento Europeo P9_TA(2024)0138) no sea excesivamente problemático, dado que la consecuencia explícitamente prevista en él es la del sometimiento a su art. 50.4, en virtud del cual los responsables del despliegue del programa que permita la producción de este tipo de contenidos audiovisuales deberán especificar que los mismos han sido modificados por un sistema de IA, excluyéndose esta obligación en casos legalmente previstos para la investigación penal y limitándose en el caso de obras creativas, satíricas, artísticas o de ficción. Sin embargo, y por las razones que se han apuntado, parece necesario ofrecer una conceptualización de *deepfake* que, incluyendo los distintos elementos y rasgos característicos mencionados, sí asegure la limitación de la respuesta jurídica sancionadora mediante la inclusión de criterios y estándares normativos que circunscriban la relevancia jurídica a aquellos casos en que realmente exista, desde una perspectiva objetiva, un cierto riesgo derivado de su clara credibilidad e imposibilidad o dificultad de verificación.

Por ello la propuesta que realizamos es la de adoptar una concepción de *deepfake* por la que como tales se entiendan todo tipo de contenidos audiovisuales generados por medio o con asistencia de un sistema de IA, en los que se representen lugares, objetos o personas reales realizando actos o en situaciones ficticias pero con una apariencia de realidad tal que esta representación resulte creíble para un espectador medio conforme a los estándares y circunstancias del ámbito en que se difunde y cuya falsedad no pueda verificarse fácil y accesiblemente mediante las herramientas disponibles para dicho espectador.

Así se tratará de una definición compuesta por cinco elementos esenciales. El primero de ellos es las ultrafalsificaciones se materializan en todo tipo de contenidos audiovisuales, lo que incluirá tanto imágenes como vídeos y sonidos, así como cualquier combinación de los anteriores. En segundo lugar, encontramos que se tratarían de contenidos que se han generado con la intervención en algún momento del proceso de un sistema de IA. En tercer término, estos *deepfakes* pueden consistir en representaciones tanto de objetos o lugares, lo que permitirá la atribución de relevancia jurídica a estos contenidos, por ejemplo, en materia de cierta modalidad de delincuencia económica relacionada con la transmisión de productos o bienes cuyo valor pueda falsearse por medio de estas técnicas; como de personas, pudiéndose optar por entender que existe una representación de una persona real cuando en ella es reconocible de forma inequívoca ‘’alguno de sus atributos más característicos, propios e inmediatos, como son la imagen física, la voz o el nombre, cualidades definitorias del ser propio y atribuidas como posesión inherente e irreductible a toda persona’’ (STC 117/1994, FJ 3), elementos que, según nuestro Tribunal Constitucional (TC) conforman el contenido del derecho fundamental a la propia imagen. En cuarto lugar, estas imágenes, vídeos o sonidos deberán ser, para considerarse como *deepfakes*, realistas, entendiendo por tal su grado de coherencia visual o la calidad y fidelidad de la representación respecto a la persona u objeto real que se representa. Finalmente, como quinto elemento se introduce el criterio de la credibilidad, matizada mediante ciertos requisitos para la valoración normativa

desde una tendencia a la objetivación de su capacidad de engaño en un público medio y conforme al contexto en que este contenido se presenta.

Debe señalarse, para finalizar, que esta concepción de ultrafalsificación, al igual que parecen realizar tanto la ofrecida por el legislador europeo como las conceptualizaciones de la mayor parte de los autores citados, excluye de su ámbito conceptual aquellos contenidos generados por medio de la IA en que se representan lugares, objetos o personas completamente ficticios, que no existen en realidad pese a que los datos que el sistema emplea para la generación de la imagen o video sí contenga imágenes de personas o entidades reales. Desde esta perspectiva, a mero título ejemplificativo, los resultados publicados de Sora, el nuevo modelo generativo de OpenAI (2024), en que se observan imágenes totalmente sintéticas en que no se representan a personas ni lugares auténticos sino completamente generadas por el sistema de IA; no tendrían cabida en el concepto de *deepfake* en tanto no se emplee generando la imagen de un elemento real. En definitiva, todo lo expuesto hace llegar a la conclusión de que en un sentido normativo no todos los productos generados por modelos generativos de IA son realmente ultrafalsificaciones, pese a que en un sentido amplio pueda identificárseles como tales.

2. ¿Estamos asumiendo que la vaca es esférica? Sobre la necesidad de partir de un modelo de criminalización para abordar la cuestión jurídico penal sobre las ultrafalsificaciones

Sea cual sea la definición definitiva por la que se opte, no es de extrañar, dados los ejemplos indicados más arriba, que distintas voces hayan señalado la necesidad de garantizar que el Ordenamiento Jurídico sea capaz de ofrecer una respuesta eficaz y coherente frente a estas amenazas, tratando de evitar espacios de impunidad frente a conductas inaceptables cometidas mediante el empleo de sistemas de IA y concretamente, por lo que aquí nos ocupa, a través de la generación de *deepfakes*. Una reflexión similar llevó a Agustina (2021) a afirmar que el ciberespacio impone la necesidad de desarrollar cambios dogmático-penales y la mutación de ciertas figuras de la teoría del delito que permitan una respuesta penal satisfactoria frente a las nuevas conductas posibilitadas por el surgimiento de este nuevo entorno. Desde la perspectiva de la creación de pornografía ficticia mediante IA, Simó (2023) considera necesario configurar un modelo que permita reaccionar integralmente contra estos fenómenos, encontrándose entre las medidas a adoptar, para la autora, la inclusión de un nuevo tipo en el Código Penal que permita su castigo como delito contra la intimidad dado que, entiende, estas conductas no podrían subsumirse en ninguno de los tipos del art. 197 CP. Por su parte, Rodrigues (2023) destaca la necesidad de expandir o resignificar el concepto de libertad sexual como una consecuencia lógica de las exigencias de protección de un Estado de Derecho democrático para dar lugar a un bien jurídico de dignidad sexual al que afectarían los *deepfakes*, pues perjudicarían al derecho a disponer de la propia imagen sexual, entendiendo que sería legítimo y justificado el castigo de las ultrafalsificaciones pornográficas como un delito de peligro abstracto.

En los casos indicados subyace una idea común: son necesarios cambios para poder castigar estas conductas y soslayar los obstáculos que pueden dificultar la aplicación de las normas penales a estos nuevos supuestos de hecho. Sin embargo, esta idea esconde implícitamente un problema distinto al de la posible impunidad y las formas para evitarla, y es que en ellas parece partirse de la presunción de que las conductas indicadas deben conllevar responsabilidad penal.

Existe una broma metafórica, referida tradicionalmente al campo de la Física, en la que se afirma que un estudioso de esta disciplina, para dar una solución a un problema planteado por un granjero, confecciona un modelo teórico que, sin embargo, tan solo funciona suponiendo que una vaca es una esfera. Esta historia pretende señalar cómo a menudo se tiende, desde las disciplinas científicas, a confeccionar teorías y modelos en que se sobresimplifican o pasan por alto ciertos problemas o circunstancias para poder dar solución a una cuestión. Salvando las distancias, cabría preguntarse si desde el ámbito de la doctrina penal no se está igualmente suponiendo que una vaca es una esfera cuando presupone que ciertos casos de elaboración de *deepfakes* deben ser criminalizados y discute si esto es posible con la teoría del delito y los tipos vigentes e indaga en las vías que posibilitarían en mejores términos su castigo.

Está claro que ciertos actos realizados en que se encuentran involucradas ultrafalsificaciones pueden ser excepcionalmente dañinas, como los autores citados no dudan en señalar, y resulta lógico que se busquen medios para vehicular el reproche penal. Pese a ello, un razonamiento en este sentido, sin más justificación que lo problemático de algunos supuestos, supondría partir de una base que no ha sido justificada. Habríamos omitido el paso previo, que es la fundamentación de la legitimidad del castigo y la reflexión sobre la naturaleza de estas conductas, lo que entraña el riesgo de castigar injustificadamente acciones novedosas que no pudieron tenerse en cuenta por el legislador, el de hipertrofiar los códigos con figuras penales simbólicas para responder a conductas que quizá podrían castigarse desde tipos existentes y, a la postre, de desatender ciertos rigores y garantías penales motivados por lo trágico de la casuística.

Lo que se pretende afirmar con lo anterior es que del hecho de que esta nueva forma de acción produzca un resultado similar o incluso idéntico al de un tipo preexistente no se sigue de modo automático que dicha conducta pueda castigarse como tal delito, pues olvidaríamos así el carácter valorativo del examen del acto para comprobar la tipicidad de este desde una perspectiva normativa y caeríamos en una noción naturalista del delito. Por otro lado, que si ignoramos que debe justificarse y fundamentarse la razón por la cual una conducta debe ser castigada penalmente quizá transgredamos ciertos principios y garantías que entendemos que deben regir el sistema penal, y es que de la comprobación de que una conducta no puede ser castigada con un tipo penal existente tampoco se colige la conclusión de que deba castigarse dicha conducta, lo cual rozaría, si no es que abraza de lleno, la falacia naturalista; se requiere para afirmar tal cosa una argumentación sustentada en aquellos criterios y límites que determinan cuándo es legítimo que actúe el Derecho penal.

Puede entreverse entonces que, aparentemente, la cuestión jurídico-penal sobre los *deepfakes* se bifurca en dos problemáticas. La primera, la relativa a la posibilidad o no de castigar los actos de generación y/o difusión de estos contenidos desde una figura delictiva existente. La segunda la de si es posible y legítima la tipificación de nuevos delitos para castigar estos fenómenos. Creemos, sin embargo, que en realidad la respuesta a ambas preguntas requiere un examen de estas conductas a la luz de un modelo de criminalización que delimite el alcance del ius puniendi de un Estado social y democrático de Derecho; y si bien respecto de la segunda cuestión lo que se ha afirmado es evidente, sí parece necesario explicar esta posición respecto de la primera de ellas.

Todo intento de subsumir una conducta novedosa, que representa un fenómeno con una entidad disruptiva para el ser humano, en un tipo penal preexistente debe tener en consideración en todo momento cuáles son los criterios que determinan qué clase de actos pueden ser castigados penalmente y asegurarse la consonancia de esta conducta que

se pretende castigar con alguno de dichos criterios porque existe una interrelación estrecha entre estos fundamentos de punibilidad y aquello que se puede considerar como una conducta típica. Esto es así porque con la tipificación de una conducta este sustrato fáctico que es la acción se normativiza, se le dota de un significado de desaprobación. Pero esta normativización sólo puede partir del entendimiento de que dicho tipo de conductas pueden ser castigadas por corresponderse con un criterio de criminalización admisible en la comunidad en que se produce. Por consiguiente, en algún lugar dentro del conjunto de significado que es la acción típica debe encontrarse un punto de anclaje con dicho criterio que lo apuntale como una figura delictiva legítima, pues todos los tipos existentes deben estar alineados con los criterios de criminalización que delimitan el alcance del ius puniendi. Todas estas acciones típicas, en consecuencia, deben en el fondo presentar un carácter que se corresponda con el criterio por el que es admisible su castigo como delito; y todo acto cuya tipicidad pretenda afirmarse deberá constituir, en primer lugar, un acto que se alinee con dicho criterio además de poseer el resto del significado de desaprobación de dicho tipo específico.

Señalar esta relación entre ambos ámbitos viene a remarcar que tanto la remisión a un criterio de criminalización admitido que permite la existencia legítima de un tipo penal como las consideraciones normativas sobre la desaprobación del riesgo que genera una conducta determinada forman parte de un mismo paso en el análisis del delito. Ambas se encuentran interrelacionadas y no pueden ser obviadas para afirmar que una conducta es típica. Por ejemplo, si podemos castigar el homicidio es porque el matar a otro constituye un daño. Por otro lado, está asentado que no todas las formas por las que causal o ‘’naturalmente’’ se produce la muerte a otro constituyen una conducta típica, sino que sólo aquellas muertes que se produzcan por la generación de un peligro jurídico-penalmente desaprobado que se concreta posteriormente en el resultado lesivo podrán considerarse como verdaderas conductas típicamente homicidas; pero también debe tenerse en cuenta que todas estas conductas típicas, para serlo, deben generar aquello que se entienda por daño, pues es el requisito necesario para que pueda existir, en primer lugar, tipo. Se concluye entonces que no todas las muertes son homicidios pero que todos los homicidios han de ser daños. Entonces, la significación normativa de desaprobación de una conducta gira en torno al elemento que permite el castigo legítimo de la misma en línea de principio.

Por consiguiente, según esta perspectiva, todo examen de tipicidad debe pasar previamente por la comprobación de que la conducta genera aquello que deba entenderse por el criterio de criminalización al que remite el tipo; y siempre y cuando el resto de la significación normativa de desaprobación que da el tipo esté presente en la misma. Si nos adentramos en este segundo paso, el de contrastar la exteriorización de la acción con un tipo delictivo, sin tener presente la cuestión de su entidad como daño o, en su caso, ofensa se corre el riesgo de castigar más allá de lo legítimamente admisible en nuestro sistema por aplicar un tipo sobre una conducta que no se corresponde con el género de acciones al que la descripción típica se refiere. Por ello, tanto los intentos por subsumir los fenómenos relacionados con las ultrafalsificaciones en tipos existentes como las respuestas a la cuestión sobre si es posible la creación de delitos ex novo para su castigo deben edificarse y sustentarse sobre un modelo de criminalización y sobre la exploración del significado y carácter más íntimo de estas nuevas actividades.

En este sentido, desde la perspectiva que aquí mantenemos, entendemos que el procedimiento para el abordaje jurídico-penal de los *deepfakes* debe ser como sigue. En primer lugar ha de partirse de contrastar estas acciones relacionadas con la IA con un modelo de criminalización en que se determinen qué clases de actos pueden ser

legítimamente castigados por el Estado, pues, de nuevo, si no se realiza previamente y se constata que efectivamente se correspondan con alguno de ellos se corre el riesgo de castigar injustificadamente una conducta, ya que todo tipo que pretenda aplicarse parte necesariamente del respeto a los criterios de criminalización admisibles.

Es en segundo lugar cuando, una vez constatado que estas conductas pueden considerarse efectivamente daños, ofensas o aquello que se entienda legítimamente castigable, se deberá comprobar si estas se alinean en su significación normativa con alguno de los tipos ya existentes, pues aventurarse a la creación ad hoc de nuevos tipos sin este paso previo podría llevar a duplicidades y a una hipertrofia de las partes especiales de los Códigos Penales.

Por último, cuando se comprobase que, constituyendo una conducta que podría ser lícitamente atacada por el ius puniendi estatal, no existe ningún tipo en el Código que podría ser aplicado al caso será posible, aunque no estrictamente necesario, y siempre con el debido y escrupuloso respeto del principio de intervención mínima y ultima ratio, el establecimiento de nuevos tipos delictivos.

Si bien sobre esta reflexión se volverá más adelante y, creemos, quedará más clara, creemos que ahora, para ahondar en este razonamiento, puede ser positivo el planteamiento de un ejemplo. Imaginemos que, como el mismo Borges (1949) haría en su propio relato, encontramos un Aleph, esto es, un punto del espacio que contiene todos los puntos y que, por consiguiente, permite observar toda la existencia de modo simultáneo. Al mirar al Aleph tendríamos la capacidad de desvelar todos los espacios de intimidad de toda la humanidad, de explorar los domicilios sin consentimiento de sus titulares, de conocer todos los secretos y ver todos los momentos y estados de todas las personas, incluso de acceder a la correspondencia privada, como haría el propio autor-personaje, Borges, al leer las cartas que Beatriz escribió a Carlos Argentino. ¿Habríamos entonces cometido algún tipo de delito? ¿Sería necesario incluir una nueva figura delictiva que castigue observar tal fenómeno?

Si constatásemos que efectivamente el Aleph ofrece la visión de todo cuanto es y en tanto tal es o, perplejos ante ello, no podemos sino meramente admitir que se trata de la creación de una imagen irreal de todo lo que existe, pero con una gran fidelidad y precisión; sólo podremos afirmar que se comete acción típica en tanto la misma entrañe la naturaleza del acto que se desvalora normativamente. Esto es, sólo si genera el peligro jurídicamente relevante —y si se une al resultado desaprobado ex post cuando sea necesario, claro está— o constituye la actividad normativamente desaprobada podemos subsumirla en el tipo existente. Pero debemos justificar el porqué de que la observación de un punto del espacio que contiene todos los puntos supone una intromisión en la esfera de intimidad de una persona, y en esta justificación deberá encontrarse de algún modo la razón por la que se entiende que esta acción constituye un daño u ofensa jurídico-penalmente relevante. En caso de encontrar que desde esta perspectiva no pudiésemos subsumir dicha visión en alguno de los tipos penales existentes en el momento de hacerlo por el defecto de algún elemento típico, la propuesta de introducir una nueva figura delictiva no puede fundamentarse en la impunidad que se derivaría de no hacerlo, sino que precisaremos una argumentación de los motivos que hacen necesaria tal criminalización y que la misma respeta los principios y garantías que exigimos al Derecho penal de un Estado social y democrático de Derecho.

Entendemos que, de algún modo, la IA ha producido un fenómeno similar a un Aleph. Nos ha ofrecido un medio para crear imágenes altamente realistas de lo que

deseemos, aunque en este caso sepamos que se trata de una representación generada artificialmente a partir de múltiples datos e imágenes. Si bien es cierto que ‘’la manipulación de fotografías, textos, audios o vídeos con múltiples finalidades no es un fenómeno nuevo’’ (Simó, 2023, p.495) sí tiene este nuevo mecanismo una entidad novedosa derivada de la instantaneidad en la generación de las imágenes y la democratización del acceso a la herramienta para ella, así como, sobre todo, la del ultrarrealismo que presenta; y estas particularidades conllevan que sus efectos tengan también una entidad mayor, como remarcan Simó (2023), Rodrigues (2023) o Citron y Chesney (2019) —la persistencia del daño, el anonimato en la generación de los contenidos, entre otras circunstancias—. Esto hace que las ultrafalsificaciones se presenten como una realidad cualitativamente distinta a las manipulaciones plásticas o digitales que se pudieren realizar anteriormente, una acción nueva. Por ello no puede darse por hecha la reflexión sobre su naturaleza jurídico-penal más profunda, sobre su carácter de daño o no, primero, y sobre su posible correlación de significado normativo con un tipo específico, después. Y esta labor doctrinal tampoco puede soslayarse simplemente con, una vez remarcados los elementos que pueden producir que los problemas generados por las ultrafalsificaciones sean de una gran entidad, proponer la reformulación de ciertos bienes jurídicos para permitir el castigo, pues se presentaría un problema de circularidad conforme al cual debe criminalizarse dicha conducta porque afecta a un bien jurídico que debe entenderse de un modo distinto para que pueda incluir los perjuicios causados por este tipo de conductas que se pretende criminalizar. En otras palabras, no podemos asumir que una vaca es una esfera.

En el fondo de este asunto, como ya se ha ido señalando tácitamente, se encuentra la cuestión relativa a la delimitación del riesgo permitido y el desvalor o desaprobación de una acción; de qué conductas deben aceptarse, aunque entrañen cierto riesgo, y cuáles deben —o pueden— ser castigadas. Esto debe canalizarse, en definitiva, por un proceso argumentativo que lleve a una u otra opción, un proceso en que se desarrolla una valoración de tal riesgo o perjuicio causado, pero el cual, por lo que ya hemos argumentado, se sustente a su vez en los límites del Derecho penal, en los criterios de criminalización que consideramos admisibles; pues sólo podremos afirmar que este debe actuar frente a una acción si lo hacemos de forma coherente con el espacio de actuación posible que se le atribuye. Sólo puede argumentarse que una conducta posee un significado normativamente desaprobado con trascendencia jurídico-penal, ya esté desaprobado por un tipo existente o deba procederse a su tipificación, si esta atribución de desvalor se apoya en un criterio de fondo que permite legítimamente desaprobar jurídico-penalmente la conducta, precisándose entonces partir de una afirmación de cuáles son estos criterios y justificar debidamente por qué la conducta observada se corresponde con alguno de ellos.

Es por todo ello que entendemos justificado el interés por abrir líneas de investigación que aborden la cuestión desde el prisma de un modelo de criminalización que sirva de base desde el que analizar la posible punición de esta nueva fenomenología de forma respetuosa con los principios que deben regir el ejercicio legítimo del ius puniendi en un contexto axiológico determinado. Se defiende aquí, por tanto, no una perspectiva tendente a explorar cómo deben mutarse ciertas figuras doctrinales generales para responder a esta nueva fenomenología —del modo que hace Agustina (2021)—; sino la de cuestionarse desde estas líneas teóricas y principios si es posible y necesaria la modificación de ciertas figuras delictivas para adecuarla a esta realidad o si las existentes ya pueden solventemente dar cuenta de la misma.

3. Trazando la hoja de ruta: ¿desde qué tipo de modelo de criminalización debemos aproximarnos a las ultrafalsificaciones?

El problema que aquí nos planteamos no es únicamente sobre el fundamento de criminalización de un tipo determinado de *deepfakes*, como podrían ser los pornográficos. Al contrario, lo que tratamos de lograr cierto criterio que permita analizar sistemáticamente la legitimidad del castigo de las distintas conductas relacionadas con las ultrafalsificaciones que pueden presentar relevancia penal. Esto sólo podrá realizarse partiendo de un modelo general que determine las líneas básicas de criminalización que permita el examen, a la luz de las problemáticas y riesgos específicos que un tipo de ultrafalsificación en particular genera, de las propuestas de criminalización que se realicen. En definitiva, debemos indagar en los motivos que, con carácter general, habilitan la movilización de la acción punitiva del Estado de forma legítima.

Sin embargo, debemos preguntarnos cómo ha de ser y qué elementos ha de integrar aquello a lo que aquí nos referimos como modelo de criminalización. Tradicionalmente se ha adoptado una perspectiva conforme a la cual la legitimidad de la intervención penal se hacía depender, tanto en la tradición jurídica anglosajona del *common law* como en la continental, de un instrumento de carácter sustantivo (Miró-Llinares, 2024), a saber, el principio de daño, en el caso de la doctrina angloamericana, o el bien jurídico en Alemania (Hirsch, 2007), sirviendo estos conceptos como fundamento del castigo legítimo por el Estado. Pese a ello, en el ámbito continental se ha destacado cómo la teoría del bien jurídico —cuya relación con el principio de daño se abordará más adelante— no ha sido una herramienta eficaz para trazar límites a la intervención penal estatal. Como ya afirmó Silva (2001):

> ‘’No es posible controlar ley penal alguna desde la perspectiva de una hipotética vulneración del principio de exclusiva protección de bienes jurídicos […] no se niega que la persistencia en la afirmación de que el Derecho penal debe proteger exclusivamente bienes jurídicos puede manifestar una cierta actitud de los autores proclive a la permanente revisión de los presupuestos de la ampliación del círculo de objetos de protección del Derecho penal. Pero sí se insiste en que no cabe asignar a la idea de bien jurídico una trascendencia que, desde luego, no alcanza el concepto tal como se ha producido su desarrollo histórico y tal como es su configuración actual’’ (p.123)

Para Hörnle (2019) el problema del modelo basado en el bien jurídico protegido es que para la autora, siguiendo a Duff (2018), constituiría un *thin principle*, un criterio o principio que, frente a un *thick principle*, no presenta un alto nivel descriptivo en cuanto a la sustantividad de su contenido, sino que destaca por su carácter abstracto y, por consiguiente, por la posibilidad de incluir dentro de su ámbito de cobertura un mayor número supuestos —aquí optaremos por una traducción libre de *thin* y *thick principles* como principios blandos o flexibles y duros o rígidos en atención a lo concretamente descrito que se encuentre su contenido—. En este sentido afirma que el bien jurídico ‘’is not only a thin concept, but also so thin as to be an empty concept. The notion of «good» does not give any guidance at all —every state of affairs could be labeled this way’’ (Hörnle, 2019, p.211).

De este modo, concebir una teoría basada en un solo principio sustantivo, como el de exclusiva protección de bienes jurídicos o el del daño, habría sido insuficiente para su labor de constreñir la acción penal del legislador. La razón de esta insuficiencia podría radicar en que estos modelos sobresimplifican su formulación y tratan de sintetizar y

combinar en un mismo nivel argumentos de muy distinta naturaleza e incluso incompatibles (Miró-Llinares, 2024) para tratar de responder a una variedad de conductas y realidades muy amplia (Hörnle, 2019) que no puede ser abarcada por un principio tan simplificado. Ahora bien, debe tenerse en cuenta que el problema que se señala aquí no es el de que la sobresimplificación del principio produzca una excesiva sustantividad de manera que no pueda dar debida cuenta a tan diversa realidad. Por el contrario, la problemática deriva de que el principio o criterio resultante es demasiado flexible o blando; la sintetización excesiva por la que se sobresimplifica el criterio lo dota de demasiada abstracción y por ello posee muy poco contenido sustantivo y, por lo tanto, un escaso o nulo poder discriminador, permitiendo al legislador su modelación para adecuarse a las más heterogéneas de las conductas. Es esta dinámica la que esconde la crítica de la excesiva simplificación señalada por Miró-Llinares (2024). Y es que, como afirma Duff (2018), cuanto más duro o sustantivo sea un principio o criterio menos inclusividad presentará, y por tanto menos conductas permitirá afirmar que son legítimamente reprimibles por el Estado. Por ello, según Hörnle (2019), la protección de la dignidad humana como único criterio sustantivo de criminalización —aunque quepa discutir su carácter duro y preciso— puede ser eficaz para explicar la legitimidad de algunas figuras delictivas, pero para otras como los delitos fiscales sería poco convincente.

Es por ello que concluye Miró-Llinares (2024) que ‘’the harm principle or the legal interest cannot be configured as a master principle or a silver bullet to determine the scope of criminal law’’ (p.12), precisamente porque tiene una nula capacidad para la exclusión de propuestas de política criminal, permitiendo que se extienda el alcance del Derecho penal tanto como el legislador desee. Ante esta conclusión podría pensarse que la solución pasa por volcarse en la búsqueda de un modelo cuyo carácter sustantivo se plasme y desarrolle de forma más precisa y dura. En definitiva, concluir que ‘’if scholars strive to have some more impact on legislative activities, they should strive for more substantive content’’ (Hörnle, 2019, p.211) pues ‘’if it were to guide decision-making in an effective way, we would need a *thick* theory of criminalization’’ (Hörnle, 2019, p.209).

Efectivamente, parece que cerrar filas en torno a un criterio o teoría que de forma taxativa delimite claramente qué puede ser objeto de intervención penal y qué no parece la solución ideal para lograr la reinstauración del que Silva (2001) refiere como *gutes, altes liberales strafrecht*, el viejo y buen Derecho penal liberal. Sin embargo, como este mismo autor o Miró-Llinares (2015) han destacado, esta solución puede llevar consigo efectos indeseables.

Debe tenerse en cuenta que la tendencia a la inflación penal no está causada por la teoría del bien jurídico protegido, a la que sólo puede recriminársele no haber podido refrenarla. Por el contrario, esta tiene su origen en una realidad material compleja, propia de las dinámicas económicas y sociales del capitalismo postindustrial, que ha producido un cambio de paradigma en el modo de concebir el papel del Derecho penal, el cual ha experimentado un giro preventivo desde el que se ha exigido de este protección ante las incertidumbres de la llamada sociedad del riesgo, donde otras instituciones sociales y jurídicas tendentes a dar salida a los conflictos sociales y ordenar la vida en comunidad han perdido su eficacia, dando lugar así a propuestas punitivas cada vez más extensas y sobre ámbitos que anteriormente no se entendían merecedores de intervención penal alguna (Silva, 2001). Por consiguiente, el establecimiento de un criterio sustantivamente perfilado de forma rígida y determinada no evitará la aparición de pulsiones sociales punitivistas que se traduzcan en propuestas legislativas tendentes a la hipertrofia del sistema penal, las cuales seguirán proliferando, pues estas no surgen ociosamente del

legislador, sino que tienen su origen en unas circunstancias materiales propias de las sociedades de nuestro tiempo que desarrollan esta tendencia sociopolítica.

Es ingenuo pensar que entonces el legislador se plegará plenamente ante el límite teórico impuesto a la intervención penal y se resignará a no proceder penalmente ante estas nuevas circunstancias. Por el contrario, es más creíble pensar que ante esta limitación se tienda a desvirtuar y pervertir el contenido del criterio impuesto o la naturaleza y cualidades de la conducta que se trata de criminalizar con la finalidad de alinear ambos elementos y permitir el despliegue punitivo, como de hecho ya tuvo lugar en Estados Unidos al debatirse sobre la criminalización del discurso del odio, tratando de afirmar que este tipo de expresiones entrañan un daño propiamente dicho, elemento que en la tradición anglosajona se ha visto como exigencia exclusiva para la legitimidad del castigo, y con ello dando lugar a una idea de daño ambigua que abre la puerta al castigo con una amplitud mayor que la que se habría logrado incorporando un criterio adicional que permitiera el castigo de actos ofensivos siempre que cumplan ciertos requisitos rigurosos y sólo permitiendo la imposición de cierto tipo de penas no privativas de libertad (Miró-Llinares, 2015).

En definitiva, optar por un solo principio sustantivo pretendidamente *thick* en un contexto social caracterizado por una amplia demanda punitiva corre el riesgo de inducir que este principio sea relativizado y desdibujado, perdiendo así el valor sustantivo que se le pretende atribuir y llevando al castigo de cierto tipo de conductas desde su definición como aquello que no son, justificando así frente a ellas el recurso a penas propias de un denominado núcleo duro del Derecho penal —para el que Silva (2001) reserva las máximas garantías exigidas por la ciencia penal clásica, precisamente por la gravedad de las penas que impone— al que, en puridad, no pertenecen.

Como muy concretamente sintetiza Hörnle (2019), ‘’the range of justifications for criminal norms cannot be reduced to one, substantive, thick master value that guides decision-making in a clear-cut, uncontroversial way’’ (p.210). Y aunque la autora afirme lo anterior con base en un argumento distinto al que aquí se ha expuesto —a saber, el de la mera amplitud y heterogeneidad de las conductas cuyo castigo debe legitimar— la conclusión es la misma: no es útil un solo principio flexible o blando porque tiene nula capacidad para limitar la intervención penal estatal, pero tampoco es deseable conformar un modelo de criminalización basado en un principio sustantivo duro, pues ello puede conllevar efectos igualmente nocivos dado el contexto socioeconómico en que nos encontramos y las exigencias de protección penal que de él derivan.

Debemos entonces cuestionarnos sobre qué clase de modelo debemos tratar de buscar que supere estos problemas. Para ello debemos tener en cuenta que la raíz de la debilidad de estos modelos señalados se encuentra en su elemento común: tratan de delimitar qué puede castigarse por medio de un sólo criterio. Esto puede identificarse con el que Marshall y Duff (como se citaron en Miró-Llinares, 2024) definen como primer enfoque para configurar una teoría de criminalización, uno en que para responder a la pregunta «¿qué debemos castigar?» se busca un principio sustantivo ‘’or an alternative combination of them’’ (Miró-Llinares, 2024, p.11). Puede entonces optarse por el enfoque alternativo señalado por Marshall y Duff (como se citaron en Miró-Llinares, 2024), una perspectiva procedimental que ponga el acento en el proceso por el cual debe tomarse la decisión de criminalización, imponiendo al legislador el cumplimiento de ciertos requisitos en este procedimiento (Miró-Llinares, 2024). Creemos que a una solución similar llega Hörnle (2019) cuando tras concluir la necesidad de disminuir el grado de abstracción de los criterios afirma que es posible lograr ‘’ somewhat thicker

theories'' (p.211) mediante varias teorías o enfoques, un acercamiento plural que se divida en distintos principios de criminalización para delitos contra intereses individuales, los cuales son analizados por la autora desde el prisma de que todos ellos comparten el denominador común de ser lesivos de derechos; y contra intereses colectivos, abriendo la puerta a posibles subdivisiones dentro de cada ámbito.

Al margen de las consideraciones sobre lo acertado o equivocado de la división propuesta por la autora, podría pensarse que esta perspectiva es distinta e incluso incompatible con el modelo procedimental apuntado porque la de Hörnle (2019) sería una mera ''alternative combination'' (Miró-Llinares, 2024, p.11) de principios que quedaría circunscrita a una teoría puramente sustantiva y no procedimental; siendo interesante esta discusión para perfilar de forma más concreta a qué nos referimos por un modelo de criminalización procedimental. Y es que sólo podría afirmarse esta distancia entre ambas propuestas si se entendiese que la de una perspectiva procedimental única y exclusivamente pudiese referirse a cuestiones estricta y puramente formales, cosa que se rechaza, pues precisamente se atribuye a los principios y criterios sustantivos el rol de configurar el marco deliberativo dentro del que ha de desarrollarse el proceso de toma de decisiones, en el cual deberán tenerse en consideración las formas tradicionales de abordar la legitimidad de la intervención penal, pero no de forma exclusiva (Miró-Llinares, 2024).

Como reconoce Douglas Husak (2008), quien elabora una teoría de marcado corte procedimental desde la que se trata de delimitar la legitimidad de la intervención penal del Estado e informar la toma de decisiones del legislador mediante una serie de criterios a los que, de forma deliberada, no se dota de un contenido sustantivo determinado por reconocer lo arduo de tal labor; ''[d]esafortunadamente, esta metodología no podría llevarnos muy lejos a la hora de combatir el problema de la sobrecriminalización. Muchas cuestiones permanecerán sin respuesta si no analizamos con cuidado los conceptos centrales de la teoría de la criminalización que defiendo'' (p.195). Se entiende entonces que un modelo procedimental con pretensión de ofrecer consecuencias prácticas no se encuentra reñido con el perfilado de contenidos sustantivos en los criterios que integra.

Lo que entonces entendemos que diferenciaría esta aproximación procedimental del enfoque en que se configura como respuesta a la pregunta sobre qué castigar una combinación o agregación de principios es la forma de ordenarlos y coordinarlos. En esta última perspectiva estos criterios se combinan de forma simultánea, se tienen en cuenta como una agregación de argumentos que conforman una idea general que lleva a que ''different arguments, moral or deontological, consequentialist or pragmatic, to be used and prevail indinstinctly, whithout being clear which one shall prevail'' (Miró-Llinares, 2015, p.13). Esto traería consigo reincidir en imprecisiones y en problemas similares a un modelo de un solo principio flexible; a lo cual no apunta el modelo de Hörnle (2019), que se presenta más bien como un diagrama en árbol en que los criterios de criminalización se bifurcan en función del interés afectado.

Entendemos entonces que, dejando a un lado las disquisiciones terminológicas sobre qué significa un modelo procedimental o uno basado en varios principios sustantivos; parece que, en el fondo, de lo que se trata es de huir de un único *master principle*, y no de su sustantividad —pues es precisamente el problema del bien jurídico su flexibilidad y poca sustantividad dura—, de huir de una teoría sustentada sobre una clave de bóveda singular; tratando de avanzar en la configuración de un esquema de proceso de toma de decisiones al que se ha de circunscribir el legislador a la hora de decidir sobre la criminalización de una conducta en el que se integre un conjunto de

principios y criterios de un carácter sustantivo tal que permita su funcionalidad como filtro de conductas que pueden ser legítimamente castigadas. Sin embargo, estos principios no deben ser operar demasiado restrictivamente, pues se corre el riesgo de no permitir canalizar adecuadamente las demandas de protección ante la nueva realidad social y el ímpetu del legislador, llevándole a forzar conceptualmente los principios para habilitar el castigo de realidades sensibles socialmente con penas legitimadas por una naturaleza que quizá no presenta realmente. Es por ello por lo que cada uno estos principios no deben ser excesivamente extensivos, permitiendo su operatividad en cascada conforme a una organización sistemática que nivele la intensidad de las consecuencias penales que habilita a la gravedad y relevancia de las conductas que se analizan, lo que convertiría el modelo en ‘’una red lógica y coherente de condiciones de criminalización que no se solapen entre sí y cuyas características y condiciones derivadas resulten claras’’ (Miró-Llinares, 2015, p.54).

Así, estos principios no deben tener por objetivo proyectarse como *master principles* hacia la totalidad de supuestos que se planteen —lo que configuraría esta teoría como una combinación sintética de principios de forma simultánea y no como una ordenación en forma de proceso para la toma de decisiones—, sino que circunscriban su función al espacio concreto y determinado que sistemáticamente le corresponde, filtrando un tipo específico de conductas en atención a su naturaleza concreta, a los problemas que suscita y al tipo de consecuencia jurídica que se le pueda atribuir por dichas dimensiones. A esto entendemos que alude Miró-Llinares (2024) con el plan de imponer al legislador el desarrollo de una mejor y más profunda justificación de su decisión mediante ‘’a series of concrete and not generic arguments’’ (p.13). [2]

Asumimos así lo que creemos que constituye la esencia de un modelo procedimental que abrace la necesidad de dotar hasta cierto punto a sus elementos de un determinado contenido sustantivo. Pero lo apuntado en estas páginas no sólo trae conclusiones puramente formales sobre la clase de modelo de criminalización que entendemos idóneo, sino que también conlleva aventurar ciertas consecuencias sustantivas sobre lo que ha de contener el mismo. La opción por acoger distintos principios hasta cierto punto sustantivos y no optar por un único criterio duro o rígido conlleva implícitamente la admisión de un modelo llamado a ser flexible en su conjunto. Esto es, se acepta que de forma explícita el Derecho penal pueda castigar legítimamente conductas que no comparten en términos absolutos una naturaleza determinada, flexibilizando las categorías que en línea de principio pueden formar parte de lo jurídico-penalmente relevante. Se abandona así un fundamento unitario o singular de la intervención penal, permitiéndose, por ejemplo, que el ejercicio del ius puniendi no se despliegue únicamente sobre conductas que constituyen daños, abriéndose las puertas a la punición de conductas de otro carácter, por ejemplo, las ofensas.

La decisión por una perspectiva que trate de armonizar la necesidad de ofrecer algún tipo de contenido sustantivo para evitar una ambigüedad inoperante con la imposibilidad de que este contenido se refiera a un elemento unitario destinado a ser disfuncional y relativizado; lleva a optar por una solución desde la que se entienda que las conductas criminalizables deben presentar algún tipo de carácter específico que habilite el castigo, pero que entienda que este carácter no es único, sino que pueden existir distintos atributos que legitimen la criminalización para adecuar en mejores términos esta decisión de política criminal a la complejidad y variedad de la realidad ante la que se

[2] En el Anexo I de este trabajo se puede encontrar una serie de diagramas que representan visualmente los distintos *idealtypus* de modelos de criminalización tratados en estas páginas y sus problemas.

enfrenta. Esto es a lo que Miró-Llinares (2015) denomina un *modelo de criminalización flexible desde mínimos*.

Si bien esto puede parecer a priori menos alineado con los principios de un Derecho penal liberal y, por tanto, un enfoque que puede permitir la inflación penal hacia ámbitos que no habrían sido, al menos nominalmente, admisibles en un modelo puramente rígido; las razones para ello ya han sido señaladas al menos parcialmente más arriba y su carácter menos garantista es tan solo aparente.

Como ya se ha indicado, las tendencias expansionistas del Derecho penal no surgen del vacío, sino de una realidad cambiante y compleja que ha llevado a que se le atribuya el papel de tratar de paliar los riesgos que de ella derivan. Al margen de las dinámicas específicas que han propiciado tal cambio, parece certero el señalamiento por Silva (2001) de que estas circunstancias, el avance tecnológico y las dinámicas de las sociedades postindustriales del capitalismo tardío puede llevar consigo la aparición de nuevos peligros y formas de afectar a intereses relevantes, el surgimiento de nuevos intereses antes inexistentes o que ciertas realidades o situaciones que anteriormente eran abundantes y no merecían especial atención ahora sean escasos y ganen una relevancia tal como para recabar la tutela penal. Esta idea es acogida expresamente en el preámbulo del Código Penal de 1995, donde se indica que

> ‘’se ha afrontado la antinomia existente entre el principio de intervención mínima y las crecientes necesidades de tutela en una sociedad cada vez más compleja, dando prudente acogida a nuevas formas de delincuencia, pero eliminando, a la vez, figuras delictivas que han perdido su razón de ser. En el primer sentido, merece destacarse la introducción de los delitos contra el orden socioeconómico o la nueva regulación de los delitos relativos a la ordenación del territorio y de los recursos naturales; en el segundo, la desaparición de las figuras complejas de robo con violencia e intimidación en las personas que, surgidas en el marco de la lucha contra el bandolerismo, deben desaparecer dejando paso a la aplicación de las reglas generales’’ (Exposición de motivos CP).

Estas nuevas realidades llevan a la construcción de nuevos intereses difusos y a una legítima mayor demanda social de protección, generando así lo que para el autor constituye un ‘’espacio de expansión razonable del Derecho penal’’ (Silva, 2001, p. 12). Admitir esto parecería inadecuado si se pretendiese lograr el sueño de negar de entrada la posibilidad de castigo conductas que no se correspondan en puridad con lo que conforme a una noción maximalista y heterodoxa del Derecho penal mínimo puede considerarse punible. Sin embargo, la construcción de una respuesta teórica que desatienda esta realidad social, sus demandas y tendencias, podría desencadenar lo que Silva (2001) denomina estallidos disfuncionales.

Estas exigencias acabarán por introducirse de un modo u otro en la legislación penal, y la opción por un modelo tajantemente rígido llevará a imponer ‘’condiciones mínimas [de criminalización] que acaban por romperse desvirtuando su sentido limitativo final’’ (Miró-Llinares, 2015, p.52). Esto es, acabará produciéndose el mismo resultado de ampliación del ámbito de intervención penal sobre un ámbito que en principio podía corresponderse con un espacio de expansión razonable, pero desde unas premisas que permiten un castigo máximo, pues los límites de entrada que en principio impone también lo son; derivando así en una intervención irrazonable en este sentido (Silva, 2001).

Lo que entendemos oportuno entonces es optar por conseguir

> ‘’una política criminal, en la que se logre la necesaria armonía entre la necesidad de adaptar las normas penales y de procedimiento al surgimiento de nuevos intereses de protección o la revalorización de algunos ya existentes y la necesidad de garantizar la vigencia de un proceso penal garantista y respetuoso de los derechos fundamentales, frente a la innegable criminalidad de nuevo “corte” (Goite y Medina, 2018, p. 2074).

Esto es, tratar que el reconocimiento de este espacio de ampliación razonable sea empleado como pretexto por el legislador para dinamitar los principios garantistas que deben regir en un Derecho penal legítimo. Así entendido, el modelo flexible desde mínimos que aquí se defiende no admitiría la inclusión de criterios adicionales que permita el castigo de conductas de naturaleza diversa con el caballo de Troya de traer consigo una rebaja de garantías. Al contrario, trataría de garantizar con la flexibilidad un respeto más escrupuloso de los límites que se imponen y explicitar la diferencia cualitativa de dichas conductas, estableciendo, como corolario de esta distinción, una igual diferenciación en el tipo de criminalización admitida que opere como elemento racionalizador; articulando para las conductas que no forman parte del núcleo duro del de la tutela penal un régimen atenuado en que se limite el tipo de sanción aplicable (Miró-Llinares, 2015). Hasta cierto punto, creemos que esto se alinea con la propuesta de configurar un Derecho penal funcional compatible con una vocación restrictiva de la intervención punitiva articulada por medio de distintas velocidades (Silva, 2001).

Dicho lo anterior, también parece idóneo adelantar aquí un pronunciamiento sobre cuáles deben erigirse como grandes principios que vertebren un modelo así concebido orientado a ser de utilidad para el asunto que aquí nos ocupa. Creemos que las ultrafalsificaciones son un fenómeno que se caracteriza principalmente, como se ha señalado a la hora de abordar su conceptualización, por producir falsas representaciones en el público sobre realidades diversas de un modo tal que, eventualmente, pueden llegar a ser indisceribles de la auténtica realidad. Esto, como queda patente cuando se atiende a la casuística que se ha mostrado, constituye un modo privilegiado para producir injerencias en esferas personales tales como la propia imagen y el honor, sin perjuicio de que pueda incidir en otros ámbitos de la vida sociopolítica de gran relevancia, como los procedimientos electorales o el discurso del odio, al tener una notable capacidad para influir en la opinión pública. Por ello un modelo de criminalización que trate de ofrecer respuestas sobre la legitimidad de su castigo debe ser capaz de explicar los motivos por los que las conductas relacionadas con estos intereses pueden ser punibles, lo que pasaría por ofrecer una caracterización de su naturaleza y una argumentación sobre la relevancia jurídico-penal de dicha naturaleza. Entendemos que este planteamiento se interrelaciona de forma muy íntima con el debate sobre el castigo de conductas ofensivas, pues en él se encarna la discusión sobre la naturaleza lesiva en sentido estricto o no de los actos que menoscaban el honor y la propia imagen de otras personas o bien los que contrarían contenidos valorativos ampliamente consensuados; así como la conjugación de su castigo con el derecho a la libertad de expresión, cuestión también imbricada en el asunto de los *deepfakes.* Por ello creemos que es favorable la opción por un modelo que orbite en torno a las ideas de daño y ofensa, que servirían de ejes principales de un modelo flexible como el defendido más arriba.

Sin embargo, pese a que aquí hayamos tratado de señalar los rasgos esenciales que debe presentar una teoría de criminalización que trate de ser funcional y útil para nuestros propósitos, somos conscientes de que tratar de dibujar un modelo completo en este sentido es una cuestión harto compleja que excede por mucho las posibilidades del presente trabajo y de quien lo redacta. Nos limitaremos por tanto a un objetivo más modesto que,

no obstante, nos permita en la medida de lo posible avanzar en la problemática que ocupa el lugar central del trabajo: la cuestión de la respuesta penal ante los *deepfakes*. Tomaremos como punto de referencia las aportaciones y avances realizados por Miró-Llinares (2015), centradas precisamente en el debate sobre la criminalización de conductas ofensivas y que presenta como líneas fundamentales de un modelo flexible desde mínimos los principios de daño y ofensa. Trataremos entonces de analizar estas líneas maestras sin desconocer que, al enfocarnos de forma primordial en ellas, lo aquí esbozado del modelo estará abocado a la incompletitud. Creemos, sin embargo, que circunscribirnos a esta tarea presenta el beneficio el evitar una extensión teórica desmesurada y nos permitirá centrar los esfuerzos en configurar ciertas premisas básicas que nos habiliten a desarrollar ciertas reflexiones prácticas y concretas sobre las ultrafalsificaciones.

III. PRINCIPALES EJES DEL MODELO DE CRIMINALIZACIÓN

1. Los principios y límites generales del ius puniendi de un Estado social y democrático de Derecho

En las páginas precedentes hemos defendido que un modelo de criminalización no puede estar ciego ante la sociedad sobre cuyo Derecho penal pretende influir y limitar si desea ser funcional y no sufrir la perversión de sus contenidos. Sin embargo, la apertura cognitiva al sustrato social sobre el que se despliega el sistema penal no puede quedar circunscrita únicamente a las tendencias y demandas punitivas y la dinámica socioeconómica.

El perfilado de un modelo de criminalización debe sustentarse sobre una determinada opción político-constitucional, pues en ella estriba buena parte de la atribución de un papel al Derecho penal y de los valores que deben inspirarlo. No se pretende afirmar entonces que un modelo de criminalización haga referencia a una serie de criterios que se derivan de unos supuestos límites absolutos y universales del Derecho penal. A lo que aquí hacemos referencia es a una aproximación que delimite qué puede castigarse en abstracto con arreglo a unos principios y contenidos circunstanciales, dependientes de la comunidad sobre la que opera. Siguiendo a Lascuraín (1998):

> ‘’El Derecho *per se* no es proporcionado, sino que debe serlo desde cierta perspectiva axiológica y sólo en parte de sus elementos. […] La idea de proporcionalidad pertenece a la deontología del Derecho, no a su ontología; no señala un elemento definicional, una propiedad o una consecuencia de la naturaleza del Derecho, sino una exigencia que impone un determinado criterio de justificación del mismo’’ (p.159).

Lo anterior, prescindiendo de discutir las implicaciones iusfilosóficas que presenta respecto del concepto mismo de Derecho, lo que excede por mucho el propósito de este trabajo; podemos suscribirlo no sólo respecto al principio de proporcionalidad, sino respecto de todo el entramado valorativo que perfila, en un contexto socio-político concreto, los límites, fines y formas del Derecho penal y determinan cuándo y cómo actúa legítimamente este. Entendemos que este acervo valorativo debe ser tenido en cuenta a la hora de tratar de construir el entramado procedimental de decisión sobre la criminalización renunciando, como ha hecho Pawlik (2023) en su teoría del Derecho penal, a toda pretensión de universalidad y asumiendo la que, según este autor, ha sido la mayor aportación de Jakobs a la filosofía del Derecho penal, la de ‘’haber abierto los ojos

de los penalistas a la dimensión sociocultural de su objeto de estudio'' (p.15). En nuestro caso, este contexto remite a la conformación de un modelo de criminalización basado en los principios del Derecho penal de un Estado social y democrático de Derecho.

Esto conlleva que la afirmación de un deber ser, como es la relativa a la decisión de criminalización de una conducta; se realiza necesariamente a partir de otra definición de un deber ser. Concretamente, la de qué debe ser el Derecho penal conforme a los criterios y fundamentos últimos que rigen en nuestros sistemas como sustrato axiológico, como conjunto de valores políticos de nuestro orden constitucional que indican los rasgos que el sistema penal debe poseer, lo cual es independiente de la forma que este presente conforme al Derecho positivo (Miró-Llinares, 2015). Por ello, pese a que tratemos aquí de definir las líneas principales que han de informar las decisiones de política criminal con carácter general, y por tanto también respecto de las de criminalización o no de los *deepfakes*; no dejamos de reconocer que un modelo de criminalización será, así entendido, relativo y circunstancial.

Afirma Hörnle que cualquier teoría de criminalización lleva implícita cierta presuposición sobre la función del Derecho penal, ya que, en sus palabras, ''the question «why should we have state punishment?» is logically prior to «what should be punished?» (Hörnle, 2016, p.301). Efectivamente, para poder afirmar que una conducta debe o no debe castigarse debe haberse admitido la posibilidad de tal castigo. Sin embargo, la imbricación entre las preguntas sobre por qué debe existir el Derecho penal, o lo que es lo mismo, por qué es legítimo el castigo de ciertas conductas por el Estado; y sobre qué debe castigar el Derecho penal es más profunda que una mera relación lógica de precedencia basada en que sin la respuesta a la primera cuestión no puede siquiera plantearse la segunda, lo que dejaría, una vez justificada la existencia del castigo estatal, total libertad para determinar qué puede o debe castigarse. Por el contrario, cuando se fundamenta el ius puniendi lo que se hace es legitimarlo, y ello se realiza supeditando la legitimidad de su ejercicio a que este se oriente hacia unos fines concretos y/o que respete unos principios y límites básicos; por lo que se está influyendo decisivamente en la subsiguiente pregunta sobre qué puede o debe ser castigado por el Estado.

No por nada afirmó Mir (1976) que aquello que se entienda por función de la pena ''ha de ser base de la política criminal tanto en orden a guiar al legislador —por mucho que este desprecie tan a menudo en nuestro país la opinión de la ciencia— como a efectos de crítica de la ley'' (p.76). Piénsese, siguiendo el ejemplo expuesto por este autor, que no será igual de extenso ni incidirá sobre los mismos fenómenos el Derecho penal cuando se atribuya a la pena la función de lograr la justicia en el mundo, lo que produciría una tendencia a ampliar lo criminalizable sobre un mayor número de conductas consideradas injustas o inmorales; que cuando se le atribuya una finalidad puramente resocializadora conforme a un modelo de prevención especial, que podría tender a limitar desde una perspectiva humanitaria el ámbito de lo punible a los supuestos de mayor gravedad. En igual sentido, Roxin (2007) cree que ''los límites de las facultades de intervención penal deben extraerse de la función social del Derecho penal'' (p.436).

Puede verse pues que, en definitiva, la labor de desarrollar un modelo de criminalización parte implícitamente de un pronunciamiento sobre tres cuestiones estrechamente interrelacionadas y cuya respuesta se verá influenciada por el paradigma axiológico desde el que se observen, a saber, la de la legitimidad material del Derecho penal; la de sus funciones y fines; y, por último, la de sus límites, siendo esta última la que de forma explícita influirá, al ser respondida de forma coherente con las anteriores,

los criterios de criminalización a tener en cuenta a la hora de decidir si criminalizar o no una conducta.

Todas estas cuestiones son relevantes a la hora de tratar de confeccionar un modelo o teoría de criminalización sistemático y completo. Sin embargo, como ya se ha indicado, el propósito de este trabajo no es el de desarrollar tal cosa, lo cual, en palabras de Husak (2008), exigiría incluso la elaboración de una Teoría del Estado; sino tan solo el de explorar las líneas maestras que de una forma más directa pueden sernos de utilidad para ser trasladadas a un análisis sobre la criminalización de ciertas manifestaciones concretas del fenómeno de las ultrafalsificaciones. Además, ciertos pronunciamientos como el relativo a la teoría de la pena, aunque teóricamente sea necesario como punto de partida por trazar la orientación más básica y fundamental del modelo, debe reconocerse que la trascendencia práctica muy limitada o nula a la hora de expulsar o permitir una propuesta de criminalización concreta. Por ello quizá baste decir a este respecto que nuestro texto constitucional asume el paradigma preventivo y, concretamente, establece que ‘’las penas privativas de libertad y las medidas de seguridad estarán orientadas hacia la reeducación y reinserción social y no podrán consistir en trabajos forzados’’ (art.25.2 CE), en lo que según Mir (2005) supone la admisión desde el paradigma del Estado social y democrático de Derecho de ‘’una función de prevención limitada de delitos’’ (p.18:10).

Por ello se señalan todos estos elementos cuestiones a desarrollar en futuros trabajos tendentes a avanzar en el desarrollo de un modelo completo. No obstante, que no podamos realizar un análisis pormenorizado de las mismas no significa que debamos dar la espalda a ciertos contenidos y principios que gozan de cierto consenso como límite o guía del modo de ejercer el ius puniendi por un Estado social y democrático de Derecho como el nuestro, que pueden tener valor como incorporaciones en el esquema de criminalización que aquí se esboza al exponer más directamente cómo debe configurar el Estado el ejercicio de su potestad sancionadora. Por ello entendemos que puede ser de gran utilidad limitarnos a prestar especial atención a los límites del ius puniendi que Santiago Mir Puig (1976) deriva de la axiología propia de nuestro régimen político-constitucional y tratar de identificar aquellos que pueden presentar una mayor trascendencia para nuestros objetivos prácticos.

Mir (1976) estructura los principios y límites del ejercicio del ius puniendi en dos grandes categorías: los derivados del fundamento funcional del Derecho penal y los derivados de su fundamento político.

Los primeros serían aquellos que se deducen de la función propia que se atribuye al Derecho penal, la cual para el autor se identifica en la prevención de delitos, y por consiguiente entiende que el fundamento funcional del ius puniendi es el de ‘’la necesidad de protección de la sociedad’’ (Mir, 1976, p.99), extrayendo de este unos límites encarnados en la idea de que el ejercicio del poder punitivo sólo puede estar orientado a la exclusiva y estricta búsqueda de cumplimiento de tal fin, siendo ilegítima toda superación de tal función. En cierto sentido, esta perspectiva concuerda con los presupuestos teóricos de Jakobs (1991) —aunque discrepe Santiago Mir en el normativismo radical del autor alemán (Mir, 2005)— quien asume una función preventiva del Derecho penal basado en la estabilización de las expectativas sociales y la confirmación de la vigencia de las normas penales infringidas mediante un conflicto que debe presentar una dimensión pública, por lo que afirma que ‘’jurídico-penalmente sólo se garantizan aquellas normas a cuya observancia general no se pueda renunciar para el mantenimiento de la configuración social básica’’ (Jakobs, 1991, p.12). Esto es, para Jakobs el Derecho penal sólo puede operar para la protección de la sociedad,

concretamente de los rasgos de su configuración básica, lo que ya de por sí podría presentar una perspectiva limitadora de la intervención penal en un sentido similar a la señalada por Mir Puig. Estas cuestiones serán retomadas más adelante.

Los segundos, esto es, los extraíbles del fundamento político, serían los que se derivarían de la específica concepción y papel que se atribuye al Estado en tanto tal, lo que supone partir de un paradigma axiológico específico y derivar de éste ciertas condiciones que se imponen al empleo de su poder punitivo. Como ya se a apuntado, en nuestro contexto sería el propio de un Estado social y democrático de Derecho. Así entiende Mir (1976) que pueden encontrarse límites derivados del carácter propio del Estado de Derecho, del Estado democrático y del Estado social.

No obstante, debe apuntarse que existe cierta confusión en los términos, ya que en la obra que acaba de referenciarse no se encuentran explícitamente los límites que derivan del Estado social, sino que estos se desarrollan en su tratado ‘’Derecho penal parte general’’ (Mir, 1984). En este último se encuentran como inherentes al Estado social unos límites que se corresponden íntegramente con los que, en su obra de 1976, ‘’Introducción a las bases del Derecho penal’’, identifica como límites derivados del fundamento funcional del ius puniendi. Buena parte de la confusión puede estribar en la metodología empleada por Mir (1976), extrayendo la función de la pena y del Derecho penal de la observación del ordenamiento jurídico-penal positivo, el cual se ha de insertar en el orden político en el que se encuadra y, por tanto, se origina como producto de la función del derecho penal propia de un estado social. Por ello indica que ‘’la idea del Estado social sirva para legitimar la función de prevención en la medida en que sea necesaria para proteger la sociedad’’ (Mir, 1984, p.114), siendo precisamente esta función preventiva la que identifica como núcleo del que derivar límites funcionales del ius puniendi. El problema entonces sería que el autor no identifica el fundamento funcional de la observación del Derecho penal como sistema en sí mismo, sino de la función positivamente atribuida al mismo desde un sistema social determinado. Esta perspectiva le lleva a asumir en un primer término como derivaciones funcionales lo que luego identifica como deducciones políticas.

Sea como fuere, y sin ánimo de profundizar más en el origen sistémico-funcional o axiológico-político de los principios y en la distinción entre ambos ámbitos, debe notarse cómo estos límites, en definitiva, son dotados de un mismo contenido y son, de hecho, los más relevantes a tener en cuenta por nosotros, pues son aquellos que inciden en el establecimiento de una determinada restricción material al ius puniendi como barrera de entrada a lo criminalizable y no meramente de carácter formal sobre un sustrato identificado como legítimamente punible. Esto es, delimitan el contorno exterior del alcance del ius puniendi, que es lo que más directamente nos interesa aquí; y no trazan limitaciones a la forma en que debe darse el ejercicio dentro de los límites exteriores marcados. En las propias palabras de Mir (1976):

> ‘’la primera clase de límites [los del fundamento funcional] es previa a los demás, pues si falta la necesidad de la pena o la medida de seguridad […] el recurso a estos medios no sólo supondría un exceso en el ejercicio de un derecho existente, sino la falta de todo derecho’’ (p.108).

Dicho esto, procederemos a indagar en los distintos límites que forman parte de cada una de estas dos grandes categorías.

La primera de ellas, la de los límites del ius puniendi derivados de su fundamento funcional enraíza, como ya se ha afirmado, con la construcción de una idea de Derecho

penal orientado a la protección de la sociedad mediante la prevención de conductas delictivas. De esto Mir (1976) extrae dos principios generales, a saber, el principio de necesidad, referido a que sólo puede entenderse como coherente con su fundamento el ejercicio de la potestad sancionadora cuando el recurso a la misma sea estrictamente necesario, el cual se subdivide a su vez en dos principios concretos, los de *ultima ratio* y el carácter fragmentario del Derecho penal, al que añadió posteriormente el requisito de utilidad de la pena (Mir, 1984); y el principio de exclusiva protección de bienes jurídicos. Puede observarse que estos dos principios generales se ordenan de modo que el de necesidad tiende a establecer cómo y cuándo puede actuar el Derecho penal para la protección de la sociedad; mientras que el segundo, de protección de bienes jurídicos, se refiere a especificar qué debe entenderse por protección de la sociedad mediante la identificación de los bienes que encarnan sus elementos más valiosos. Si bien ambos presentan una gran relevancia sustantiva aquí nos limitaremos a estudiar los principios específicos de *ultima ratio*, utilidad y carácter fragmentario, corolarios del principio de necesidad; creemos necesario, debido a su complejidad, relegar el análisis de la cuestión del bien jurídico a un apartado específico.

El primer principio específico que hemos mencionado es el de *ultima ratio*. Según este principio, la legitimidad del castigo estatal pasa necesariamente porque el recurso al Derecho penal sea la última opción de entre los distintos medios al alcance del poder político para lograr el fin de protección de la sociedad y su buen funcionamiento, debiéndose entender como una rama del ordenamiento jurídico de carácter eminentemente subsidiario respecto de otros órdenes o medios disponibles. Se indica así como primera opción del legislador el recurso a instrumentos con carácter no sancionador, ‘’como una adecuada política criminal’’ (Mir, 1984, p.128); para dar paso a continuación a mecanismos jurídicos que puedan imponer sanciones no penales, como el del Derecho civil o el Derecho administrativo sancionador; descansando este principio por tanto en la idea central de que ‘’no está justificado un recurso más grave cuando cabe esperar los mismos o mejores resultados de otros más suaves’’ (Mir, 1976, p.109).

Este principio encarna la idea de que no basta la utilidad del Derecho penal para entender necesario el recurso al mismo, sino que para que esto se dé debe ser el único medio útil o eficaz. Este razonamiento no puede sino descansar en una vocación axiológica, y no meramente funcional, tendente a reducir en la máxima medida la intromisión estatal en la libertad individual; la cual responde a los fundamentos políticos del liberalismo clásico. Así, se puede ver que ‘’ esta última opción es, por su propia naturaleza, la menos deseable y, por lo tanto, sólo debe emplearse cuando las demás se evidencien como manifiestamente inidóneas para la consecución del objetivo’’ (Ozafrain, 2017, p.18); y esta indeseabilidad radica precisamente en la voluntad de erigir la libertad individual en un elemento cuasi sagrado que sólo puede ser vulnerado en extraordinarias condiciones. Aparece entonces esta imposición como una forma de garantía de las esferas de libertad del individuo frente al poder político, esto es, como la búsqueda de la consagración de la que Constant (1819) denominó libertad de los modernos, basada en el establecimiento de barreras protectoras del individuo frente al poder heterónomo del Estado. Constituye, pues, un principio liberal que antepone al fin utilitario preventivo de la pena la libertad individual que esta quebranta; y, en tanto esto se concebirá como un acto extraordinariamente lesivo para el individuo, se entenderá que para que sea necesaria la pena no debe ser socialmente útil; sino que debe, además, ser en este sentido inevitable.

Por ello es tan acertado entender que este principio

‘’expresa idea heredada ilustrada por la que se entiende que el Derecho penal nunca puede ser legítimo cuando los poderes públicos tienen a mano otras posibilidades de regulación igualmente adecuadas para el fin propuesto y menos drásticas que el delito y la pena’’ (García de la Torre, 2021, p.137).

El acierto estriba en acentuar el origen histórico-axiológico de la idea de *ultima ratio*. Su identificación como una expresión heredada del pensamiento de la Ilustración pone énfasis en que es el resultado penal del ímpetu del constitucionalismo por lograr ‘’la limitación de los poderes públicos y la consolidación de esferas de autonomía garantizadas’’ (Fioravanti, 2014, p.17) que estaba llamado a consagrarse con el surgimiento del Estado liberal de Derecho (García de la Torre, 2021). Pero con esto no se pretende, en absoluto, desvirtuar o criticar su contenido, sino más bien negar que se trate en puridad de una limitación derivada del fundamento funcional del Derecho penal, como afirma Mir (1976), y resaltar que constituye un contenido que encuentra un fundamento político. Parece dar la razón a ello el mismo Mir (1984) cuando lo propugna entre los principios impuestos por el Estado social, lo cual podría deberse, como ya se ha señalado, a la metodología empleada por el autor. Por ello, no puede obviarse bajo el pretexto de su carácter funcional un análisis de este desde la evolución histórica de la función atribuida al Derecho penal.

Este principio, propio de un Derecho penal puesto al servicio de la garantía de los derechos de los eventuales delincuentes, seria puesto en entredicho con la irrupción del Estado social intervencionista, centrado eminentemente no en la protección del individuo sino en exacerbar la utilidad en el combate contra el delito, lo que permitiría que la degeneración en un Derecho penal social-autoritario (Mir, 1976).

> ‘’Si el fin es únicamente la máxima seguridad social alcanzable contra la repetición de futuros delitos, servirá para legitimar de un modo apriorístico los máximos medios, las penas más severas incluida la pena de muerte, los procedimientos más antigarantistas incluida la tortura y las medidas de policía más autoritarias e invasivas: desde el punto de vista lógico, el utilitarismo, entendido en este sentido, no es de ningún modo una garantía frente a la arbitrariedad del poder’’ (Ferrajoli, 1995, 261)

No sería hasta el resurgimiento del valor de la dignidad humana tras la Segunda Guerra Mundial, y con ella de la necesidad de establecer límites que garanticen la libertad individual frente a excesos por parte del poder político, que recobre el principio de *ultima ratio* su relevancia capital como límite al poder político en el ejercicio del *ius puniendi*. Conforme a su significación histórico-política actual, surge como síntesis de los valores inspiradores del Estado social, con la inclusión de los requisitos y límites propios del Estado democrático, con vocación humanitaria. Desde entonces, ‘’si el fin es también el mínimo de sufrimiento necesario para la prevención de males futuros, estarán justificado sólo los medios mínimos, y por consiguiente el mínimo de prohibiciones, el mínimo de penas […]’’ (Ferrajoli, 1995, 261).

Así, concluye Mir (1976):

> ‘’Para quien, como nosotros crea que el Derecho penal sirve a la función de protección de los bienes jurídicos a través de la prevención de delitos, un Derecho penal actual debería incorporar los postulados del planteamiento social, porque la justificación del Derecho penal subjetivo se halla condicionada a su capacidad para satisfacer del modo más eficaz posible la necesidad de protección de la sociedad. Pero la experiencia histórica —y presente— obliga a destacar con el

mismo valor la necesidad de que el *ius puniendi* respete el ejercicio de su función los *límites* que impone la garantía del individuo. El Derecho penal *social* no debe sustituir sino completar la unilateralidad del Derecho penal *liberal*. La síntesis habrá de alcanzarse en un Derecho penal *democrático*, que impondrá a su vez límites propios a la facultad punitiva del Estado'' (p.107).

En este pasaje queda patente la combinación e interacción de los principios políticos de nuestro Estado y cómo esto lleva a la operatividad de ciertos principios limitativos del *ius puniendi* que pudieron ser debilitados en el vigor del Estado social intervencionista. Sin embargo, también deja clara la artificialidad de la distinción realizada por el autor entre límites funcionales, que en sentido estricto desde una perspectiva utilitario-preventiva se verían reducidos a un núcleo esencial en que se identifique necesidad con utilidad; y límites políticos. Su noción de límites derivados del fundamento funcional no viene sino a ser la de los límites propios de un fundamento funcional políticamente orientado, que parte de la matización de la forma de entender la funcionalidad propia del Derecho penal a partir del marco axiológico en que se encuentra.

Esta constatación quizá nos sea útil para concretar la forma específica en que este principio debe operar. Esto es, responder a la pregunta de cuál es el grado de eficacia Derecho penal respecto de otros instrumentos que permite acudir a él sin vulnerar su carácter subsidiario. Desde una interpretación heterodoxa basada en la garantía de la libertad individual como un derecho sagrado podría defenderse que recurrir al Derecho penal sólo estaría justificado cuando otros medios para la resolución de los conflictos relevantes para la sociedad se presenten como manifiestamente inútiles, configurando entonces el *ius puniendi*, literalmente, como el último y único recurso adecuado. De este modo, mientras otras vías como el Derecho administrativo sancionador, puedan presentar alguna utilidad, aunque menor, deberá optarse por ellas y no huir hacia el Derecho penal.

No parece, sin embargo, que esta solución sea realmente adecuada ni realista. El sentido histórico del principio de subsidiariedad en un Estado social y democrático de Derecho parece conducir, por el contrario, a una vocación sincrética, basada en ''una exigencia de economía social coherente con la lógica del Estado social, que debe buscar el mayor bien social con el menor costo social'' (Mir, 1984, p.128), debiendo valorar este costo, por exigencias democráticas, el perjuicio que ocasiona la intervención penal en un individuo cuya dignidad humana debe ser respetada.

Desde esta perspectiva la búsqueda de la '' máxima utilidad posible para las posibles víctimas debe combinarse con el mínimo sufrimiento necesario para los delincuentes. Ello conduce a una fundamentación utilitarista del Derecho penal no tendente a la mayor prevención posible, sino al mínimo de prevención imprescindible'' (Mir, 1984, p.128). Esto parece aludir a que no podrá ser legítimo el uso de los instrumentos penales cuando ''cabe esperar los mismos o mejores resultados de otros más suaves'' (Mir, 1976, p.109); pero sí cuando, aunque otras vías no punitivas puedan ser útiles, el Derecho penal se presenta como un medio privilegiado y más eficaz para el logro de su cometido preventivo. Así las cosas, debe defenderse la necesidad de tomar en consideración las aportaciones de estudios empíricos sobre el efecto de las penas y los delitos dados por la ciencia penal y la criminología, pues resulta esencial para afirmar con seguridad la mayor o menor utilidad de la criminalización.

Para finalizar con este principio, parece interesante señalar que parte de la jurisprudencia española ha visto en el principio de *ultima ratio* no sólo un límite de política criminal, sino también uno que debe ser tenido en cuenta a la hora de valorar una

conducta normativamente para su calificación como típica. Así ha afirmado el TS que el principio de última ratio impone que ‘’sólo han de entenderse incluidas en el tipo las conductas más graves e intolerables, debiendo acudirse en los demás supuestos al Derecho administrativo sancionador, pues de lo contrario el recurso a la sanción penal resultaría innecesario y desproporcionado’’ (STS 2037/2024, FD 17). Sin embargo, creemos que esta afirmación no responde en puridad al núcleo conceptual del principio de subsidiariedad, sino al carácter fragmentario del Derecho penal, íntimamente relacionado con aquel, precisamente, por compartir ambos su raíz en el principio general de necesidad de la intervención penal. Creemos, por tanto, que pese a esta profunda conexión debe trazarse una distinción conceptual clara entre los principios de *ultima ratio*, recién examinado, y el del carácter fragmentario del Derecho penal, que se aborda más adelante.

Sea como fuere, el contenido de este primer principio se encuentra estrechamente vinculado, como se ha ido señalando, con el del principio de utilidad de la pena, pues el principio de *ultima ratio* se sustenta sobre la presuposición de que la pena constituye un medio eficaz para lograr el fin propio del Derecho penal, el de la prevención de delitos para la defensa de la sociedad.

Este principio de utilidad supone la interdicción del empleo de la represión penal ‘’cuando se demuestre que una determinada reacción penal es inútil para lograr su objetivo protector’’ (Mir, 1984, p.127). Esta conclusión se obtiene del razonamiento según el cual ‘’si el Derecho penal […] se legitima sólo en cuanto protege a la sociedad, perderá su justificación si su intervención se demuestra inútil […]. El principio de necesidad conduce, pues, a la exigencia de utilidad’’ (Mir, 1984, p.127).

No obstante, está claro que el principio de subsidiariedad o *ultima ratio* está subordinado al de utilidad, pues sólo puede afirmarse que el Derecho penal es la última opción viable para la prevención de delitos y que la logra de un modo que otros medios jurídicos no han podido desde el entendimiento de que el Derecho penal, o más bien la pena, es útil. Esto se debe, en definitiva, a que ambos principios tienen su centro de gravedad en la absoluta y estricta necesidad del recurso a la represión penal conforme al ideal liberal-democrático. Mientras el de utilidad establece un límite mínimo para permitir su despliegue, operando desde fuera del Ordenamiento Jurídico, pues sólo puede instituirse el castigo si cumple la función por la que es necesario; el de *ultima ratio*, sentada aquella utilidad, opera como límite máximo endógeno, al establecer la frontera hasta la que el *ius puniendi* puede emplearse y más allá del cual debe ceder ante otros remedios sistémicos. Debemos insistir nuevamente en la necesidad de atender a los estudios empíricos penales y criminológicos, pues resulta esencial para confirmar la utilidad de la criminalización, y con ello, de forma indirecta, el respeto de su subsidiariedad, al sólo poder ser el Derecho penal un último o final recurso útil si es útil en absoluto.

La indisoluble unión de estos principios ya se encuentra en Beccaria (1764), cuando tras fundamentar el origen del *ius puniendi* mediante una tesis contractualista, desde el pretexto de que el poder punitivo surge como una cesión de libertad individual para garantizar la convivencia, afirma que ‘’todas las penas que sobrepasan la necesidad de conservar este vínculo son injustas por su naturaleza’’ (Beccaria, 1764 p.44), y concluye de lo anterior, junto con otros principios como el de legalidad, que

> ‘’cuando se probase que la atrocidad de las penas fuese, si no inmediatamente opuesta al bien público y al fin mismo de impedir los delitos, a lo menos inútil,

> también en este caso sería no solo contraria a aquellas virtudes benéficas que son efecto de una razón iluminada que prefiere mandar a hombres felices más que a una tropa de esclavos, en la cual se haga una perpetua circulación de temerosa crueldad, sino que lo sería a la justicia y a la naturaleza del mismo contrato social'' (Beccaria, 1764, p.46)

Desde el Derecho penal liberal se conforma entonces la utilidad no como un elemento funcional independiente y ajeno a cualquier consideración axiológica, lo cual sería coherente desde una visión pura y estrictamente funcional que busque la imposición de límites inherentes al funcionamiento del Derecho penal como sistema; sino como un elemento puesto al servicio del cumplimiento del fin político propio desde el que se legitima el uso del *ius puniendi*. El principio de necesidad, como una exigencia política propia del liberalismo, primero, y del Estado social y democrático de Derecho, después; absorbe el principio de utilidad y lo atraviesa con un *telos* específico: la búsqueda del punto máximo común entre lo útil y lo respetuoso con el individuo. En la conjugación de ambos límites puede observarse entonces una vocación racionalizadora del ejercicio de la potestad sancionadora que radica en una noción garantista del concepto de necesidad. En este sentido, la represión penal sólo puede emplearse cuando sea el instrumento más eficaz que el resto de medios para evitar la delincuencia —no el único eficaz—; pero, adicionalmente, ante la constatación de que una determinada conducta antisocial no puede ser paliada por el Derecho civil ni el administrativo, y frente a la cual el Derecho penal no ofrezca tampoco expectativas de éxito, no quedaría habilitado el empleo legítimo de la sanción penal.

Pese a lo aparentemente simple de la conclusión anterior, debe notarse que la cuestión de la apreciación de la utilidad o inutilidad de la pena presenta una más profunda complejidad de la que a priori podría presentar superficialmente. La valoración de que la imposición de una pena no es útil para lograr el fin protector depende en gran medida, primero de la función que se atribuya a la pena, y luego de la forma que se entienda que esta pena logra tal función. Efectivamente, dado que ''proclamar la función retributiva de la pena supone entender que la finalidad esencial de ésta se agota en el castigo del hecho cometido'' (Mir, 1976, p.49) no podrá negarse la utilidad de la pena en ningún caso en tanto esta se imponga de forma legítima frente a un acto injusto, pues con la imposición del mal de la pena como castigo del injusto se perfecciona plenamente su cometido. Los retribucionistas se preocuparían entonces tan sólo por el aseguramiento del carácter injusto de la conducta castigada, del merecimiento de la pena en términos de justicia, y no en consideraciones ulteriores sobre los efectos que esta logra.

Pero, como hemos dicho, aquí partimos de una óptica preventiva, que opera desde un prisma utilitario desde el cual parece claro que se debe acoger el principio de utilidad. Sin embargo, es posible que la forma específica que concebir la función preventiva de la pena que tengan algunas teorías de esta corriente produzca que, pese a que se reconozca este principio, este sea superfluo y carezca de valor sustantivo alguno por entender que la pena es siempre útil. Dado que no es nuestro propósito posicionarnos más allá del paradigma preventivo y resolver la cuestión de la teoría de la pena en favor de la prevención general, en alguna de sus formas; o de la prevención especial, la exploración de esta cuestión puede sernos de ayuda para tratar de determinar si la exigencia de la utilidad de la pena, elemento lógico común de toda perspectiva preventiva, puede convertirse en un principio vacío desde cierta noción utilitario-preventiva por la particular forma de concebir su función.

Concretamente, la problemática parece surgir en torno a la visión preventiva de Jakobs. Podría criticarse que si, como Jakobs (1991), entendemos que la prevención desarrollada por el Derecho penal se da por medio de una pena entendida como expresión de significado cuya misión es el ‘’mantenimiento de la norma como modelo de orientación para los contactos sociales’’ (p.14) precisamente porque se identifica como bien jurídico protegido la propia norma penal cuya vigencia garantiza —explícitamente afirma: ‘’la pena debe proteger las condiciones de tal interacción y tiene, por tanto, una función preventiva. La protección tiene lugar reafirmando al que confía en la norma en su confianza’’ (Jakobs, 1991, p.18)—; difícilmente podría entenderse que una pena, en tanto reacción jurídica reestabilizadora frente a la realización de una conducta normativamente negatoria del contenido expresivo orientador de la norma vigente, puede ser en absoluto inútil, pues siempre logrará su fin: estabilizar la confianza y las expectativas respecto de la norma identificada como bien jurídico a proteger cuya vigencia reafirma mediante una pena que por su mera imposición ya culmina su transmisión de significado.

Por ello, podría pensarse, como máximo, que el autor sólo se acerca al reconocimiento de un principio de *ultima ratio* aludiendo a la posibilidad de sustitución de las penas por ‘’equivalentes funcionales’’ (Jakobs, 1991, p.14), especificando que ‘’en la teoría de la prevención general positiva no se trata, pues, de considerar en todo caso adecuada únicamente la pena y no otra reacción’’ (Jakobs, 1991, p14, nota 14a); pero que no acoge en absoluto el requisito de utilidad de la reacción por la que se opte. Con esto vendría a indicar que si bien la pena puede verse como inherentemente eficaz o útil para su cometido de prevención por medio de la protección de la vigencia y confianza en la norma, no es siempre la vía más adecuada para su consecución y, en estos casos, debe cederse el paso a métodos menos gravosos, si bien en todo caso, pero no sin vocación crítica, señala que ‘’tanto para los equivalentes funcionales de la pena como para la evitación de los conflictos […] surgen costes que deben repartirse’’ (Jakobs, 1991, p.15) y dado que entiende que la subsidiariedad es el producto jurídico-penal del principio constitucional de proporcionalidad, sostiene que este recurso al medio sustitutivo del equivalente funcional sólo puede darse

> ‘’cuando los costes de la medida alternativa afectan a una persona que es responsable del conflicto a resolver […] Del principio no ha de deducirse que la pena se convierta en ilegítima cuando el conflicto se pueda prevenir o resolver a costa de cualquiera en lugar de con la pena. Dicho llanamente: todo conflicto puede solucionarse mediante la renuncia al contacto social, así como otros muchos mediante la autoprotección de la víctima, pero la obligación de asumir estos costes no se puede fundamentar por lo general aduciendo que son menos gravosos que la pena’’ (Jakobs, 1991, p.61).

Así, podría entenderse que en algún sentido Jakobs supedita la aplicación del principio de subsidiariedad a que el método alternativo a la pena respete el principio de culpabilidad, la responsabilidad por el hecho propio que conlleva la obligación normativa de asumir los costes tendentes a la reestabilización de la norma defraudada. No parece, sin embargo, que a este respecto la opción por remedios jurídicos frente al conflicto producido como la responsabilidad civil o el ámbito administrativo sancionador supongan vulneraciones de este requisito, si bien agudamente indica que esta opción ‘’puede también conducir a que entonces las nuevas vías estigmaticen a todo lo que se solucione a través de ellas’’ (Jakobs, 1991, p.14, nota 14).

Una vez cerrada la cuestión sobre la subsidiariedad en la teoría de Jakobs, debe notarse que la crítica de prescindir del principio de utilidad con base a la eficacia inherente de la pena que supondría su teoría de la prevención general positiva, creemos, se basa en la construcción de un hombre de paja con el que se identifica la postura jakobsiana, táctica en la que ahonda Mir (2005) cuando dice que Jakobs ‘’contempla el Derecho como un sistema normativo cerrado, autorreferente, y limita la dogmática jurídico-penal al análisis normativo funcional del Derecho positivo, con exclusión de consideraciones empíricas no normativas y de valoraciones externas al sistema jurídico-positivo’’; pues precisamente de esta perspectiva se pretende inferir que, al prescindir de toda observación de los efectos empíricos que la pena tiene sobre la sociedad y centrarse únicamente en la confirmación de la norma, no puede de esta más utilidad que, precisamente, esta confirmación que realiza.

Pero es que el autor no sólo reconoce explícitamente el requisito de la utilidad cuando afirma que ‘’una pena inútil no puede legitimarse de ningún modo en un Estado secularizado’’ (Jakobs, 1992 p.1052). Sino que, además, la utilidad a la que alude no puede estribar sólo en la pena, sino que debe volver la vista a la sociedad, pues Jakobs (1991) apuntala precisamente sus postulados en las dinámicas del contacto social y considera que la eficacia de la pena sí trasciende lo meramente normativo, que sería la confirmación de la vigencia de la norma penal promulgada, e invade lo puramente social-empírico a través de su función normativa. En sus palabras, ‘’la pena tiene una función que debe surtir efectos finalmente en el nivel en el que tiene lugar la interacción social, y que no se agota en significar algo: la pena debe proteger las condiciones de tal interacción y tiene, por tanto, una función preventiva’’ (Jakobs, 1991, p.18). Entiende que la reafirmación de la confianza en la norma propia de la pena opera *erga omnes* comunicando que la norma sigue vigente, y con ello indaga en la confianza generalizada en las expectativas que esta genera; que su quebrantamiento conlleva costes, sembrando en todos la idea de que la defraudación de estas expectativas es negativa; y que estos costes que se imponen como reacción deben asumirse por quien produce la ruptura normativa; sintetizando todos estos mecanismos en una función de la pena entendida como ‘’prevención general mediante el ejercicio en el reconocimiento de la norma’’ (Jakobs, 1991, 18). Por consiguiente, deberá Jakobs reconocer que cuando la pena no logre fácticamente restituir la confianza en la vigencia de esta norma y generar expectativas sociales de cumplimiento habrá que rechazarse la posibilidad de empleo legítimo del *ius puniendi* —por ejemplo en una situación de crisis de la comunidad política en que el Estado es incapaz de lograr el reconocimiento de las normas y, en última instancia, de mantener la vigencia del Ordenamiento Jurídico, caso en que, seguramente, defendería Jakobs la movilización de su Derecho penal del enemigo, lo que es coherente con lo que aquí se ha defendido pues no puede entenderse por tal un verdadero Derecho penal legítimo (Miró-Llinares, 2006)—.

La conexión de su teoría con la realidad social también queda clara cuando explicita cómo la existencia de las normas cuya estabilización persigue la pena radica en el fondo en la afección social a la misma cuando, explicando la causa de la despenalización de ciertos supuestos, indica que lo que sucede es que ‘’la evolución social rechaza total o parcialmente ciertas normas: ya desaparece la decepción de expectativas, o se reduce’’ (Jakobs, 1991, p.61). Y desde el entendimiento de que la afección por la institución que la norma protege se deriva la necesidad de su protección, la vincula con su utilidad, pues entiende que ‘’la pena debe ser necesaria para el mantenimiento del orden social -sin esta necesidad sería a su vez un mal inútil-’’ (Jakobs, 1992, 1052). Se ve así el paralelismo que Jakobs traza entre necesidad y utilidad como corolario conjunto del

fundamento preventivo por medio de la protección de las instituciones básicas socialmente vigentes.

Se debe observar entonces que el principio de utilidad, desde el prisma de la prevención general positiva que propone Jakobs, opera en dos sentidos. En primer lugar, de forma comunicativa en un contexto social determinado donde existen expectativas en el cumplimiento de la norma penal, confirmando su vigencia la pena, que estabiliza la confianza y expectativas depositadas en ella. Cuando no exista tal contexto, cuando no se dé confianza alguna y la norma misma carezca de coherencia normativo-social no existirá utilidad. La utilidad de la pena descansa entonces, en este primer sentido, en la utilidad de la norma penal en tanto generadora de expectativas como modelo de conducta para la resolución de conflictos de una forma alineada a los valores sociales. En segundo lugar, la utilidad debe darse también en el momento de la estabilización misma, que no se da únicamente con la mera imposición de la pena, sino que está llamada a culminarse con la efectiva comunicación de vigencia y restitución de las expectativas. Si se da un entorno en que pese a la imposición de la pena esta es ineficaz para restituir su papel como modelo de conducta vigente y generador de expectativas de cumplimiento el Derecho penal no puede estar legitimado. Puede verse entonces que el normativismo de Jakobs no está reñido con el establecimiento de los principios de *ultima ratio* o subsidiariedad y el de utilidad; sino que estos se derivan del elemento nuclear de su teoría, que es la norma penal como bien jurídico-penal en sí misma y los efectos que esta, posteriormente, despliega en el medio social por medio de la dialéctica negación-confirmación.

Así, tanto en la postura de Mir (1984), que exige de la pena una función preventiva bien desde la disuasión por medio de la conminación penal, que se satisfaría con la comprobación de cuántos ‘’no han delinquido y acaso lo hubieran hecho de no concurrir la amenaza de la pena’’ (p.127), bien desde una perspectiva preventivo-especial tendente a evitar la reincidencia; como en la de Jakobs (1991), con su modelo de prevención general positiva de marcadísimo corte normativista; puede entenderse vigentes los principios de *ultima ratio* y utilidad.

Dicho esto, pasemos entonces al tercero de los subprincipios que se derivan del principio general de necesidad, a saber, el del carácter fragmentario del Derecho penal. Según Mir (1976) este principio ‘’significa que el derecho penal no sanciona todas las conductas lesivas de bienes jurídicos, sino sólo las modalidades de ataque más peligrosas para ellos’’. Este requisito viene a encarnar una vez más la idea de proporcionalidad, que exige orden jurídico-penal llamado a ser empleado cuando sea única y exclusivamente necesario, con fundamento en la gravedad de las penas que impone y el grado de injerencia en la libertad que estas suponen. Conforme a esta filosofía, el uso legítimo del *ius puniendi* no puede identificarse con el castigo de cualquier conducta, por nimia que pueda ser, que afecte de algún modo a lo que se identifique como un bien jurídico; sino que debe darse en todo caso respecto de conductas que perturben este bien jurídico de forma excepcionalmente grave. Incide, por tanto, en la delimitación de los daños o peligros jurídico-penalmente relevantes contra bienes jurídicos protegidos.

No se refiere, entonces, a una limitación al ámbito sobre el que puede incidir el Derecho penal, sino a las concretas conductas que dentro de este ámbito pueden castigarse. Esto es, no impone el carácter fragmentario que sólo pueda actuar frente a las acciones que atacan las estructuras básicas de convivencia y por ello son las más graves, labor que se apoya en la idea de bien jurídico; sino que de entre las distintas formas de atacar aquellos bienes jurídicos sólo se consideren relevantes penalmente las más gravosas, pues sólo así se garantizaría, de nuevo, el ideal de un Derecho penal liberal,

constituyendo una ‘’característica de un Estado de Derecho respetuoso para con la libertad del ciudadano’’ (Mir, 1976, p.110).

Puede notarse, pues, la estrecha relación que presenta este principio con el contenido normativo que se pretenda atribuir a la conducta típica, exigiendo entender por tal conducta típica sólo aquellas acciones que presentan una entidad suficientemente grave. Por ejemplo, no impedirá el principio que se tipifique el delito de lesiones, pero sí establecerá límites a la hora de delimitar el concepto normativo de lesión, pues si bien una gran variedad de acciones puede suponer algún tipo de menoscabo de la integridad física o la salud, sólo las que sean más graves pueden ser constitutivas de una acción típica de lesiones. Lo que aquí se pretende señalar es que esta valoración de la gravedad presenta una naturaleza normativa que enraíza en las disquisiciones sobre los criterios materiales que se impongan como necesarios para permitir el castigo de una conducta, como el daño y, en su caso, la ofensa; lo que entronca con la discusión sobre el carácter material del concepto de bien jurídico y su relación con los principios de daño y ofensa, cuestiones que se explorarán más adelante.

Mir (1976) no ofrece profundas reflexiones sobre este principio, si bien deja patente la estrecha relación de la exigencia de este principio, de nuevo, con la teoría de la pena que se maneje, entendiendo que

> ‘’para quien, como Binding, el Derecho penal esté destinado a la realización de la justicia, es lógico considerar defectuoso que no se castiguen todos los hechos lesivos de unos mismos bienes, con independencia de la peligrosidad […]. Una concepción preventiva del Derecho penal, para la cual el límite del ius puniendi deba ser la absoluta necesidad de defensa de la sociedad, deberá, en cambio, excluir de reacción penal los ataques menos peligrosos’’ (Mir, 1976, p.111).

Sin embargo, caben de nuevo dudas sobre si esta limitación es predicable también de una perspectiva preventiva como la de Jakobs (1991). Sin embargo, más que en la cuestión de su concepción de la prevención como prevención general positiva, parece que los problemas surgen de su noción de bien jurídico penal, el cual se identifica con la misma norma que la pena reafirma. Efectivamente, si el fin protector de la sociedad se da por medio de la estabilización de la norma y la vulneración del bien jurídico que constituye dicha norma ya produce su negación plena, es lógico pensar que no puede hacerse distinción entre ataques más o menos graves al bien jurídico, pues todos ellos producen un mismo efecto: la negación de la norma y la frustración de las expectativas de su cumplimiento. Sin embargo, esto no tiene por qué ser así, precisamente porque Jakobs (1991) coloca en el centro de la contradicción de la norma el significado de la conducta. Entiende que, dado que la acción de un sujeto controlada y dominada por este no es una simple exteriorización de unos efectos en el mundo, sino que ‘’significa también algo’’ (Jakobs, 1991, p.13), esta significación será el contenido normativo de la conducta desplegada y por la cual se desautoriza la norma. Por esta se convierte la mera acción exterior en una infracción normativa que desautoriza y niega la norma penal vigente. El carácter fragmentario operará entonces en la concepción que se tenga sobre la conducta que defrauda la norma, sobre la configuración de su significación jurídico-penalmente relevante, sobre su significado normativo. Para Jakobs (1991) esta configuración normativa es la que produce efectivamente la infracción, y no otras posibles configuraciones, y ‘’la determinación exacta de cuándo concurre una contradicción a la norma es el problema de la teoría de la imputación, en especial de la imputación en calidad de comportamiento típico y antijurídico’’ (Jakobs, 1991, p.13). Por tanto, sobre lo que el carácter fragmentario sí se tendrá influencia es en el momento de delimitar cuál es

precisamente el contenido normativo necesario para entender que una acción posee un significado contrario a la norma, una significación típicamente antijurídica.

Esto se alinea con la reflexión hecha más arriba sobre la relación entre fragmentariedad del Derecho penal y tipicidad de la conducta punible. Creemos entonces, en resumidas cuentas, que el carácter fragmentario aborda la legitimidad del Derecho penal como una suerte de ‘’principio penal en blanco’’ que remite a las cuestiones sustantivas que subyacen en la cuestión del bien jurídico y su relación con los principios de daño y ofensa, lo cual se abordará de forma más clara más adelante.

Estos tres subprincipios que se derivan del principio de necesidad vienen a conformar en su conjunto una idea general llamada a informar el carácter del Derecho penal: este debe limitarse a ejercer una intervención mínima, reducido a lo más estrictamente necesario, en la esfera de la libertad individual de los ciudadanos. Así lo señala Mir (1984) cuando afirma que ‘’ambos postulados [*ultima ratio* y carácter fragmentario] integran el llamado principio de intervención mínima’’ (p.128), al que debemos añadir el de utilidad por descansar precisamente sobre él, como hemos señalado, el mismo principio de subsidiariedad.

El segundo de los principios funcionales del Derecho penal que se erigen como límite a su uso legítimo es el principio de exclusiva protección de bienes jurídicos el cual, como se ha mencionado, será objeto de análisis en el siguiente epígrafe. Baste aquí, entonces, señalar que en el sentido político-criminal que aquí nos interesa, por ser precisamente el que permite su funcionalidad limitativa del poder punitivo, este principio exige que el Derecho penal sólo puede operar para la protección de los intereses sociales que por su relevancia y esencialidad se consideran bienes jurídicos (Mir, 1984), constituyendo por consiguiente el bien jurídico la reificación o cosificación de la función de protección de la sociedad. Por medio de él se concreta qué ha de entenderse por aquellas esferas que conforman las dimensiones y estructuras más necesarias de la sociedad, y por consiguiente se identifica su protección con el cumplimiento de la función propia atribuida al orden jurídico-penal: la protección de la sociedad mediante la evitación de los delitos.

Dicho esto, podemos abandonar el ámbito de los límites impuestos por el fundamento funcional del *ius puniendi* y entrar en el de los derivados de su fundamento político. Sin embargo, como ya se ha señalado, pese a que los anteriores sean pretendidamente una derivación pura de la función que cumple el Derecho penal en absoluto son asépticamente ‘’científicos’’, sino que buena parte de su contenido limitador proviene de una determinada base axiológica propia de un Estado social y democrático de Derecho, desde la que se configura un modo específico de entender la función que cumple el *ius puniendi* basada en la síntesis de los ideales del derecho penal liberal, la intervención social y el respeto a la dignidad humana. En un sentido estrictamente funcional, carente de valoraciones ulteriores al cumplimiento del fin utilitario que se le confiere, sólo podrá definirse como límite funcional un principio de necesidad identificado con la utilidad. Los restantes contenidos expuestos, si bien de gran valor y absolutamente necesarios en un contexto político como el nuestro, no dejan de ser, precisamente, exigencias de nuestro modo particular de entender lo que debe ser el Derecho penal conforme a un conjunto de contenidos axiológicos — lo cual, de nuevo, no desmerece un solo de ápice su valor—.

Puede verse entonces que hemos adelantado mucho en la cuestión de los límites derivados del fundamento político en nuestra exploración de los límites funcionales al *ius*

puniendi. En cualquier caso, el fundamento político del *ius puniendi* en nuestro contexto ‘’ha de ser, como se dijo, triple: el de Estado de Derecho —o liberal—, que se refiere al aspecto *formal* de sujeción a la ley, y el social y democrático, que apuntan al contenido material del Derecho penal’’ (Mir, 1976, p.124). Se compone, por consiguiente, en torno al modelo triádico del Estado social y democrático de Derecho.

Comencemos, pues, con los límites impuestos por el Estado —liberal— de Derecho. Para Mir (1976) ‘’los límites derivados del Estado de derecho son consecuencia de lo que se conoce como «principio de legalidad».

Si bien pueden encontrarse antecedentes históricos de este principio den la Carta Magna inglesa de 1215 y en la *Constitutio Criminalis Carolina* de 1532, su significación moderna aparece con el movimiento intelectual de la Ilustración que alcanzará su plena consagración con el triunfo de la Revolución Francesa (Mir, 1976).

Desde las tesis contractualistas, donde se fundamenta precisamente el *ius puniendi* sobre la base del consentimiento de los individuos en la conformación de un poder político dotado del derecho a castigar sólo y exclusivamente en la medida en que sea necesario para conservar la comunidad así creada y su paz, entenderá Beccaria (1764) ‘’que sólo las leyes pueden decretar las penas de los delitos, y esta autoridad debe residir únicamente en el legislador, que representa toda la sociedad unida por el contrato social’’ (p.44). Se verá en el principio de legalidad, especialmente desde la óptica de Rousseau (1762), también una exigencia democrática, en tanto la sólo la ley, como manifestación de la voluntad popular, puede establecer las prohibiciones, y las penas cuyo incumplimiento acarrea, que rigen en una sociedad respetuosa de la libertad y autonomía de los individuos que la conforman.

Conforme a estos ideales liberales se conformará un principio de legalidad condensado en la forma latina de *nullum crimen, nulla poena sine lege*, acuñada por Feuerbach (Mir, 1976). Este se verá como ‘’la garantía política de que el ciudadano no podrá verse sometido por parte del Estado ni de los jueces a penas que no admita el pueblo’’ (Mir. 1984, p.115); pero especialmente, por lo que aquí nos ocupa, como una ‘’exigencia de seguridad jurídica’’ (Mir, 1984, p.115), pues sólo podrá castigarse aquello que legalmente se establece como delito mediante una pena que no podrá extenderse más allá de lo que la misma ley dispone, pues ‘’una pena extendida más allá del límite señalado por las leyes contiene en sí la pena justa más otra pena adicional’’ (Beccaria, 1764, p.45).

De este principio se derivan, según Mir (1984) cuatro tipos de garantías, a saber, una garantía criminal consistente en que el delito esté determinado por una norma con rango de ley; una garantía penal, conforme a la cual la pena impuesta sólo puede ser la señalada por la ley; una garantía jurisdiccional, que exige que la pena se imponga por un juez; y una garantía de ejecución, tendente a exigir que el modo de ejecutar la pena impuesta sea igualmente un modo legalmente establecido.

Estas garantías se concretan, a u vez, en tres requisitos que debe presentar la ley penal. Estos son la exigencia de una *lex praevia*, que prohíbe la retroactividad de las normas penales no favorables como imposición de la seguridad jurídica; una *lex scripta*, debiendo ser esta ley penal una norma con rango de ley y excluyendo la posibilidad de castigo con base en la costumbre; y una *lex stricta*, requerimiento referido a la precisión de la ley, que debe respetar un mandato de determinación, debiendo establecer claramente cuál es el contenido típico de la conducta que se castiga y su pena sin desvirtuar la garantía legal mediante cláusulas generales o tipos penales en blanco y sin poder emplearse la analogía *in malam partem* (Mir, 1984).

Si bien estos límites al *ius puniendi* tienen una gran relevancia, puede notarse que constituyen límites de carácter formal, pues inciden en el modo en que la conducta criminalizada se formaliza en una ley penal; y si bien también impone que esta presente un contenido material suficientemente determinado, nada aportan sobre cuál debe ser este contenido material para entender que la represión penal sea legítima. Por ello, los límites derivados del Estado de Derecho no son especialmente interesantes para la labor que aquí perseguimos, que es la delimitación del alcance material de la legitimidad del *ius puniendi*.

En cuanto a los límites del Estado social, estos son abordados por Mir (1984) identificándolos como el principio de utilidad de la intervención penal, el principio de *ultima ratio* y el carácter fragmentario del Derecho penal, todos ellos corolarios del principio de necesidad, junto con el principio de exclusiva protección de bienes jurídicos. Estos ya han sido explorados como límites funcionales, calificados como tales por el propio Mir (1796). Las razones de esta ambivalencia ya han sido expuestas.

Basta decir entonces que, si bien pueden constituir principios funcionales, estos sólo lo son desde el entendimiento de que constituyen derivaciones de una función específica y modulada perfilada por la síntesis del Estado social intervencionista con una vocación garantista que busca lograr un Derecho penal al servicio del ciudadano, y por consiguiente respetuoso con sus derechos y libertades, que ponga frenos al ímpetu expansionista del Estado social (Mir, 1976). Esta síntesis se encarna en el acogimiento de los principios de un Estado democrático llamado a lograr que el Derecho penal, sin desmerecer las exigencias de su utilidad como medio de protección de la sociedad a través de la prevención de delitos; vuelva a erigirse en la *Magna Charta* del delincuente.

De este modo, el Derecho penal de un Estado que democrático debe dotarse ‘’de un contenido respetuoso de una imagen del ciudadano como dotado de una serie de derechos derivados de su dignidad humana, de la igualdad (real) de los hombres y de sus facultades de participación en la vida social’’ (Mir, 1984, p.133). Este prisma, para Santiago Mir (1984) viene a poner en funcionamiento los principios de humanidad de las penas, culpabilidad, proporcionalidad y resocialización.

El resurgimiento tras la II Guerra Mundial del valor del humano en tanto tal como poseedor de una dignidad propia supuso el renacimiento, tras el triunfo de los totalitarismos europeos, del respeto por los derechos y libertades del individuo. Así, la exigencia de la humanidad de las penas entronca en los mismos valores que hicieron surgir el principio de intervención mínima. Así lo afirma también Mir (1984) cuando entiende que este principio es una reivindicación alineada con el núcleo del programa penal de la Ilustración.

El principio de humanidad de las penas viene a exigir, en primer lugar la desaparición de las penas corporales así como la pena de muerte; pero también ha supuesto una progresiva búsqueda de sustitución de las penas privativas de libertad por otras consecuencias jurídicas del delito menos lesivas (Mir, 1984). Así, este principio llevaría a la entender que las penas más graves previstas en el ordenamiento jurídico-positivo, debiendo entender por estas en nuestro contexto actual las penas privativas de libertad; no sólo deben cumplirse en unas condiciones mínimas de humanidad y respeto por el reo, sino que además deben concebirse como una *ultima ratio* dentro de la *ultima ratio* que constituye el Derecho penal.

Esta idea es capital para la configuración del modelo de criminalización de varias «velocidades» basado en varios principios que aquí nos proponemos, pues lleva a la

necesaria conclusión de que solo aquellas conductas que se correspondan con un criterio de criminalización basado en la gran lesividad y gravedad de mismas puedan conllevar penas privativas de libertad; mientras que el marco criminal atenuado basado en otros criterios más leves nunca podrá traspasar el límite de la pena de prisión.

Con este principio de humanidad de las penas se encuentra estrechamente vinculado el principio de resocialización, pues la exigencia humanitaria incide en el modo que la pena debe cumplir su función preventiva. La sincretización del utilitarismo social con la dignidad humana exigirá así que la pena no esté enfocada a la marginación del individuo penado, sino a la promoción de su participación en la vida social, prefiriendo, de nuevo, las penas que no lo aíslen de la sociedad y orientando todo el conjunto del catálogo punitivo al trato del delincuente como sujeto y no objeto de la búsqueda de resocialización, tratando de que reincorpore a la vida libre en sociedad mediante la ampliación de sus posibilidades de participación a través de ‘’ofertas de alternativas al comportamiento criminal’’ (Mir, 1984). Esto, como ya se ha indicado, es exigido por el art. 25 de nuestra Constitución Española.

Por otro lado, por medio del principio de culpabilidad se consagra el principio del hecho, conforme al cual debe rechazarse de forma tajante el Derecho penal de autor y propugnar un modelo desde el que se castiguen las conductas cometidas, y no ‘’formas de ser, personalidades, pues su configuración por parte del sujeto es difícil de determinar’’ (Mir, 1984, p.135). En definitiva, debe imperar la máxima de *cogitationis poenam nemo patitur*, pues sólo pueden castigarse actos exteriorizados, nunca los meros pensamientos, voluntades ni tan solo la resolución interna de delinquir (STS 13505/1988), pues sólo ‘’son delitos las acciones y omisiones dolosas o imprudentes penadas por la ley’’ (art. 10 CP). Pero no basta con que se castiguen conductas, y no pensamientos, sino que además, sólo puede el Derecho penal determinar responsabilidad por el hecho propio, que debe basarse en una imputación personal de un hecho doloso o imprudente, por lo que el castigo debe basarse, en virtud de la exigencia de culpabilidad, en un ‘’principio de personalidad de las penas’’ (Mir, 1984, p.135).

En último lugar debemos ocuparnos del principio de proporcionalidad. Reconoce Mir (1984) que

‘’el principio de culpabilidad no basta, entendido en sus justos términos, para asegurar la necesaria proporcionalidad entre delito y pena. Aquel principio sólo exige que pueda «culparse» al sujeto de la lesión por la que se le castiga, lo cual requiere sólo ciertas condiciones que permitan *imputarle* la lesión […]. Nada dice esto de la *gravedad* de la lesión ni, por tanto, de que deba ajustarse a ésta la cuantía de la pena’’ (p.139)

Sin embargo, siguiendo a Lascuraín (1998):

> ‘’Si el criterio democrático de legitimidad del Derecho y del Estado es la consecuencia de la proclamación y la vigencia de ciertos valores, y entre ellos, de modo muy significado […], del valor de la autonomía personal o libertad genéricamente entendida, resultará que ciertas normas restrictivas de la misma sólo podrán encontrar justificación en su funcionalidad para generar más libertad de la que sacrifican’’ (p.160).

Esta exigencia de proporcionalidad se bifurcará en dos sentidos. ‘’Por una parte, la necesidad misma de que la pena sea proporcionada al delito. Por otra parte, la exigencia de que la medida de la proporcionalidad se establezca en base a la importancia social del hecho’’ (Mir, 1984, p.139). El primero de los sentidos nos interesa por cuanto ahonda en la reflexión hecha sobre la necesidad de graduar las penas a imponer en función de las

«velocidades» que se configuren en el modelo de criminalización que tracemos, así como también por suponer la necesidad de una distinción de las penas a imponer dentro de cada uno de estos marcos penológicos en función de la gravedad de los distintos hechos que se correspondan con un mismo criterio de criminalización. Esto es, impone una graduación penológica tanto inter como intra «velocidades» del modelo.

El segundo de los sentidos impone que sólo puedan considerarse como hechos delictivos aquellos posean un carácter nocivo socialmente considerable. Parece que este sentido es el que lleva principalmente a Lascuraín (1998) a entender que ‘’el principio de proporcionalidad integra el muy restringido grupo de principios que informan la actividad penal sustantiva en un ordenamiento democrático’’.

Efectivamente, este carácter sustantivo es el que lleva al principio de proporcionalidad a incidir no sólo en el tratamiento penológico de lo criminalizado, sino también en la delimitación de lo criminalizable. Sin embargo, entendemos que esto vendría a suponer la equiparación de la proporcionalidad con las exigencias del principio de intervención mínima —y los principios que lo integran y configuran—así el principio de exclusiva protección de bienes jurídicos.

Parece estar de acuerdo el propio Lascuraín (1998), pues deriva del principio de proporcionalidad la necesidad de comparar la intervención penal ‘’con otras medidas alternativas y sopesar si con una medida de menor intensidad coactiva, con menos gasto de libertad, podemos alcanzar metas similares’’ (p.162), entendiendo por tanto que la proporcionalidad se ‘’escinde, pues, en dos juicios, el de necesidad y el de proporcionalidad en sentido estricto’’ (p.163). Y, de hecho, vincula expresamente este juicio de necesidad con la consideración de la ‘’calidad de libertad del bien protegido’’ (Lascuraín, p.163), con lo que se refiere al requisito de que el *ius puniendi* se emplee con la finalidad de protección de un bien jurídico protegido entendido como una condición de libertad. Puede verse así, claramente, la analogía que para Lascuraín (1998) puede trazarse entre el principio de proporcionalidad, al menos de una de sus dos vertientes, con el principio de necesidad aquí abordado como parte del fundamento funcional —y político-social— del *ius puniendi*.

En este sentido, tanto este juicio de necesidad propio de la proporcionalidad expuesto por Lascuraín (1998), como la exigencia de Mir (1984) para confirmar la proporcionalidad del recurso al poder punitivo de que este sólo castigue los hechos más nocivos socialmente; remiten a las exigencias propias del entendimiento de que el Derecho penal sólo puede proteger bienes jurídico-penales, que son tales por su esencialidad e imprescindibilidad social, así como el carácter fragmentario de tal protección, actuando sólo frente a los actos más graves contra dichos bienes. Por tanto, en pos de la claridad de la exposición, entenderemos aquí por principio de proporcionalidad sólo su dimensión penológica y circunscribiremos sus consecuencias sustantivas para la delimitación de las conductas castigables por el *ius puniendi* al ámbito propio del análisis la teoría del bien jurídico protegido y del carácter fragmentario del Derecho penal, sin dejar de reconocer que estos son también una exigencia propia del principio democrático que informa nuestro Estado.

2. La teoría del bien jurídico protegido: sobre el bien jurídico de Schrödinger

Cuando se habla de bien jurídico puede hacerse en dos sentidos. El primero de ellos es un ‘’sentido dogmático, que alude a los objetos que de hecho protege el Derecho penal vigente’’ (Mir, 1984, p.131), cumpliendo una función que Hirsch (2007) denomina intrasistemática que de los bienes jurídicos protegidos ‘’uno de los *topoi* esenciales de

cara a su interpretación'' (p.33), sirviendo de ''ayuda a la aplicación de los preceptos recogidos en la Parte Especial, en tanto funciona como criterio orientativo de la interpretación'' (Kahlo, 2007, p.49). En este sentido, el bien jurídico es entendido como el objeto sobre el que recae la conducta delictiva tipificada y que debe servir para poder determinar cuándo una conducta enjuiciada se corresponde con la conducta típica que se castiga. La segunda de las acepciones del bien jurídico protegido sería el que Mir (1984) denomina su ''sentido político-criminal'' (p.130), una función crítica que ofrecería la ''posibilidad de declarar la inexistencia de un bien jurídico merecedor de protección frente a supuestos de penalización de conductas meramente inmorales […] lo que a su vez habría de llevar a concluir que dicha conducta no debería ser criminalizada'' (Hirsch, 2007, p.33). La noción de bien jurídico que aquí nos interesa es precisamente en su sentido político-criminal o crítico, pues se erige conforme a este significado en un intento de limitar materialmente el poder punitivo estatal, identificando el ''objeto que puede reclamar protección jurídico-penal'' (Mir, 1984, p.130). Constituye, así un mínimo sustantivo que debe presentar un elemento para poder ser tutelado jurídico-penalmente, permitiendo un juicio crítico de la *lege data* en atención a si esta es respetuosa con este límite, más allá del cual no puede extenderse el poder punitivo, desde la idea de que ''la penalización de una conducta tiene que poseer una legitimación distinta de la que le otorga la mera voluntad del legislador'' (Roxin, 2007, p.433).

Mir (1976) extrae el principio de exclusiva protección de bienes jurídicos como una exigencia funcional del *ius puniendi*, esto es, como una imposición derivada de la misma función que el Derecho penal ha de cumplir en un Estado social y democrático de Derecho. Entiende que, en España al menos, esta función es ''la prevención de los delitos por razón de su gravedad y del peligro representado por los medios empleados y por la posibilidad de repetición'' (p.86) dado que sólo con esta función se satisface el fundamento del *ius puniendi*, que ''sólo puede hallarse en la necesidad de protección de la sociedad'' (p.99). Esta protección de la sociedad se dará, por tanto, por medio de la protección aquellos intereses sociales ''que por su importancia merecen la protección del Derecho'' (Mir, 1984, p.130) que son lo que se identifica como bienes jurídicos, debiendo ser de una importancia tal como para que su protección se identifique con la protección de la sociedad misma, pues este es el fundamento funcional del *ius puniendi* que se maneja; y por consiguiente, el poder punitivo se haya materialmente limitado funcionalmente por la exigencia de que sólo actúe para la protección de aquello que constituya un genuino bien jurídico, pues sólo así puede proteger la sociedad mediante la prevención de delitos entendidos como acciones que menoscaban estos bienes. Por ello hemos afirmado más arriba que estos bienes jurídicos suponen la reificación de la protección de la sociedad.

De igual modo, Roxin (2007) parte de que ''los límites de las facultades de intervención penal deben extraerse de la función social del Derecho penal'' (p.436), entendiendo por tal función la de ''procurar a los ciudadanos una existencia pacífica, libre, socialmente segura, en la medida en que tales objetivos no puedan conseguirse mediante otras medidas socio-políticas menos intrusivas en la esfera de libertad de los ciudadanos'' (p.436), de tal suerte que ''las normas penales sólo pueden perseguir la finalidad de asegurar a los ciudadanos una coexistencia libre y pacífica garantizando al tiempo el respeto de todos los derechos humanos'' (p.437). Acorde con lo anterior, concluye que debe entenderse por bien jurídico ''a todos los objetos que son legítimamente protegibles por las normas bajo estas condiciones'' (p.437).

Puede verse entonces que la idea común que subyace en el principio de exclusiva protección de bienes jurídicos es ''que no pueden ser amparados por el Derecho penal

intereses *meramente* morales […] exige que tengan *algo más* que los haga merecedores de protección jurídico-penal'' (Mir, 1984, p.131), por lo que el Derecho penal no tendría la función de proteger ''creencias políticas o morales, doctrinas religiosas, ideologías sobre el mundo o meros sentimientos'' (Roxin, 2007, p.434). Debe circunscribirse únicamente, pues, a la protección de los elementos, estructuras, instituciones, estados, realidades o intereses que constituyan la base esencial de la organización sociopolítica de que se trate, pues ''sólo puede considerarse bien jurídico, como objeto merecedor de protección jurídico-penal, aquello que sea necesario para la subsistencia, en ciertas condiciones, de la sociedad'' (Mir, 1976, p.116).

En realidad, a una conclusión similar, pese a su conocida posición escéptica respecto de la teoría del bien jurídico, llega Jakobs (1991). Si bien se le critica que entiende ''desde un principio que la finalidad del Derecho penal no es la protección de bienes jurídicos, sino la confirmación de la vigencia de la norma'' (Roxin, 2007, p.435), lo cierto es que su identificación de la norma como bien jurídico-penal en sí misma se realiza desde la presuposición de que desde un Derecho penal legítimo ''jurídico-penalmente sólo se garantizan aquellas normas a cuya observancia general no se puede renunciar para el mantenimiento de la configuración social básica'' (Jakobs, 1991, p.12), y por ello toda vulneración de la norma constituye un conflicto público. De hecho, explicita que ''la legitimación material reside en que las leyes penales son necesarias para el mantenimiento de la forma de la sociedad y del Estado'' (Jakobs, 1991, p.44). Si bien es lógico que se le reproche que ''el sistema social no debe ser conservado en su propio beneficio, sino en beneficio de las personas que viven en tal sociedad'' (Roxin, 2007, p.446) y que ''evita cualquier afirmación sobre la legitimidad o ilegitimidad del contenido de las normas, ya que considera que este tipo de afirmaciones no son científicas'' (Roxin, 2007, p.447); lo cierto es que, en cierto modo, sí lo hace implícitamente cuando, tras afirmar que ''no existe ningún contenido genuino de las normas penales'' (Jakobs, 1991, p.44) establece que el contenido legítimo posible de las mismas ''se rige por el respectivo contexto de la regulación'' (Jakobs, 1991, p.44-45). Es cierto que sólo establece un criterio legitimador ''para la institución de la pena como tal, pero no para normas penales concretas'' (Seher, p.68), pero reconoce que estos límites sustantivos derivan del marco sociopolítico específico en que se dé la norma penal y que determinen su legitimidad o no. Lo único que puede recriminársele entonces, y quizá con razón, es que no trate en absoluto de aportar ningún avance en este sentido por considerarlo una cuestión ajena a la labor científica de la doctrina penal y que se circunscribe únicamente a la deliberación política, pero es algo que en cierto modo también asumen Roxin (2007), que reconoce que ''los bienes jurídicos no tienen una eterna validez iusnaturalista, sino que se ven afectados por los cambios en la estructura constitucional y las relaciones sociales'' (p.448); y Mir (1984), para quien ''esta determinación de los bienes a proteger penalmente depende de los intereses y valores del grupo social que en cada momento histórico detenta el poder político'' (p.173).

Creemos, en definitiva, que es positivo pronunciarse, al menos de forma genérica, a aquello que se determina la legitimidad de la norma penal por medio del concepto de bien jurídico, precisamente para no dejar ''a los juristas a merced del capricho y la arbitrariedad del legislador'' (Roxin, 2007, p.446); pero también creemos que la distinción entre colocar el concepto de bien jurídico en un elemento relevante de la configuración social que la norma penal protege y que sólo puede ser legítima si se refiere a este tipo de elementos; y entender que la norma penal es el bien jurídico en sí mismo, pero que sólo puede serlo si se orienta a la protección de elementos relevantes de la

configuración social, es una cuestión que presenta poca trascendencia práctica para los fines que aquí nos proponemos.

Dicho lo anterior cabe preguntarse entonces qué naturaleza concreta tienen y cómo pueden extraerse cuáles son efectivamente estos bienes jurídicos en un contexto sociopolítico determinado. Sin embargo, respecto de estas cuestiones, más allá de la que es, más o menos, la idea básica a la que se orienta la función crítica del bien jurídico protegido, no hay consenso. Si el bien jurídico se crea mediante la valoración plasmada en la propia ley penal que lo protege o si se corresponde con una realidad social anterior al Derecho, si se trata de un valor, un elemento con entidad corpórea, o algo que pese a no ser físicamente tangible sí pertenece al campo de una realidad, al menos, social; es algo que no se ha conseguido convenir satisfactoriamente, como se muestra en la evolución del concepto y naturaleza concreta del bien jurídico expuesta por el propio Mir (1976, pgs. 112 y ss.).

Sin embargo, más allá de las consideraciones sobre su naturaleza y entidad concreta, desde el punto de vista axiológico que impone el contexto del Estado social y democrático de Derecho, sí se ha podido avanzar a una aproximación de cuál ha de ser la orientación que debe darse al sentido del bien jurídico como elemento crítico. Así, entiende Mir (2005) que

> ‘’si el Derecho penal ha de estar al servicio de los seres humanos, habrá de proteger intereses reales de éstos, ya sean directamente vinculados a su individualidad –como la vida, la integridad física, la libertad sexual, el patrimonio, etc.–, ya sean mediados por instituciones de las que dependen intereses individuales –como la Administración de Justicia u otras instituciones estatales–. Los bienes jurídico-penales han de verse como concreciones de estos intereses reales de los individuos, directos o indirectos, que merecen por su importancia fundamental la máxima protección que supone el Derecho penal’’ (p. 18:13)

En igual sentido, Roxin (2007) entiende, como ya se ha señalado, que ‘’en un Estado democrático de Derecho […] las normas penales sólo pueden perseguir la finalidad de asegurar a los ciudadanos una coexistencia libre y pacífica garantizando al tiempo el respeto de todos los derechos humanos’’ (p.47), lo cual se consigue mediante la protección no sólo de ‘’las condiciones individuales necesarias para tal coexistencia (como la protección de la vida y la integridad física, de la libertad de actuación, de la propiedad, etc.), sino también [de] las instituciones estatales que sean imprescindibles a tal fin’’ (p.437). Concluye entonces una definición de bien jurídico penal como ‘’realidades o fines que son necesarios para una vida social libre y segura que garantice los derechos humanos y fundamentales del individuo, o para el funcionamiento del sistema estatal erigido para la consecución de tal fin’’ (Roxin, 2007, p.437).

Parece también referirse a esta idea Lascuraín (1998) cuando establece como requisito para la proporcionalidad de la norma penal, en el sentido de necesidad que el autor le da como criterio legitimador de carácter material, que esta tenga por fin la protección de ‘’una libertad o una condición de libertad […] que representan las condiciones externas —materiales— del ejercicio de la libertad’’ (pp. 163-164).

Se puede ver entonces que estos autores hacen orbitar la idea del bien jurídico en torno al individuo y sus derechos en tanto humano, orientando el Derecho penal a proteger los elementos que permiten su libre desarrollo en el marco de la sociedad; una perspectiva que marcadamente se alinea con el acervo valorativo proclamado por los Estados sociales y democráticos de Derecho actuales. Por consiguiente, esto llevará a pensar que los bienes

jurídico-penales más esenciales y que conformen el núcleo duro de protección penal son intereses primordialmente los de carácter individuales, que garantizan ciertas esferas personales corpóreas, morales y patrimoniales; pero también permite extraer la existencia de ciertos bienes jurídicos colectivos, referidos al funcionamiento de instituciones sociales y políticas, siempre y cuando su necesidad se pueda fundamentar a través y desde el individuo, siempre que ‘’puedan funcionalizarse desde el individuo’’ (Hassemer, 2000, p.167, como se citó en Seher, 2007, p.69).

Sorprendentemente, esta perspectiva parecería ser del agrado de Jakobs (1991), que explica cómo se ha dado la tendencia a identificar con bien jurídico ‘’la relación (comprendida en el interés, valorable positivamente) entre una persona y una situación, pero sin imponer al concepto la función de designar a las fuentes de estas relaciones valorativas (vida, cultura, ordenamiento constitucional, etc.)’’ (p.51), siendo lo que se protege ‘’la posibilidad de que una persona realice sus intereses […] el uso y disfrute de una situación valorada positivamente’’ (p.51), lo que presenta claras similitudes con las ‘’posibilidades de participación del individuo en los sistemas sociales’’ de Mir (1984, p.174). Es decir, se pone el foco en que el bien jurídico alude a una funcionalidad para el individuo en la sociedad, pero no se pide de este que identifique la razón por la que tal disponibilidad de recursos es positiva, labor que se remite al orden cultural y/o constitucional, que en nuestro caso deberá estar orientada a la preservación de la dignidad humana, los intereses y valores que de ella derivan, y la comunidad que permite su protección y desarrollo. ‘’Con esta inclusión de la persona en el concepto de bien jurídico se puede encontrar lo valioso del bien jurídico mejor que mediante la mera enumeración de objetos’’ (Jakobs, 1991, p.51).

Mas allá de esta orientación básica, la especificación de qué es un bien jurídico protegido y cuáles son concretamente en un contexto como el nuestro es una cuestión que, como no puede ser sorprendente, no quedará resuelta aquí, dada su complejidad, considerando incluso algunos que un logro en tal sentido sería lograr la cuadratura del círculo (Stratenwerth 2007). Nos basta para los fines de este trabajo haber esbozado el bien jurídico desde una perspectiva tendencial, especificando la idea general de hacia qué debe orientarse la protección penal, a saber, que esta ‘’no ha de ocuparse de respaldar mandatos puramente formales, valores puramente morales, ni intereses no fundamentales‘’ (Mir, 1984, p.134) y que la delimitación de qué es y qué no un interés fundamental debe radicar en la promoción del desarrollo libre del individuo en la sociedad y la estabilidad de la misma como requisito para tal desarrollo; sin perjuicio de que deban continuarse los trabajos tendentes a perfilar un contenido más sustantivo.

Si bien concluye Wohlers (2007) que el bien jurídico no habría cumplido su cometido teórico, que es, según Seelman (2007), el de responder por sí mismo la pregunta de si hay razones para castigar más allá del Derecho; por precisar para el desarrollo de su potencial crítico ser enriquecido desde ‘’baremos que se tienen que incorporar externamente’’ (Wohlers, 2007, p.393) de carácter teórico-sociales, esto no supone que debamos rechazarla de plano, pues podríamos acoger una teoría del bien jurídico limitada que reconozca que no es autosuficiente y que, dando indicios de qué debe considerarse un bien jurídico-penal, se abra a la necesidad de acudir al entorno social para su determinación. Desde esta óptica podemos poner de relieve el carácter dinámico del bien jurídico en atención a la evolución social, cuestión que, como se ha señalado, admiten tanto Roxin (2007) como Jakobs (1991) o Mir (1984) entendiendo este último que es una exigencia del principio democrático que ‘’sean los propios ciudadanos quienes decidan qué objetos reúnen las condiciones requeridas para constituir bienes jurídico-penales’’ (p.132). Esta idea entronca con los postulados de Stratenwerth (2007), en los que se incide

en la importancia determinante a la hora de decidir estas cuestiones de la voluntad manifestada por los parlamentos. Ante esto, puede ser interesante la elaboración de criterios y requisitos de carácter procedimental más enfocados en el señalamiento de las condiciones desde las que esta labor deliberativa debe tener lugar que en indicar contenidos sustantivos específicos, siempre y cuando esta acción legislativa democrática se mueva dentro de los límites materiales impuestos constitucionalmente.

Nada hay que objetar, creemos, a una perspectiva apuntalada sobre el centro de gravedad del desarrollo del individuo en sociedad aunque no podamos delimitar claramente qué consecuencias tiene esta idea, pues lo queda latente en el fondo es en el mandato de que en esta deliberación sobre lo que es y lo que no objeto de protección no se pretenda la realización de una mera voluntad personal o colectiva, sino que se tienda a la conservación de las convenciones más básicas y elementales sobre las que se erige y configura la sociedad en un momento determinado. No creemos que esto sea, como afirma Roxin (2007), renunciar a todo el potencial crítico de la teoría del bien jurídico protegido, pues creemos que puede aportarse desde este prisma cierto carácter limitativo del poder punitivo, si bien desde el reconocimiento de las limitaciones explicativas de la teoría.

De hecho, el mismo Roxin (2007) extrae de este limitado marco teórico la idea básica de hasta nueve límites que se imponen al legislador penal. Algunos de los más reseñables son el que ‘’son ilegítimas aquellas normas penales que vienen exclusivamente motivadas por la ideología o atentan contra los derechos humanos y fundamentales’’ (p.438); que ‘’la mera delimitación de la finalidad de la ley no constituye todavía un bien jurídico […] con ello sólo se indica lo que quería el legislador. Lo importante es, sin embargo, si se ve perjudicada la coexistencia libre y pacífica de las personas’’ (p.439); que ‘’ni la autolesión consciente ni su posibilitación o favorecimiento pueden legitimar la amenaza de pena’’ (p.440); o que ‘’las leyes penales simbólicas no sirven a la protección de bienes jurídicos’’ (p.441), entre otros que señala.

Más allá de ello, merece especial atención uno de estos límites señalados por el autor para poder explorar desde otra perspectiva las limitaciones del bien jurídico. Este límite es el de que ‘’la mera ilicitud moral no basta para justificar una disposición penal. En tanto no lesione la libertad y la seguridad de nadie, no lesiona un bien jurídico’’ (Roxin, 2007, p.4369). Este límite nos resulta interesante porque produce un salto cualitativo en la ubicación sistemática en que se pretende que opere el bien jurídico, pues no se está tratando de delimitar específicamente qué puede proteger el Derecho penal y sólo justificar qué puede castigarse de forma indirecta por medio de la identificación de aquello que afecta a lo protegido; sino que se trata de dar una directriz adicional sobre la valoración normativa de la conducta lesiva del bien jurídico.

Cuando dice que una conducta ‘’en tanto no lesione la libertad y la seguridad de nadie no se lesiona el bien jurídico’’ (Roxin, 2007, p.439) se está pretendiendo decir que la función crítica del bien jurídico, que lleva a decir que sólo puede protegerse un bien en tanto sea esencial para las bases de una sociedad determinada, puede también determinar cuándo una conducta contra este bien jurídico, ya positivizado en su caso como bien jurídico dogmático-sistemático, es relevante jurídico-penalmente. Esto es extralimitarse en la función crítica del bien jurídico, que tan sólo limita el poder penal mediante la delimitación de las esferas sobre las que puede desplegarse su función protectora y nada dice de la conducta específica que puede castigarse. Aunque se trate de distinguir claramente, como Hefendehl (2007), el bien jurídico en sentido estricto —crítico— con el objeto sobre el que recae la acción delictiva —bien jurídico dogmático—; debe concederse que no toda conducta que exteriormente produce un resultado de menoscabo

de un bien jurídico constituye una conducta jurídico-penalmente relevante, porque, aunque mantengamos certeramente esta distinción, no todo curso causal desarrollado por una acción humana que produce una muerte puede considerarse una conducta típica de homicidio. Y esta muerte, que podría verse como un menoscabo del mero objeto del tipo penal, no puede producirse sin constituir también una afectación a la entidad conceptual que se entienda por el bien jurídico protegido crítico en virtud del cual puede castigarse el homicidio. La distinción entre el acto que afecta al bien jurídico protegido de un modo desvalorado y la que lo hace sin trascendencia penal es un trabajo que desborda las capacidades explicativas del bien jurídico, crítico o dogmático-sistemático.

Se puede ver entonces que, pese al valor que puede aportar esta perspectiva aquí defendida del bien jurídico, el único problema que atenaza al bien jurídico no es el de su vaguedad a la hora de ofrecer un marco explicativo completo de los límites que este supone para el *ius puniendi*, sino que, en palabras de Seelman (2007), carece de criterios de legitimación del castigo de una conducta específica (p.366). Dicho de otro modo, carece de fuerza explicativa para satisfacer el carácter fragmentario del Derecho penal, para delimitar cuándo una conducta que menoscaba un bien jurídico posee, por sí misma, una relevancia y gravedad suficiente para justificar su punición. Por ello Wohlers afirma que la ‘’legitimidad de un tipo penal no puede aclararse [exclusivamente, apuntillamos nosotros] a través de la remisión a un bien jurídico de por sí digno de protección. Decisiva es la relación en la que se sitúa el comportamiento cuestionado con respecto al «algo» protegido’’ (p.395). En definitiva, no sólo debe una teoría de legitimación de la intervención penal centrarse en qué se protege, sino en qué determina la relevancia penal del comportamiento que incide sobre lo que protege.

En el núcleo material de la antijuricidad penal, como señala Mir (1984), se encuentra, con carácter general, la existencia de una lesión de una esfera protegida y que, por tanto, se encuentra desvalorada —desvalor de resultado—; y junto a ello, el hecho de que dicho resultado desvalorado se haya producido como consecuencia de una conducta desvalorada —desvalor de acción— desde ‘’un punto de vista *ex ante*’’ (Mir, 1984, p.177) que se centra en determinar la ‘’peligrosidad para el bien jurídico que un espectador objetivo (el hombre medio) puede advertir en la conducta […]. El desvalor de conducta es, en este sentido, desvalor intersubjetivo ex ante’’ (Mir, 1984, p.177). Pero este desvalor intersubjetivo ex ante que pesa sobre la acción no puede radicar, si queremos ser coherentes con la búsqueda de un Derecho penal respetuoso con la libertad y dignidad de los individuos en un contexto políticamente plural, en que se trate de una conducta meramente molesta o que choca con nuestras convicciones morales. Creemos que se requiere entonces disponer de algún tipo de criterio que señale cuándo es legítimo el reproche jurídico-penalmente la conducta en sí misma, además, es claro, de la valoración de la adecuación social general de la forma en que se desarrolla la conducta, los estándares normales exigibles en el contexto y otras cuestiones. El bien jurídico nada puede aportar al respecto.

Podría parecer que esto es una cuestión que se debe responder sólo en sede del enjuiciamiento de una conducta respecto de un tipo preexistente y no a la hora de determinar la posibilidad de la formulación de una conminación penal. Sin embargo, lo que tratamos de defender aquí es que está definición de qué es una conducta castigada debe fundamentar los dos elementos principales que determinarán la antijuricidad de tal conducta: se debe tratar de una conducta desaprobada que produce una lesión del bien jurídico. Para la legitimidad de una propuesta de política criminal esta debe respetar un mínimo en la definición de ambos elementos. Tanto el desvalor de acción como de resultado deben partir de un pilar básico que permita decir que esta acción puede ser

considerada penalmente relevante, y que ese resultado lesivo también puede serlo. Se trata, pues, de identificar el núcleo duro de ambos ámbitos, cuál es la raíz del desvalor de acción y cuál es la del desvalor de resultado. En cuanto al resultado, podríamos partir de lo ya expuesto sobre la teoría del bien jurídico. Nos resta entonces avanzar en la consecución de un elemento mínimo común a todas las formas de desvaloración jurídico-penal de una acción.

Con una perspectiva únicamente basada en el bien jurídico protegido, corremos el riesgo de que, ante una conducta indeseada socialmente, problemática en algún sentido, simplemente se conceptualice y constituya un bien jurídico como materialización de aquello sobre lo que la conducta incide. El bien jurídico-penal se convierte así en un fetiche dogmático, la reificación del desvalor dado a la conducta, soslayando la pregunta de si es legítimo dar una significación penalmente relevante a tal conducta, dando por justificado el ejercicio del *ius puniendi* frente a cualquier conducta que se desee, máxime si se tiene en cuenta la notable indeterminación del carácter limitativo de la teoría del bien jurídico como criterio exclusivo. En este sentido, ante la disyuntiva que plantean Hefendehl, Hirsch y Wohlers (2007) en el título de su libro: ‘’La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?’’, no puede responderse en uno u otro sentido. El bien jurídico es ambas cosas, manteniéndose en un estado de superposición como un ‘’bien jurídico de Schrödinger’’, dado el carácter legitimador aquí reconocido y su posible configuración como mero fetiche si no se reconocen sus limitaciones a este respecto.

Desde estos presupuestos puede entenderse plenamente la necesidad defendida más arriba de conformar un modelo de criminalización completo para poder comprobar la legitimidad del castigo de una conducta completamente novedosa desde un tipo penal preexistente. Si exigimos que, además de proteger un bien jurídico cuya lesión encarne el desvalor de resultado, exista algún criterio que sirva de base para delimitar el desvalor normativo de una acción, todo tipo penal legítimo se sustenta, en definitiva, sobre la existencia de ambos criterios, separados pero interrelacionados. Si sólo se analiza el castigo de la conducta sobre la base de que esta puede incidir sobre un bien jurídico se obvia un paso previo y puede darse el castigo sólo desde el valor de resultado, sin garantías en lo relativo al desvalor de acción.

En cualquier caso, la conjugación de cumplir con esta necesidad apuntada de un criterio de legitimación centrado en la justificación del castigo acción en sí misma para evitar la consagración del bien jurídico como un mero fetiche y la idea aquí defendida de mantener, aunque en un sentido limitado, las aportaciones hechas por el principio de exclusiva protección de bienes jurídicos, vendrá por reconocer el valor legitimador de este al tiempo que la necesidad de un criterio adicional complementario enfocado en la acción.

¿Cuál podría ser este o estos criterios complementarios? Creemos que puede ser favorable seguir la propuesta de Seher (2007), que entiende que ‘’la teoría del bien jurídico necesita ayuda externa’’ (p.66) y que esta ‘’podría prestársela el discurso angloamericano’’ (p.66). Lo positivo de este ámbito anglosajón es que se concentra precisamente, como destaca Miró-Llinares (2015), no en determinar qué proteger, sino en qué castigar, por medio del *harm principle* y, además desde posiciones alineadas con las propuestas de Feinberg (1985), desde el *offense principle*, cuestiones que, como ya se ha indicado, son especialmente adecuadas para el debate sobre la criminalización de las ultrafalsificaciones.

En conclusión, puede verse que los criterios de daño y ofensa no son incompatibles con la idea del bien jurídico, sino que operan en un estadio previo al mismo. El daño y ofensa se centran en los caracteres de la conducta, pero su normativización como una conducta jurídico-penalmente relevante pasará, desde la perspectiva aquí desarrollada, por el requisito adicional de que estos afecten a aquello que se pueda considerar como un genuino bien jurídico protegido.

3. Los principios de daño y ofensa y su relación con el bien jurídico protegido

Suele tomarse como punto de partida del concepto de *harm principle* los postulados de John Stuart Mill (Hirsch, 2007; Miró-Llinares, 2015). Y es que Mill (1859) sentó las bases principales seguidas en el Derecho angloamericano para la justificación de la intervención punitiva del Estado desde presupuestos políticos liberales al tratar en su ensayo de explorar la libertad política y, con ello, ‘’la naturaleza y los límites del poder que la sociedad puede ejercer de forma legítima sobre un individuo’’ (p.37).

El autor entenderá que la vida en sociedad exige por sí misma cierta limitación de la libre autodeterminación de la conducta, una restricción de la más extensa libertad que se entenderá en la modernidad, desde sus presupuestos epistemológicos, como la plena autonomía (Cianciardo, 2000), constituyendo esta limitación de la libertad el que ‘’el hecho de vivir en sociedad hace indispensable que cada cual se comprometa a observar una cierta línea de conducta para con los demás. Esta conducta consiste, en primer lugar, en no perjudicar los intereses de otro’’ (Mill, 1859, p.171).

De tal modo, sólo podrá, para Mill (1859), entenderse legitimado el poder político a interferir en la esfera de libertad del individuo para la evitación de estos perjuicios a los intereses de los demás y sólo podrá castigarse efectivamente aquello que constituye un verdadero daño a otros, una auténtica lesión a intereses relevantes de titularidad de otra persona. Cuando se dé que una acción no presenta la entidad necesaria para ser considerada un auténtico daño para con los derechos e intereses de otro ‘’el culpable, entonces, podrá ser justamente castigado por la opinión, aunque no por la ley’’ (Mill, 1859, p.172).

‘’La que, sin embargo, es unánimemente considerada la versión más depurada del principio del daño es la del jurista americano Joel Feinberg’’ (Miró-Llinares, 2015, p.10). Por ello, sin ánimo de analizar extensamente qué es específicamente lo que entiende Mill (1859) por daño, sí es relevante destacar que sus planteamientos le llevarán a entender que el daño es ‘’un principio muy sencillo, capaz de regular por completo las relaciones de la sociedad con el individuo en todo lo que de obligatoriedad y control representan’’ (Mill, 1859,p.52), un principio exclusivo para permitir el castigo estatal y que, por consiguiente, este no estará autorizado para su despliegue por acciones que sólo dañan al que las realiza ni frente a aquellas que tan solo constituyan actos contrarios a las opiniones o moral mayoritaria y no verdaderos daños. Se construirá así el dogma clásico en el Derecho penal anglosajón de que sólo el daño en sentido estricto puede servir como criterio único de criminalización.

Como se ha indicado, será Feinberg (1984) quien habría realizado el desarrollo y conceptualización del *harm principle* más acogido en nuestros días, el cual es enunciado del siguiente modo:

> ‘’It is always a good reason in suport of penal legislation that it would probably be effective in preventing (eliminating, reducing) harm to persons other than the actor (the one prohibited from acting) *and* there is probably no other means that is equally effective at no greater cost to other values’’ (p.26)

Sin embargo, como señala Miró-Llinares (2015), este principio del daño de Feinberg (1984) contrasta con el *one very simple principle* de Mill (1859), y es que frente a la concepción de este último, como único y exclusivo principio que debe llevar a la criminalización de ciertas conductas; conforme al entendimiento del principio del daño de Feinberg (1984) este se presenta como una mera buena razón para apoyar una ley penal, lo que abre a que, en primer lugar, sea una simple posibilidad el castigo de los daños, y no una necesidad, y, en segundo lugar, que no tiene por qué ser el único criterio que permite el castigo. De hecho, para Feinberg (1984, 1985), cabrá complementar el principio del daño con el *offense principle*. En cualquier caso, esta cuestión de la exclusividad de un principio ya ha sido abordada más arriba, desde la perspectiva de la determinación de la clase de modelo de criminalización que buscamos, y será retomado posteriormente con el examen del principio de ofensa.

En este momento debemos tener presente cuáles son los motivos que nos llevan a indagar en el principio de daño. Como hemos indicado, lo que buscamos en él es algún tipo de criterio que nos sirva como base esencial del desvalor jurídico-penal de una acción al margen del objeto sobre el que recae, ámbito en el que nos hemos movido en torno al concepto de bien jurídico-penal. Es decir, buscamos un principio que legitime la criminalización de una conducta en sí misma como límite mínimo inicial referido al carácter y naturaleza de dicha acción que debe presentar cualquier conducta que pretenda ser castigada y al que se unirán elementos legitimadores adicionales.

Conforme al principio del daño que se ha expuesto por Feinberg (1984) podemos ver fácilmente que este se compone de distintos elementos, a saber, la exigencia de que la criminalización sea una vía eficaz para prevenir daños; el mismo concepto de daño, del que ya se excluye *ab initio* los daños autoinfligidos; y la exigencia de que no existan otras vías menos lesivas o costosas con igual eficacia para lograr esta prevención. En estos elementos se observan claras alusiones a principios que nosotros hemos exigido como consustanciales al Derecho penal de un Estado social y democrático de Derecho, como el de *ultima ratio* y la utilidad de la pena. Por ello, tratando de depurar lo máximo posible el principio para extraer de él aquello directamente alineado con nuestros intereses prácticos, debemos centrarnos exclusivamente en la conceptualización del *harm* en tanto tal y prescindir del análisis de otras cuestiones distintas a esta.

Para Feinberg (1984), las conductas que generan daños son el objeto sobre los que predilectamente opera el Derecho penal, entendiendo por tales conductas aquellas que causan ‘’harmed states or conditions in people’’ (p.31), pero entendiendo este daño desde una perspectiva normativa distinta de la mera producción de un detrimento físico de ciertos elementos exteriores presentes en el mundo exterior. Son, por tanto, actos que dañan o perjudican los intereses de otras personas, no simplemente un elemento fáctico, pero los cuales no son producidos por mero azar o eventos estrictamente naturales, sino que se producen por medio de un *wrong*, de una acción injusta, ilegítima o ilícita.

Puede verse de este modo que el concepto feinbergiano de daño se compone por dos elementos. El primero de ellos es el interés legítimo que se ve vulnerado, un interés entendido en un sentido distinto al concepto de interés como precio del dinero o como atención, preocupación o curiosidad, sino como una relación para con un elemento sobre el cual se deposita cierta esperanza en la obtención de un cierto rédito en términos de mejora de su bienestar general en la vida en sociedad, de tal modo que este bienestar de la persona que posee este interés ‘’flourish or languishes as they flourish or languish’’ (Feinberg, 1984, p.34). Este interés debe ser, además, un interés legalmente tutelable, entendiendo que sólo aquellos más directamente relacionados con la consecución del

bienestar de una persona a través de su mantenimiento indemne son dignos de protección legal, y dentro de estos sólo algunos merecen tutela penal; mientras que los más lejanos a estos, que tan sólo pueden relacionarse indirecta o remotamente con el bienestar individual, ‘’are for the most part not directly protectable by the law’’ (Feinberg, 1984, p.62) porque ‘’the law cannot protect me by interfering with the Liberty of those whose carácter and life-style falls below my standards’’ (Feinberg, 1984, p.62).

Desde una perspectiva similar Hirsch (2007) trata de especificar más en qué consiste un interés, definiéndolo como un ‘’recurso sobre cuya integridad tiene una pretensión la persona involucrada’’ (p.37), entendiendo por recurso como un ‘’medio o una capacidad que, en el caso normal, posee un vierto valor para el mantenimiento de un estándar de calidad de vida’’ (p.38) y asumiendo que esta pretensión de indemnidad debe poder serle atribuida normativamente acudiendo ‘’al Ordenamiento Jurídico primario —por ejemplo al Derecho civil—, o bien a presupuestos éticos, o a ambos’’ (Hirsch, 2007, p.38). Respecto de la noción, para Hirsch (2007), más vaga e imprecisa de Feinberg (1984) basada principalmente en la percepción del sujeto de qué son los intereses que se relacionan con su bienestar; el primero avanza en la exigencia de que la calificación de algo como un interés de alguien debe pasar por un juicio normativo prejurídico o a su reconocimiento por una norma jurídica extrapenal.

Sin embargo, pese a la crítica de Hirsch (2007) de la vaguedad del concepto de interés en la elaboración de Feinberg y su intento por concretarlo desde una perspectiva normativa, una propuesta similar ya la realizó el propio Feinberg (1984), quien entiende que el principio del daño no podría en ningún caso ‘’support the prohibition of actions that cause harms without violating rights’’ (p.36), por lo que vincula indisolublemente el interés, cuyo menoscabo será un daño, con un derecho, pero un derecho no entendido en el sentido de poder jurídicamente exigible ante los tribunales, que otorga una acción procesal, desde el que se concebiría el derecho subjetivo en sentido estricto desde el intento de Ockham de defender el uso por los franciscanos de bienes de primera necesidad sin poseer un auténtico derecho a ello, salvaguardando así el cumplimiento de su voto de pobreza (Hervada, 2000); sino como un ‘’moral right merely, that is a claim directed against one’s fellow citizens prior to and independent of any claim of enforcement against the state’’ (Feinberg, 1984, p.111), esto es, aquello que Guillermo de Ockham denominaría *ius soli* (Hervada, 2000), una facultad meramente moral para exigir aquello que conforme a Justicia y la recta razón le es debido a alguien. Con esto se llega entonces, al normativizar el interés tutelable por medio de la exigencia de que su lesión debe ser también la lesión de un [no]derecho, a una solución muy similar a la que llega Hirsch (2007).

El segundo de los elementos que compone el *harm principle* será el propio concepto de *harm*, el modo en el que se relaciona una conducta con el interés digno de protección de forma que el primero daña el segundo. Para Feinberg (1984) sólo puede existir un *harm* cuando este se produzca por una conducta con sentido normativo coherente con la idea de daño, y esto presupone en todo caso una conducta humana, de modo que, pese a que puedan existir eventos naturales que lesionan los intereses de las personas, las intromisiones en estos intereses que pueden constituir un verdadero daño deben presentar un elemento adicional, ‘’they can only be ‘’invaded’’ by human beings, either by myself, acting negligently or perversely. It is only when an interest is thwarted throught an invasion by self or others, that its possessor is harmed in the legal sense’’ (Feinberg, 1984, p.34), y esto será así aunque ‘’obviously an earthquake or a plague can cause enourmous harm in the ordinary sense’’ (Feinberg, 1984, p.34). Así, con la introducción del daño en el terreno de lo normativo se deberá entender como componente

necesario de todo daño la concurrencia de una acción mediada por la agencia humana, pues sólo las conductas de sujetos morales son capaces de ser motivadas por un orden normativo y ser juzgadas conforme a la exigencia dicho estándar de conducta. Ahora bien, no toda acción humana controlada y voluntaria se corresponde con el concepto normativo de daño. Esta valoración exige que se trate de un *wrong*, una conducta injusta o ilícita, de tal modo que ‘’one person *wrongs another when his indefensible (unjustificable and inexcusable) conduct violates the other's right”* (Feinberg, 1984, p.34); da suerte que se excluye del concepto normativo de daño ‘’those to which the victim has consented. These include harms voluntarily inflicted by the actor upon himself, or the risk of which the actor freely assumed […]” (Feinberg, 1984, p.35). Esto es, exige la existencia de daño necesariamente la conducta de un agente humano distinto del propio titular del interés que lesionado y que esta pueda ser considerada normativamente *wrongful*. En este sentido Hirsch (2007) entiende que desde los presupuestos feinbergianos el daño exigirá que sea ‘’realizada con dolo o imprudencia […] sólo puede ser condenado por las consecuencias lesivas de sus acciones cuando hubiera podido evitarlas’’ (p.36) porque la pena se entendería como ‘’una comunicación con el autor, entendido como una persona capaz de emitir juicios morales [y] ese contenido de reproche presupone necesariamente que la conducta punible sea merecedora de dicha censura’’ (p.36).

Estos dos elementos, en suma, se sintetizan por Feinberg (1984) de tal modo que ‘’only setbacks of interests that are wrongs, and wrongs that are setbacks to interest, are to count as harms in the appropriate sense’’ (p.36). Sólo puede, entonces, existir un daño conforme al *harm principle* en cuanto exista una conducta humana que lesiona injustamente un interés relevante digno de protección de otra persona.

Debe notarse por consiguiente que la perspectiva relacional entre acción e interés del principio del daño introduce en el ámbito del daño un elemento externo a la valoración de la mera conducta por sí misma, que es lo que nosotros buscamos principalmente. Pone énfasis en la necesidad de que la conducta dañosa lo sea respecto de aquello que se considere un auténtico y genuino interés digno de protección, pero esto ya se pone de manifiesto desde la perspectiva del principio de exclusiva protección de bienes jurídicos tal y como se ha conceptualizado más arriba. Por ello, para Hirsch (2007), ‘’podría afirmarse que a partir del *harm principle* puede construirse algo parecido a un bien jurídico’’ (p.39). Entonces, lo que buscamos del principio del daño, más que la exigencia de que sólo se castiguen como daños aquello que lesiona un interés más o menos identificable con nuestro bien jurídico-penal; es la concreción de la construcción normativa de la acción dañosa, siendo esta cuestión precisamente a la que la noción de bien jurídico no llega. Por ello estamos de acuerdo en la afirmación de que ‘’si no existe un *harm to others*, la teoría del bien jurídico no puede legitimar la criminalización de una conducta’’ (Hirsch, 2007, p.43), pero para encontrar este criterio de legitimación limitativo de la intervención punitiva debe indagarse precisamente en el acto de dañar y no en el interés que se daña.

Feinberg (1984) afirma tajantemente que ‘’not everything that we dislike or resent, and wish to avoid, is harmful to us. […] These experiences can distress, offend, or irritate us, without harming any of our interests’’ (p.45). De nuevo, sólo aquellas acciones que constituyen un ataque contra un interés son verdaderamente daños. Surge entonces la pregunta de cómo una acción puede ser lesiva para estos intereses. Feinberg (1984) explora en primer lugar, para responder a esta pregunta, ‘’the manner in which acts and other events affect interests when they harm them’’ (p.51), y en esta exploración constata que para identificar este ataque pueden emplearse hasta 8 verbos distintos:

‘’The acts (and sometimes but not always the events) that harm people have been said (alternatively) to be those that (1) violate, (2) invade, (3) impair, (4) set back, (5) defeat, (6) thwart, (7) impede, and (8) doom their interests. These terms for the effect on interests are not all exact synonyms, so it is worth sorting them out in the expectation that some may be more apposite than others’’ (p.51).

Pese a esta variedad, el autor pone especial énfasis en los verbos 4 a 7, identificando de algún modo todos ellos con una noción general de daño como *setbacks to interest*, basada en la frustración del interés, en la producción de condiciones en que este interés existente no se verá cumplido, no evolucionará hasta su perfección y no culminará la mejora del bienestar de su titular que sí se daría de no ser por las injerencias de un tercero. Para ilustrar esta idea emplea una analogía económica que parte de ‘’the idea of a starting point or ‘’baseline’’from which the direction of advance or retreat is charted and measured’’ (Feinberg, 1984, p.53), de tal modo que las acciones de terceros influyen en la dirección y aceleración de la curva de posibilidad de beneficio para el bienestar que genera aquello sobre lo que se tiene el interés. Desde esta perspectiva sólo existirá un daño ‘’when one’s interest is brought below the centerline, and thus put into a harmed condition’’ (p.54).

Sin embargo, esta analogía presenta problemas. Como reconoce el autor, ‘’these puzzles are complicated enough in the economic real. When it comes to human well-being in general, they are without answers’’ (Feinberg, 1984 p.54). Pero no es ese el único problema que entender cómo se produce un daño de este modo presenta para con nuestros propósitos en este trabajo. Concebir el daño como un *setback of interest* sigue poniendo el foco en el resultado de la acción desvalorada, muestra cómo esta produce un perjuicio a aquello que debe protegerse, y si bien es necesario explorar las formas en que desde un perspectiva normativa puede entenderse que una acción desvalorada causa un resultado también desvalorado —cuestión que, a nuestro juicio, queda más satisfactoriamente resuelta desde la perspectiva de la generalmente aceptada en nuestro entorno teoría de la imputación objetiva—; es una línea que no nos aporta respuestas sobre el elemento nuclear del desvalor de acción, que conforme a lo aquí expuesto será el *wrong*, ya que ‘’not all invasions of interest are wrongs, since some actions invade another’s interests excusably or justifiably, or invade interests that other has no right to have respected’’ (Feinberg, 1984, p.35). Esta frustración de intereses debe entenderse, entonces, como el resultado de una conducta desaprobada para que pueda constituir un verdadero daño en sentido normativamente propio. Debe verse, más que cómo el daño es un *setback interest*, cómo el autor entiende que una conducta es *wrongful.*

En este sentido, ‘’a harm is a *wrongfully* set-back of interest’’ (Feinberg, p.105), y esta acción de daño injusto se produce cuando se da una unión de varias condiciones, a saber, que se dé una acción y que esta se dé ‘’in a manner which is defective or faulty in respect to the risks it creates […] with the intention of producing the consequences […] or with negligence or reckleslessness in respect to those consequences and […] acting in that manner is morally indefensible, […] is the cause of a setback interest, which is also a violation of [a] right’’ (Feinberg, 1984, pp. 105-106). Esto es, debe darse una acción humana realizada con conocimiento del riesgo que esta genera —preferimos esta concepción del dolo a la que lo identifica con intencionalidad—, o con imprudente o temerario desconocimiento de dicho riesgo, que no está justificada o es moralmente indefensible y que produce una lesión de un interés a cuya indemnidad tiene derecho el que la padece.

Conforme a lo anterior ya podemos dejar claro un primer elemento del desvalor de la acción: sólo pueden desvalorarse jurídico-penalmente acciones que consideraríamos dolosas o imprudentes. Nunca será legítima la criminalización de eventos fortuitos, producidos por accidente imprevisible y que, desde esta perspectiva, sería inevitable o no podría exigirse normativamente que el agente lo evitase. Está claro que si lo que se buscan son elementos que permitan, desde un juicio normativo, el reproche o censura de una conducta esta sólo se podrá dar respecto de aquellas que podrían haberse evitado por entenderse dolosas o imprudentes (Hirsch, 2007), pero no puede exigirse una inevitabilidad absoluta pues, como indica Jakobs (1991), la evitación de todo contacto social también evita plenamente cualquier tipo de conflicto, pero esto no es admisible en la vida humana, que se desarrolla en sociedad y con necesaria interacción con otros. Este es un primer elemento extremadamente básico y elemental, pero sienta en primer lugar las bases de lo que debe ser la desvaloración penal legítima de una acción.

El otro elemento nuclear del concepto de *wrong* es el que sea una acción moralmente indefendible, injusta e injustificable, y que genera una vulneración de un derecho de otra persona a que no se lesione un interés. Como hemos expuesto más arriba, Feinberg (1984) no se refiere exactamente a un derecho subjetivo en sentido estricto cuando exige que la conducta *wrongful* lesione un derecho, sino que alude a un concepto de derecho como lo justo o lo que le es debido a alguien por criterio de Justicia. Por ello creemos que, en definitiva, la cuestión de lo moralmente indefendible y la de la afectación a un derecho son dos elementos que, en los términos feinbergianos, se pueden reconducir a un mismo ámbito de valoración moral de la realidad empírica. No obstante, parece también adecuado incluir la posibilidad señalada por Hirsch (2007) de que este derecho a no sufrir injerencias ilegítimas en un interés determinado venga dispuesto por una norma del Ordenamiento Jurídico primario, como el Derecho constitucional, el civil o, incluso, el administrativo. Sin embargo, como apunta Feinberg (1984), esta atribución de un derecho legal sólo puede darse, para evitar un argumento circular, sobre la base de una valoración moral que determina que efectivamente una persona *debe tener* este derecho. Por consiguiente, no existen grandes impedimentos a circunscribir ambos ámbitos, lo injusto por injustificable de la conducta y la existencia de un «derecho» a la indemnidad del interés, al campo de la valoración normativa prejurídica de las dinámicas y contactos sociales.

Aunque Feinberg (1984) les da un tratamiento en orden inverso, nos parece más adecuado para nuestros fines expositivos abordar en primer lugar la cuestión del derecho moral, o en mejores términos, del reconocimiento a alguien de la facultad para exigir a los demás un cierto respeto a sus intereses desde una perspectiva moral.

Reconoce Feinberg (1984) que tratar de identificar estos «derechos morales», ‘’that are independent of and antecedent to law, a conception strongly reminiscent of the doctrine of «natural rights» (p.111), supone cargar con la labor de ‘’distinguishing those of our interests that ground morally valid claims to respect and noninterference from our fellows, from those of our interests that do not’’ (p.111). No es nuestra voluntad abrazar una construcción iusnaturalista sobre estos derechos y tratar de construir a partir de ella todo un catálogo que nos permita identificar cuáles son concretamente, siendo notablemente escépticos acerca de la posibilidad de lograr tal cosa. Sin embargo, como indica Feinberg (1984), esto no es estrictamente necesario, pues entiende que de forma intuitiva podemos identificar que ‘’certain kinds of morally disresputable interests can be ruled out, straightaway, as posisible grounds for valid moral claims’’ (p.111) y, a partir de ellos, encontrar qué otros intereses, en principio, pueden ser protegidos por la exigencia moral de su respeto o pueden ser objeto de una pretensión moral legítima de respeto.

> If there are any interests in causing pain and suffering for their own sakes, for example, such interests cannot be the grounds of claims against others. Cruel and sadistic interests, mobid interests, wicked and sick interest, if there are such things, can be peremptorily ruled out of court, and put aside. No one has a moral right to the protection of such interests, if, indeed, such things exist at all'' (Feinberg, 1984, pp. 111 y 112).

De tal modo se da una delimitación negativa de los posibles candidatos a ser reconocidos como derechos morales a exigir de los demás consideración y no menoscabo de un interés determinado. En un sentido inverso, pues, queda que ''any interest at all (apart from the sick and wicked ones) is the basis of a valid claim against others for their respect and noninterference'' (Feinberg, 1984, p.111).

Dentro de este campo delimitado negativamente parece que no existen certezas ulteriores sobre cuáles son los criterios para determinar cuándo existen tales derechos, quedando sólo el recurso a la argumentación y la contraposición de ideas en casos concretos para la búsqueda de ciertos consensos, una labor que aquí no puede desarrollarse, si es que se puede en absoluto encontrar una verdad a este respecto. Pese a ello, Feinberg (1984) sí que apunta a un determinado criterio que parece razonable a la hora de determinar el respeto de qué intereses puede ser exigencia una exigencia moral. Este se basa en la distinción entre *welfare interest* y *ulterior interests*, esto es, entre aquellos intereses cuyo mantenimiento y perfección se vincula directa y fuertemente con la consecución del bienestar de la persona y aquellos otros que sólo se relacionan indirecta o vagamente con este bienestar y que tan sólo pueden existir si los primeros también permanecen indemnes. Entiende Feinberg (1984) que los *welfare interests* son los más evidentes fundamentos de la exigencia moral de respeto, de modo que ''if we can speak of moral rights at all, then, each of us has a moral right to life minimal health, economic suffiency, political liberty, and so one'' (p.112), concibiéndolos como ''the grounds for valid claims against others (moral rights) *par excellence"* (p.112). En cambio, reconoce las posibles dudas respecto de ulteriores intereses, pero afirma que ''at least one class of ulterior interests are directly vulnerable: those that consist of the extensión of welfare interests to transminimal levels'' (p.112). Por consiguiente, aunque quepa el debate y discusión respecto de otros intereses no básicos y elementales para el bienestar de la persona, sí entiende el autor que fuera de toda duda puede considerarse que la no lesión de intereses que se derivan de los intereses primarios y constituyen una mejora de ese bienestar elemental es algo moralmente debido a todos.

Sea como fuere, dado que esta perspectiva se basa en la fundamentación de las exigencias morales de respeto y no intromisión en las cualidades de estos intereses, en definitiva, en que se traten de auténticos intereses dignos de protección, entenderá Feinberg (1984) que de ella se sigue que ''any indefensible invasion of another's interest (excepting of course the sick and wicked ones) is a wrong'' (p.112). Efectivamente, si la existencia de cualquiera de estos intereses es el fundamento mismo del derecho moral a exigir su respeto, el centro de gravedad del carácter injusto o ilícito de una conducta, de su *wrongfullness*, se desplaza hacia la naturaleza moralmente injustificable de tal intromisión en el interés de otro, pues no puede existir acción humana que menoscabe un interés que sea verdaderamente tal sin que exista, en estos términos expuestos, en quien la sufre un derecho moral a exigir que tal conducta no tenga lugar, y sólo podrá decirse que esta acción es un *wrong* en tanto se añada el otro elemento esencial señalado, la ausencia total de justificación moral.

Feinberg (1984) utiliza el concepto de ‘’indefensible conduct as the most generic term for actions and omissions that have no adequate justification or excuse’’ (p.108). Alude así a las causas de justificación y de exclusión de la responsabilidad penal como criterios a tener en cuenta desde una perspectiva anterior a la tipificación para poder valorar una conducta por sí misma como algo injusto, reprochable; pues entiende que ‘’excused or justified wrongdoing is not wrongdoing at all, and without wrongdoing there is no harming, however severe the harm thay might have resulted’’ (Feinberg, 1984, p.109).

Por *excuse* entiende el autor aquella circunstancia que produce que, aunque un hecho pueda considerarse injusto, malo, o indeseable, ‘’it is not quite fair or correct to say baldly, or without qualification, that one did the thing at all, that it was one’s action’’ (Feinberg, 1984, p.108). Por otro lado, cuando se alude a la justificación de la conducta se trata de decir que la conducta, que es propia de una persona y de la cual se le puede considerar responsable, ‘’was a good thing, or the sensible thing, or a permissible thing to do, either in general or at least in the special circumstances of the occassion’’ (Feinberg, 1984, p.108).

Es decir, sólo permitirá entender como legítima la propuesta de criminalización que parta de desvalorar una acción que, al margen de los elementos y la configuración del tipo específico que eventualmente pudiere tener, no estuviere mediada por circunstancias que en condiciones normales determinarían la eliminación de su valoración negativa y que, si se diese en los términos ideales en los que se concibe normativamente, sería una acción culpable, se podría imputar como conducta desvalorada a título de demérito al autor.

Esto es, si en términos generales se concibe la legítima defensa como una causa de justificación generalmente aceptada y aplicable que determina —en términos de la teoría de los elementos negativos del tipo— que una conducta es atípica, o la inimputabilidad como una circunstancia que suprime la culpabilidad; no podría entenderse aceptable criminalizar «la producción por medios congruentes, oportunos y proporcionales de lesiones a aquel que ataca ilegítimamente al autor del hecho» o «las lesiones producidas por una persona en una situación de total enajenación mental», interpretando la acción misma y las circunstancias justificantes o de excusa en las que se da de forma holística como una unidad, pues de forma independiente al interés que se lesione, su relevancia y consideración como bien jurídico; no puede darse desde una perspectiva normativa el desvalor jurídico-penal de la acción, porque no constituiría una acción moralmente indefensible y, por tanto, no puede ser una conducta injusta, un *wrong*.

Puede ser ciertamente confuso tratar de aplicar figuras que pertenecen al sistema que compone la teoría del delito, y que además se encuentran sistemáticamente en lugares de una naturaleza muy distinta, como son las causas de justificación y de exclusión de la responsabilidad penal; al ámbito de la legitimidad del castigo, pues estas presuponen la existencia de la conminación penal y operan en el proceso de concluir si una acción se corresponde con el delito previsto por el legislador o no. Sin embargo, y sin dejar de reconocer la complejidad dogmática del asunto, parece un modo lógico de proceder a identificar qué conductas pueden considerarse *wrongs*, acciones que se desvaloran de tal modo que pueden ser consideradas penalmente relevantes; pues estas causas precisamente a lo que aluden es a ciertas excepciones a la significación negativa de una conducta —justificación— o bien a la posibilidad de imputar moralmente un hecho a un agente y reprocharle su realización —exculpación—. Pese a la posible utilidad de tener en cuenta ambos tipos de circunstancias, creemos sin embargo que por lo general será más eficaz

remitir únicamente, para delimitar el concepto de *wrong*, a las causas de justificación, que son precisamente las que versan sobre la significación normativa de la conducta, sobre su desvalor o no; mientras que las causas de justificación aluden a criterios para delimitar el merecimiento o no de la pena en base a si es posible reprochar al autor la realización de la conducta. Mientras las primeras operan sobre la acción en sí misma las segundas no delimitan qué es la acción desvalorada, sino cuándo puede existir responsabilidad criminal derivada de las acciones típicamente antijurídicas. Es aceptable la alusión de Feinberg (1984) al requisito de que la acción pueda considerarse como una acción de alguien, que ese alguien ha realizado una acción, pero quizá esta es una cuestión que ya queda suficientemente cubierta con la exigencia de la agencia humana en la producción del *harm*, esto es, que constituya una acción libre y voluntaria de un ser humano, sin necesidad de acudir a las causas de exclusión de la responsabilidad penal.

Ahora bien, quizá lo que buscamos no sea simplemente remitir a estas causas de justificación, por así decirlo, dogmáticas existentes. Parece que, en esta perspectiva previa de valoración de una conducta no criminalizada con el fin de decidir su proscripción penal o no, en realidad, se tendría que atender a los motivos y argumentos que subyacen a estas causas de justificación, a la forma en que operan para justificar las conductas, el modo en que se entiende que su concurrencia elimina el significado de desvalor social que se atribuye a la conducta típica. Es decir, por ejemplo, qué es lo que determina que la defensa o el ejercicio de un derecho sean legítimos y por qué, por ello, se elimina el desvalor de acción de la conducta, su carácter injusto, su naturaleza de *wrong*. Esta perspectiva nos ofrece ciertos criterios y argumentos aceptados, pues generalmente aceptada está la legítima defensa y otras causas de justificación, que permitan concluir ante la observación de una conducta que esta no es injusta.

En cualquier caso, como la identificación de los *wrongs* es en todo caso una cuestión harto compleja, ‘’Feinberg introduce toda una serie de *mediating principles* que han de ser tenidos en consideración antes de que intervenga el Derecho penal’’ (Hirsch, 2007, p.37), pudiendo sernos de utilidad estos criterios a la hora de realizar esta tarea. Algunos de los más relevantes para este fin se refieren a la magnitud y probabilidad del daño que pueden provocar las conductas, exigiendo un cierto grado de gravedad en las posibles consecuencias y que el riesgo de su producción no sea excesivamente remoto o vago; la comparación entre el grado en que la libertad de acción queda mermada por la prohibición penal y la entidad del daño que con esta restricción se pretende evitar y, especialmente, la valoración de la importancia social de la conducta que produce un perjuicio y los intereses que la motivan frente a la del interés que resulta perjudicado (Feinberg, 1984).

A este último criterio o principio se refiere Hirsch (2007) como ‘’el valor social del comportamiento’’ (p.37) y creemos que en él radica, precisamente, la cuestión sobre la identificación de una conducta como *wrong* o no. Si tratamos de determinar cuándo existe una conducta injusta por ser moralmente indefensible, y hemos dicho que la alusión a las causas de justificación previstas legalmente y generalmente admitidas por la dogmática penal puede ser una estrategia útil para ello si se alude a los motivos y argumentos que subyacen en la aceptación de tales causas como eliminatorias del carácter injusto de esta acción; estamos remitiendo entonces la evaluación de lo *wrongful* al campo de las valoraciones ético-morales desde la perspectiva de una sociedad determinada, donde se concreta si una acción es adecuada o inadecuada a nivel social.

La adecuación o inadecuación social nos parece, en realidad, el elemento más relevante a tener en cuenta y que, en definitiva, engloba las causas de justificación, al

menos en el sentido que hemos indicado antes, pero no se agota en ellas. Lo indefensible moralmente no puede definirse meramente desde la ausencia de causas de justificación sistemáticas y admitidas dogmáticamente, pues así sólo bastaría para decir que una acción puede ser portadora del significado desaprobado necesario para ser desvalorada penalmente con que se tratase de una acción que, conforme a su aparición normal en la dinámica socioeconómica, no constituye una legítima defensa, el ejercicio legítimo de un deber, derecho, oficio o cargo, ni un estado de necesidad justificante. Nos permitiría, en realidad, desvalorar un amplio abanico de acciones, pues, como señala Mir (1984), ‘’no siempre que se realiza una conducta socialmente adecuada que en absoluto determina reprobación social […] puede afirmarse el ejercicio de un derecho, ni menos el cumplimiento de un deber’’ (p.535), y, por consiguiente, *sensu contrario*, cuando no exista un derecho —o legítima defensa, etc.—, pese a existir adecuación social de la conducta, sería legítimo el castigo. Lo relevante entonces es entender que un *wrong* sólo puede ser aquella acción respecto de la cual no quepa justificación en términos morales generalmente aceptados, que no son en absoluto adecuadas en el desarrollo normal de la vida en sociedad tal y como lo entendemos, debiendo tener en cuenta que no puedan entenderse como ejemplos de causas de justificación jurídicas pero no sólo ellas, sino que tampoco debe existir ninguna otra buena razón a nivel social para defender y entender justificada tal conducta.

Esto significa, en definitiva, que sólo puede considerarse injusta, como base del desvalor de acción, aquella conducta que es inadecuada socialmente, siguiendo así el principio de que ‘’no puede ser voluntad de la ley, al delimitar las conductas penalmente relevantes —función propia de los tipos (penales)—, el incluir actividades socialmente adecuadas’’ (Mir, 1984, p.534). Por debajo de este umbral de la adecuación social sólo puede existir riesgo permitido, aunque sea en efecto un riesgo que pueda producir daños en un sentido amplio a intereses relevantes; y la materialización del riesgo permitido, en tanto la conducta que lo despliega no puede ser un *wrong* por estar justificada, no puede constituir un *harm* en los términos de Feinberg (1984).

Este umbral entre la adecuación e inadecuación social de una conducta, en definitiva, radica en lo esperado socialmente en un ámbito de la vida determinado y en condiciones normales. Desde este prisma pueden existir conductas peligrosas o lesivas de intereses que, sin embargo, se entiendan como algo normal, aceptable o incluso deseable en atención a los beneficios agregados que comporta. Sólo aquellas acciones que se alejan de las expectativas sociales de ordenación y orientación de la conducta en un espacio de la vida socioeconómica pueden considerarse *wrongs* y, por tanto, sus consecuencias *harms.* Por el contrario, todo aquello que se realice conforme a los estándares básicos de conduta exigibles desde una perspectiva normativa, al no poder calificarse como un mal, como un injusto, no podrán ser considerados *wrongs* y, por tanto, no podrán constituir en absoluto un daño. En consecuencia, no podrán ser criminalizados. Así, la creación de un delito debe partir entonces de la base de que la acción que toma como referente es absolutamente inadecuada socialmente, contraria a la forma de actuación exigible conforme a la ordenación de los intereses en la sociedad y afecta por su desorganización a un elemento central del funcionamiento social. Este es el componente común de cualquier conducta desvalorada jurídico-penalmente cuya presencia habilita su tipificación siempre que se unan los restantes requisitos indicados. Sin este mínimo común divisor de la acción injusta no puede existir propuesta de criminalización legítima.

Se trata entonces de un requisito mínimo basado en un juicio moral informado por la estructura social. Pero con este paso se normativiza completamente el criterio y nos devuelve a la vacuidad del bien jurídico como criterio único de criminalización, pues no

es nada preciso ni materialmente sustantivo. Sin embargo, el hilo conductor de las causas de justificación señalado por Feinberg (1984) sí puede darnos una directriz en tanto nos da el punto de apoyo para un juicio comparativo. Las distintas formas de acción adecuada socialmente deberían presentar cierto grado de semejanza con los motivos que nos llevan a la justificación de ciertas conductas en el ámbito de la teoría del delito, pues estas deben presentar una raíz común a todas las formas de adecuado desarrollo de la acción en sociedad. Más allá de esta aportación de Feinberg (1984) nos resulta imposible, sin que quepa que ello despierte sorpresa alguna, lograr una respuesta clara y materialmente sustantiva a la cuestión de la conducta injusta y el desvalor de acción. Sólo podemos, de nuevo, remitir a un campo social complejo y abstracto en el que se debe deliberar, argumentar y razonar en lo relativo la delimitación de lo adecuado o inadecuado socialmente en un contexto social, político y económico determinado, pero del que difícilmente pueden extraerse verdades últimas y universales.

Sólo podemos concluir entonces que esta inadecuación social, que en suma se basará en la ausencia de criterios morales aceptados en la comunidad que permitan entender que existen buenos motivos para actuar de tal modo, constituirá así el núcleo más elemental del desvalor de la acción, el requisito esencial para poder considerar una conducta como jurídico-penalmente desaprobada. Pero está claro que no basta por sí solo, es el requisito mínimo y elemental de que constituya una acción humana controlada, dolosa o imprudente e injustificable moralmente. Sin embargo, para que la posibilidad de reproche penal se perfeccione esta debe estar efectivamente relacionada con un interés que igualmente consideremos digno de protección que se lesiona pese a que su titular se ve amparado por razones socio-morales que indican que dicho interés debe permanecer indemne frente a las acciones de terceros.

Como ya se ha indicado, el principio del daño o *harm principle* ha sido visto tradicionalmente desde las posiciones mayoritarias de la tradición jurídica anglosajona como el criterio de criminalización fundamental de un Derecho penal de corte liberal respetuoso con la libertad individual. Este principio de daño ‘’at least in that vague formulation it is accepted as valid by nearly all writers. Controversy arises when we consider whether it is the *only* valid liberty-limiting principle, as John Stuart Mill declared’’ (Feinberg, 1985, p.IX). Sin embargo, pese a esta posible controversia, que puede radicar ya en la obra del propio Mill, que admite la posibilidad de criminalización de ciertas ofensas desde su definición como inmoralidad pública (Miró-Llinares, 2015); lo cierto es que se ha aceptado generalmente que el *harm principle* sería el único criterio válido para legitimar la intervención penal (Miró-Llinares, 2015; Hörnle 2019), alineándose de cierto modo con la tradición de influencia alemana basada en la teoría del bien jurídico protegido (Hirsch, 2007).

Más arriba, en la discusión sobre la tipología de modelo de criminalización por la que creemos que debe optarse, se ha concluido que puede ser favorable el establecimiento de varios principios de criminalización básicos y que permita responder a las pulsiones sociales tendentes a exigir el castigo de ciertas conductas de naturaleza distinta dando debida cuenta de esta diversidad y de la diferencia del grado de gravedad que estas presentan y, por tanto, limitar respecto de aquellas menos graves la posible respuesta punitiva legítima, configurando de este modo una suerte de Derecho penal de distintas velocidades, al estilo de lo propuesto por Silva (2001). Asimismo, también se ha señalado que tomar como uno de estos criterios el principio de ofensa puede sernos de utilidad para nuestros fines porque pueden relacionarse íntimamente con el fenómeno de los *deepfakes* y porque, de hecho, en nuestros códigos penales han aparecido delitos que pueden subsumirse en aquello que se entienda por ofensas y que no constituirían verdaderos y

auténticos daños en sentido estricto, como los delitos de maltrato animal (Wohlers, 2007), los de utilización de símbolos nacionalsocialistas o de negación del genocidio nazi en Alemania (Hörnle, 2007), o el discurso de odio en España (Miró-Llinares, 2017).

Si lo que tratamos de defender aquí es que la inclusión del principio de ofensa como criterio de criminalización legítimo permite dar respuesta a las nuevas sensibilidades sociales y explicar claramente cuál es el elemento común de estos delitos, podría reprocharse que se está concediendo demasiado a la voluntad del legislador y que sería más idóneo mantener un único principio de daño y explicar la legitimidad del castigo de ciertas conductas sólo desde este daño, admitiendo únicamente la posibilidad de castigo de alguna de estas nuevas figuras delictivas desde la premisa de que son, en realidad, algún tipo de daño, aunque menos cualificado o difuso.

Con la inclusión de esta suerte de «daño impropio» podríamos igualmente graduar el marco penológico admisible en función de si se trata del castigo de un daño en sentido estricto o uno impropio y mantendríamos la necesidad simbólica y sentimental de negar el castigo de conductas ofensivas como exigencia de un Derecho penal liberal. Sin embargo, el riesgo que se corre sigue siendo el mismo al señalado por Miró-Llinares (2015), a saber, la desvirtuación del concepto del daño y la difuminación de su naturaleza, apareciendo así el problema de tener que determinar qué daño es más daño, lo que supone en definitiva la eliminación práctica de cualquier barrera de contención dentro de la categoría conceptual del daño que impida el castigo más severo a la conducta menos grave, puesto que todo es daño y la frontera entre el daño puro y el irregular está desdibujada. Sólo creando dos compartimentos estancos, el del daño y la ofensa, que por supuesto permiten gradaciones intraclase de su gravedad, que permitan su distinción clara, tajante y heterodoxa; podremos asegurar que aquello que no se puede considerar en sentido estricto como daño pero sí consideramos normativamente como merecedor de algún tipo de reproche se castigue con una pena cuantitativa y cualitativamente inferior —y siempre que se respete el principio de intervención mínima y última ratio— y que sea imposible su confusión con una conducta dañina en un sentido estricto.

En definitiva, si todo es daño, pero de distinta clase, habremos desvirtuado, para tal ampliación, el concepto de daño hasta tal punto en que será profundamente complejo diferenciar estas clases, corriendo el riesgo de imponer penas de prisión a conductas que conforme al principio de la ofensa, en los términos indicados por Miró-Llinares (2015), no podrían considerarse más que, en el peor de los casos, delitos leves; o para las que el propio Feinberg (1985) sólo excepcionalmente admitiría este castigo y que, cuando se estableciera, ‘’it should be measured in days rather than months or years’’ (p.4). De este modo, aunque a priori pueda resultar simbólicamente chocante, la admisión del principio de ofensa puede resultar más garantista, canalizando a un marco penológico atenuado este tipo de conductas sin forzar conceptualmente la naturaleza de estas para mantener la funcionalidad del sistema penal ante las nuevas exigencias sociales.

Qué es, entonces, aquello que debemos entender por principio de ofensa es la pregunta lógica que se sigue de lo anterior. Conforme a la noción de Feinberg (1985), este principio se formula de la siguiente forma:

> ‘’It is always a good reason in support of a proposed criminal prohibition that it would probably be an effective way of preventing serious offense (as opposed to injury or harm) to personas other than the actor, and that it is probably a necessary means to that end’’ (p.1).

Al igual que en su formulación del *harm principle*, Feinberg (1985) incluye en el principio ciertas nociones limitativas del poder punitivo del estado relacionadas con las exigencias de utilidad de la pena o el principio de intervención mínima. Estos aspectos, en el modelo de criminalización que tratamos de esbozar aquí, se localizan sistemáticamente en otro lugar, por lo que no nos son interesantes en este momento. Del *offense principle* nos interesa, ante todo, qué es esta *serious offense* que el autor opone al concepto de daño.

El núcleo esencial del *offense principle* es, evidentemente, la ofensa, la cual es concebida por Feinberg (1985) como un conjunto de ‘’disliked mental states’’ (p.1), pero que en un sentido jurídico estricto sólo puede referirse a aquellos estados mentales de disgusto que son causados ‘’by wrongful (right-violating) conduct of others’’ (pp.1-2). De este modo, se sintetiza entonces que ‘’the offense principle then cites the need to prevent some people from *wrongfully offending* (offending and wrongdoing) others as a reason for coercive legislation’’ (Feinberg, 1985, p.2), entendiendo que si bien una ofensa, como creación de un estado mental de disgusto, ultraje o afrenta, no puede ser nunca un daño, pues no supone un menoscabo injusto de los intereses relevantes de otra persona, pueden ser criminalizados, pues estos estados ‘’are surely evils, thought not as great as evils as actual harms’’ (Feinberg, 1985, p.25), son verdaderos males que se causan a personas, personas ‘’with genuine grievances and a right to complain against determinate wrongdoers’’ (Feinberg, 1985, p.25).

Puede verse entonces que el concepto aquí manejado de ofensa como criterio para desvalorar jurídico-penalmente una acción debe componerse por dos elementos esenciales, por un lado, la producción de la ofensa en sí misma, lo que para Feinberg (1985) se determinará por la producción de estados mentales de disgusto o desagrado de cierta entidad; y una determinada entidad valorativa de la acción por la que se produce que haga que esta pueda considerarse un *wrongdoing*, una actuación injusta que vulnera un derecho —entendemos que para el autor, al igual que en el principio de daño, se tratará de derechos morales— de otra persona a no ser ofendido de tal modo.

Esto se descompone a su vez por Feinberg (1985) en un total de tres elementos de la ofensa, a saber, ‘’(a) I suffer a disliked state, and (b) I attribute that state to the wrongful conduct of another, and (c) I *resent* the other for his role in causing me to be in the state’’ (p.2). Sin embargo, creemos que puede mantenerse la redirección a dos elementos principales, el estado mental y el *wrong*, puesto que la cuestión sobre la recriminación al actor de su rol en la producción de este estado mental se puede circunscribir al ámbito de la calificación de esta conducta como un *wrongdoing*.

Respecto del primero de los elementos, esto es, el estado ofendido a que se induce al sujeto pasivo de la ofensa, se ha criticado de la postura de Feinberg (1985) que es excesivamente subjetivista, y que se enfoca simplemente en identificar este con un estado mental estresante o ultrajante del individuo particular, prescindiendo de la observación «socialmente objetiva» de la ofensa, como se propone por Hirsch (2007) o por Miró-Llinares (2015). Si bien es cierto que Feinberg (1985) parece acercarse a abrir esta posibilidad de objetivación, pues indica que ‘’offense in the sense of the offense principle specigies an objective condition —the unpleasant mental state must be caused by conduct that really is wrongful—‘’(p.2) así como que estos estados mentales no sean ‘’the producto of abnormal susceptibilities’’ (p.33); esta tendencia hacia la objetivación por medio de la normativización de la ofensa es algo confusa en la propuesta de Feinberg (1985) y parece referirse no a la necesidad de que la ofensa sea en cierto modo objetivada, sino a que lo sea el *wrong*, en sentido de una conducta injusta grave, dando por sentada

la validez del estado mental causado siempre que no se deba a una especial susceptibilidad del receptor del mensaje ofensivo, pero sin aludir explícitamente a que lo relevante es el carácter ilícito, socialmente injusto, de la actuación.

Por ello concordamos con la opinión de Miró-Llinares (2015) de que

> ‘’sobre la base de dichos estados subjetivos, es posible objetivar o generalizar el carácter ofensivo de algunas conductas. Es decir, es posible afirmar que una conducta, por su potencialidad para afectar a la sensibilidad de cualquier persona independientemente de las características personales que tenga, es ofensiva. Esta idea objetiva de ofensa va más allá de la propia sensibilidad de la persona: es posible que alguien sea objeto de una ofensa y no se sienta ofendido y también que alguien se sienta ofendido y no sea objeto de una ofensa’’ (p.56).

Lo esencial será, entonces, no el disgusto producido a una persona en concreto, sino la valoración de la conducta en sí misma, independientemente de los sentimientos que esta produzca. Lo que se debe realizar para identificar una ofensa será un juicio en que se concluya ‘’no sólo sobre la base de normas jurídicas concretas sino, incluso, de valoraciones sobre lo socialmente permitido, que la conducta es, en sentido general, ofensiva’’ (Miró-Llinares, 2015, p.57).

En esta labor pueden ser útiles los *medating principles* que propone Feinberg (1985) a modo de criterios adicionales para la aplicación del principio de ofensa. Estos son concebidos por el autor como dos bloques de principios que deben contraponerse para decidir si está justificada la intervención penal, y en definitiva pueden ser un buen indicador de si la conducta analizada es un *wrong*, un ilícito moral, o no.

El primero de estos bloques es el de ‘’the seriousness of the offense’’ (Feinberg, 1985, p.25). Este se encuentra integrado por tres principios, a saber, el de la magnitud o gravedad de la ofensa, que viene determinada, a su vez, por la intensidad de la misma —que conforme a lo aquí expuesto deberá valorarse desde una perspectiva de adecuación y admisibilidad social, remitiendo el propio Feinberg (1985) a la visión de un ‘’stantard observer’’(p.27)—, por la duración de la misma, y, aunque lo mencione como principio separado, por que esta no se deba a una susceptibilidad anormal del observador; en segundo lugar, por las posibilidades razonables de evitar presenciar estas conductas ofensivas; y, por último, por la que llama ‘’Volenti maxim’’ (pp.32 y 35), que viene a indicar que las conductas toleradas o consentidas por el sujeto pasivo nunca podrán ser consideradas como auténticas ofensas.

Pese a que la valoración de la seriedad o gravedad de la ofensa en atención a la magnitud de su desaprobación social, la duración de esta y a que esta gravedad no pueda residir en absoluto en la especial sensibilidad de un individuo particular, sino que debe residir sobre cierto estándar objetivo normativamente establecido desde una perspectiva social, nos parece plenamente oportuno, cabe plantearse dudas sobre la relevancia de las variables del grado de evitabilidad de la ofensa y del consentimiento del afectado. Si la idea que se ha defendido es la de prescindir de toda consideración subjetivista basada en los sentimientos individuales y optar, en su lugar, por una perspectiva normativa tendente a la objetivación de lo ofensivo, el hecho de que el receptor pueda evitar recibir estos tipos de mensajes sin especiales costes o que se haya puesto voluntariamente en una situación en que asume que puede sufrir este tipo de conductas es una cuestión que puede presentar una importancia moderada a la hora de valorar la ofensividad objetiva de la conducta. Es cierto que el hecho de que cierto tipo de conductas se den típicamente en un entorno determinado y conocido por todos, lo que hará que aquel que se introduzca en tal

contexto sepa que es posible que escuche o vea ciertos hechos que pueden serle molestos así como que pueda evitarlos de forma más o menos sencilla, puede tener cierto valor a la hora de especificar el desvalor social que se puede atribuir en general a estas conductas, atribuyendo menor gravedad a aquellas conductas generalmente realizadas de forma muy localizada en un contexto en que los intervinientes generalmente lo consideran admisible, pero, pese a ello, este no puede ser un criterio determinante. La gravedad de lo ofensivo que resultaría, por ejemplo, el que un grupo de ideología nacionalsocialista quemara cruces, ahorcara muñecos que representan a personas negras y griten eslóganes racistas que aludan a las personas de distintas razas de una forma especialmente despectiva y denigrante o que ensalcen la realización de conductas racistas y discriminatorias contra ellas, no puede quedar desmerecida por el hecho de que se hiciera en un local donde este grupo se suele reunir o en un barrio donde su presencia es notable y al que una persona ajena a esta organización ha accedido libremente y puede, en todo caso, abandonar sin ningún coste extraordinario. Por ello, aunque reconozcamos que pueden ser criterios a tener en cuenta, creemos que los elementos esenciales que componen la seriedad de la afrenta son los de la gravedad socialmente determinable de la ofensa en sí misma y, en su caso, de la duración en el tiempo que esta presenta.

El segundo de los bloques de principios que propone Feinberg (1985), a nuestro parecer, es el que de una forma más plena se enfoca el núcleo del problema de la valoración de la ofensa como *wrongdoing* desde una perspectiva normativo-social, pues con él el autor alude a una serie de criterios que permitirán esbozar, precisamente, la razonabilidad de la conducta ofensiva. Así, ‘’the careful legislator will proceed to balane that seriousness against the reasonableness of the various kinds of conduct that can produce it’’ (Feinberg, 1985, p.37), estructurándose esta razonabilidad de la ofensa, para el autor, en torno a seis elementos principales, a saber, la importancia personal de la conducta para el autor, el valor social que se reconoce a la posibilidad de su realización, la protección por la libertad de expresión —que como él reconoce no es un principio autónomo sino un corolario de los dos primeros mencionados—, la posibilidad de realizar esta acción de un modo alternativo no ofensivo pero igualmente valioso para la sociedad y el autor, la naturaleza del lugar en que se ha producido —como corolario específico de la existencia o no de alternativas de actuación enfocado en la adecuación de la conducta al contexto específico—y la malicia con la que se realiza la acción ofensiva.

En este caso sí nos parecen relevantes la existencia de vías alternativas abiertas al actor para valorar lo ilícito moralmente de su actuación y el contexto específico donde se realiza, pues estos son elementos que se relacionan con la entidad normativamente objetivada de la ofensa en atención no a su gravedad intrínseca, sino a la adecuación o inadecuación social del comportamiento, la cual está atravesada por el contexto en que se envuelve y por las distintas posibilidades de vehicular el desarrollo de la persona y la autodeterminación de la voluntad en el marco de la vida en sociedad. Como indica Feinberg (1985) ‘’if the offending person, by doing this thing at another place or time, can avoid causing offense to a captive audience without los or unreasonable inconvinience to himself, then his offending conduct is unreasonable if done in circumstances that permit offense’’ (p.40).

Parecen, sin embargo, todavía más interesantes las cuestiones sobre la importancia personal del acto para aquel que lo realiza y el valor social que esta presenta. Respecto de la primera de ellas Feinberg (1985) entiende que ‘’if the conduct in question is part of the activity by which the actor earns his living so that its curtailment would harm his economic interest, then obviously it is important to him, whatever others may think of it’’ (p.37). Conforme a este razonamiento, entiende el autor que cuando se trate de acciones

que, pese a poder suponer una molestia para otros, son relevantes para mejorar la salud, bienestar, conocimiento o intereses económicos, no podría considerarse, en principio, como una conducta normativamente ofensiva. Sin embargo, lo que nos parece realmente determinante no es la relación individual de interés para con la acción que tenga el actor, sino el segundo de los elementos a tener en cuenta señalados por Feinberg (1985), a saber, el del valor o utilidad social de la conducta. En efecto, poca relevancia presentará que exista un auténtico interés del individuo en realizar una acción ofensiva si este interés es completamente denostado a nivel social y se considera completamente superfluo o, cuanto menos, que debe prevalecer ante los derechos de otras personas a ser tratados con respeto o no ser sometidos a determinadas experiencias. En definitiva, lo esencial para determinar si una conducta constituye una auténtica afrenta y falta de consideración para con los demás que pueda merecer reproche penal será el reconocimiento desde la estructura social y jurídica de la comunidad de tal importancia del acto ofensivo para el sujeto que lo realiza. Sólo cuando esta importancia para el individuo se alinee con los intereses generales de la sociedad podrá decirse que es una conducta adecuada socialmente y que debe ser plenamente permitida. Por consiguiente, lo esencial para determinar la legitimidad o ilegitimidad del castigo no es la relevancia de la conducta para el actor, sino la valoración social que se dé a la posibilidad de actuar de tal modo.

Esto se relaciona claramente con el criterio de ponderación que es el respeto a la libertad de expresión que introduce Feinberg (1985), que entiende que

> ‘’Expressing opinions openly in spontaneous conversation, writing, or through more powerful media of communication is also of great importance to private individuals themselves, since self-expression is valued both as an end in itself and as a means of effecting desired changes. Bur it is also a necessary condition for the satisfactory functioning of any government that relies heavily on enlightened public opinión in its decision making’’ (p.38).

En un régimen democrático como el nuestro no puede negarse que la garantía de la libre expresión es un elemento fundamental para construir aquello a lo que Oliver Wendell Holmes (Corte Suprema de los Estados Unidos, 10 de noviembre de 1919) se refirió como el libre mercado de las ideas, un espacio en que pueda desarrollarse hacia su plenitud el pluralismo político de una sociedad llamada a gobernarse a sí misma por medio decisiones tomadas en el marco del diálogo y la confrontación de opiniones e ideologías. Es por ello que concluye Feinberg (1985) que ‘’no amount of offensiveness in an expressed opinión can counterbalance the vital social value of allowing unfettered personal expresion’’ (p.39), pero teniendo en cuenta que lo que en ningún caso puede constituir una ofensa en sentido jurídico es la opinión o contenido ideológico de la misma, por la relevancia política que presenta garantizar la libertad para expresarla, pero sí puede resultar jurídicamente desaprobado, por ofensivo, ciertas formas especialmente despectivas, agresivas o denigrantes de manifestar tales contenidos ideológicos. En este sentido, ‘’a devout Christian might me oddended by the bare assertion of atheism; or the audience might be ofended instead by the manner in which the opinión itself is expressed, for example, as a caption to an obsecene poster of Jesus and Mary’’ (Feinberg, 1985, p.39). Sólo cuando se trate de expresiones ofensivas en su forma puede darse el castigo como un acto ofensivo, aunque en absoluto pueda decirse que esto, por sí mismo, determine que necesariamente este deba darse; pero nunca podrá ser legítimo el castigo de la expresión de la opinión que subyace en la forma. La inadecuación social de este tipo de expresiones es inadecuación social del medio y del modo, nunca del contenido sustantivo, pues ello supondría dinamitar las bases del pluralismo político propio de un

Estado social y democrático de Derecho, el cual aconseja, adicionalmente, una especial cautela incluso en el castigo de estas formas ofensivas de exteriorizar la opinión.

Pese a la relevancia de estos elementos anteriormente mencionados, las consideraciones sobre la motivación de la conducta, sobre la cual se concluiría que ‘’when the motive is merely malicious or spiteful it deserves no respect at all’’ (Feinberg, 1985, p.41), nos parecen prescindibles. Cuando se trata de valorar la razonabilidad de la ofensa, una vez constatado que existe un cierto interés público en la permisión de la conducta y que el interés particular por realizarla es desde esta perspectiva tutelable, así como que no existen vías alternativas de actuación igualmente satisfactorias para estos intereses colectivos-particulares; es irrelevante la relación emocional o afectiva que exista entre la conducta y el agente que la desarrolla. Si se encuentra asistido por un valor social que aconseja decir que la conducta es permisible, aunque pueda ser molesta, no parece razonable eliminar este carácter de adecuación social por el hecho de que a nivel interno este sujeto disfrute de la molestia que genera, en tanto no se dé un ejercicio ilegítimo de este derecho que le es reconocible y siempre que no se produzca una extralimitación en el mismo. Si precisamente buscamos la objetivación de la ofensa por la vía de su normativización para evitar la imprecisión inevitable que deriva de apoyarse en elementos puramente subjetivos, debemos permanecer indiferentes frente a la malicia o maldad con la que se realice el acto en tanto este entre dentro de la dinámica social normal en que se da y que se considera aceptable.

Sea como fuere, del contraste de la razonabilidad de la ofensa y su gravedad se conseguirá concluir si la conducta ofensiva puede entenderse como tal en un sentido jurídico estricto, lo cual sólo podrá darse cuando la seriedad y gravedad de la ofensa no se vea contrarrestada por la relevancia social de este tipo de acciones en un contexto social, político y económico determinado. Por el contrario, cuando pese a que pueda ser considerada una acción gravemente ofensiva se entienda que existen intereses sociales y políticos que exijan la permisión de este tipo de conducta y tengan un valor superpuesto al posible perjuicio individual que puedan causar, en ningún caso sería legítima la intervención penal. Creemos que buena parte del valor explicativo de esta contraposición de principios queda condensado en el juicio moral sobre si esta ofensa vulnera un derecho de otro, tanto en sentido jurídico como moral, bajo la máxima de que ‘’aquello que hiere la sensibilidad o los sentimientos de otro, pero que no quebranta ninguno de sus derechos no es una ofensa’’ (Miró-Llinares, 2015, p.57); sumado al de si existe un derecho del actor a realizar tal conducta por la existencia de motivos morales, jurídicos y políticos relevantes que determinan el valor de la posibilidad de realizar tales acciones.

Ahora bien, la determinación de cuándo existe un derecho, en sentido moral, a no ser ofendido puede ser compleja. Para facilitar esta labor podría tener cierto valor orientativo la noción limitada de bien jurídico-penal que hemos esbozado más arriba. En este sentido, si bien no cabrían dudas en los casos en que existan derechos subjetivos en sentido estricto a no ser ofendido de determinado modo, para resolver los casos complejos en que el Ordenamiento Jurídico positivo no especifique nada, podría tratarse de extraer derechos en sentido amplio a partir de estas estructuras, instituciones, estados de cosas o realidades que se consideran básicas en el orden axiológico específico de que se trate, lo cual, en nuestro caso, llevaría a un razonamiento basado en la dignidad de la persona y la búsqueda de garantizar su libre desarrollo en una sociedad inspirada en los valores de libertad, justicia, igualdad y pluralismo político (art. 1 CE). Cuando pueda funcionalizarse algún tipo de bien jurídico desde el individuo libre en sociedad cuyo contenido se identifique claramente con el derecho a no sufrir determinado tipo de ofensas podríamos hablar de que existe un derecho que sería lesionado por una conducta gravemente

ofensiva; lo cual concurre igualmente, de forma evidente, cuando estas conductas afecten directamente a derechos fundamentales previstos en nuestro texto constitucional, otros derechos meramente constitucionales o derechos subjetivos reconocidos en otras normas jurídicas no constitucionales. Por el lado contrario, cuando estos principios y estructuras esenciales del orden político amparen directamente la realización de ciertas acciones, aunque puedan resultar molestas u ofensivas para otros, deberá reconocerse la existencia de un gran valor social en la mera posibilidad de desarrollarlas, y por consiguiente se deberá tender a su no criminalización.

Esta conjunción del bien jurídico con ambos principios, el de daño y el de ofensa, consagraría la función crítica de la teoría del bien jurídico y permitiría su operativización a nivel sistemático, garantizando cierto estándar limitativo del poder punitivo del Estado en todas las vías por las que esta puede darse. Además, la vinculación de la ofensa con la dimensión axiológica básica y esencial de una comunidad política por la que podría entenderse el bien jurídico-penal en el sentido amplio indicado más arriba, creemos, podría asimilarse a la exigencia de Duff (2009, como se citó en Miró-Llinares, 2015) de que la ofensa —y también el daño— constituya, en todo caso, un injusto público entendido ‘’no como un injusto que lesiona al público o que se perpetra fuera del ámbito privado, sino como un actuar inmoral que concierne propiamente a lo público’’ (Miró-Llinares, 2015, p.39); o, en igual sentido, de Husak (2008), que remarca la necesidad de la repercusión pública de los actos castigados por el Derecho penal, e incluso del propio Jakobs (1991), que subraya el carácter público del conflicto que ha de tratar de solventar la norma penal por medio del establecimiento de un estándar de conducta que sirva como expectativa de comportamiento de los demás. Se daría la reconceptualización, de este modo, de los principios de daño y ofensa como dos especies de un mismo género, dos formas legítimas de ejercer el poder punitivo de forma legítimo siempre dentro de un marco común de exclusiva protección de bienes jurídicos.

Así entendido, el injusto debe presentar una determinada entidad pública para poder ser tal injusto jurídico-penalmente relevante. El desvalor de la acción debe recaer desde su examen inspirado por los principios fundamentales del orden político-social, y la contradicción o menoscabo de estos elementos estructurales del régimen político deberán constituir el núcleo esencial del desvalor de resultado. El núcleo común de la antijuricidad penal debe sustentarse, entonces, en cuestiones radicalmente constitucionales, de las que se derivará qué es un bien jurídico-penal y qué acciones son injustas como sustrato mínimo para permitir su reproche.

Sin embargo, este planteamiento suscita una cuestión interesante. Si la diferencia fundamental entre los principios de daño y ofensa según Feinberg (1984, 1985) es que, mientras que el daño supone un menoscabo injusto de un interés relevante que debe permanecer indemne, la ofensa constituye un *wrongdoing* en que no se produce un verdadero perjuicio a un interés relevante sino que presenta un significado social notablemente ofensivo y reprochable como tal; pero al tiempo tratamos de emplear el concepto de bien jurídico como elemento que permita en cierto sentido orientar la valoración sobre el carácter injusto o ilícito de la conducta ofensiva, resulta que debe trazarse una distinción clara y esencial entre el bien jurídico y el interés relevante, y no podemos asumir plenamente la práctica identificación entre principio de daño y exclusiva protección de bienes jurídicos que defiende Hirsch (2007). En caso de que se identificara directamente este bien jurídico con el interés digno de protección que queda lesionado por un daño en sentido estricto sería imposible el empleo del bien jurídico como criterio orientador de la valoración de la ofensa como *wrong*, pues ello llevaría a que se produjera

la identidad entre daño y ofensa, pues esta también sería un *wrong* que menoscaba un derecho deducido de un bien jurídico protegido identificado con el interés feinbergiano.

Por consiguiente, deberá entenderse que no todas las situaciones, realidades o estados que puedan considerarse como un bien jurídico en sentido crítico son representaciones o manifestaciones de un auténtico interés individual en el sentido que se le da por Feinberg (1984). Así una forma de conjugar ambos criterios de forma satisfactoria tanto para harm como para offense es desdoblar el concepto de interés en dos, uno referido a que este constituya un interés legítimo de una persona con una importancia que permita su tutela legal por el Estado de algún modo y otro carácter adicional que pueden tener algunos intereses, a saber, que se puedan derivar de los pilares básicos y elementales que configuran la sociedad. Sólo cuando un interés sea además de tutelable, un elemento cuya protección se pueda identificar con la protección de la razón de ser de la sociedad en que existe podrá ser tutelado penalmente. Todo *harm* en sentido estricto deberá ser un *wrong* contra un interés relevante, pero sólo el wrong a un interés tan relevante como para considerarse un bien jurídico-penal y su protección identificarse con la protección de la sociedad podrá ser tutelado por vía punitiva. La distinción entre intereses legítimos tutelados penalmente y los no tutelados parece intuirse en el propio Feinberg (1984) cuando distingue entre aquellos que se relacionan directamente con el bienestar de la persona y los *ulterior interests*. Por otro lado, cuando no exista interés relevante alguno, pero exista una conducta ilícita que suponga objetivamente una ofensa para otros por producir un desajuste del modo normal de funcionamiento de la comunidad conforme al estándar axiológico existente en la misma diremos que existe una ofensa en sentido estricto.

En definitiva, esta perspectiva exigiría una delimitación clara y precisa del concepto de interés relevante y su diferenciación respecto del bien jurídico. Pero es que, al eliminar la dimensión axiológica propia del bien jurídico en sentido crítico del concepto de interés relevante, este último queda vaciado de contenido y desdibujado, precisándose una mayor concreción y desarrollo, pues por sí sola la exigencia de que estos intereses relevantes sean aquellos cuyo mantenimiento indemne redunda en una mejora del bienestar de su titular no responde a la pregunta de si una determinada realidad o valor considerado esencial para una comunidad es un interés relevante, susceptible de ser dañado o, por el contrario, sólo podría ser ofendido. En este sentido, queda fuera de toda duda que el honor constituye un bien jurídico en sentido crítico y su protección es necesaria para el mantenimiento del oren político-constitucional básico. Pero entonces, ¿constituye el honor un auténtico interés de la persona? En caso afirmativo diríamos que las conductas que lo lesionan son daños, y por tanto permitiría una respuesta punitiva que podría alcanzar las penas de prisión. En caso contrario diríamos que la vulneración del honor es una conducta ofensiva, lo cual permitiría cierto tipo de respuesta penal pero sustancialmente debilitada.

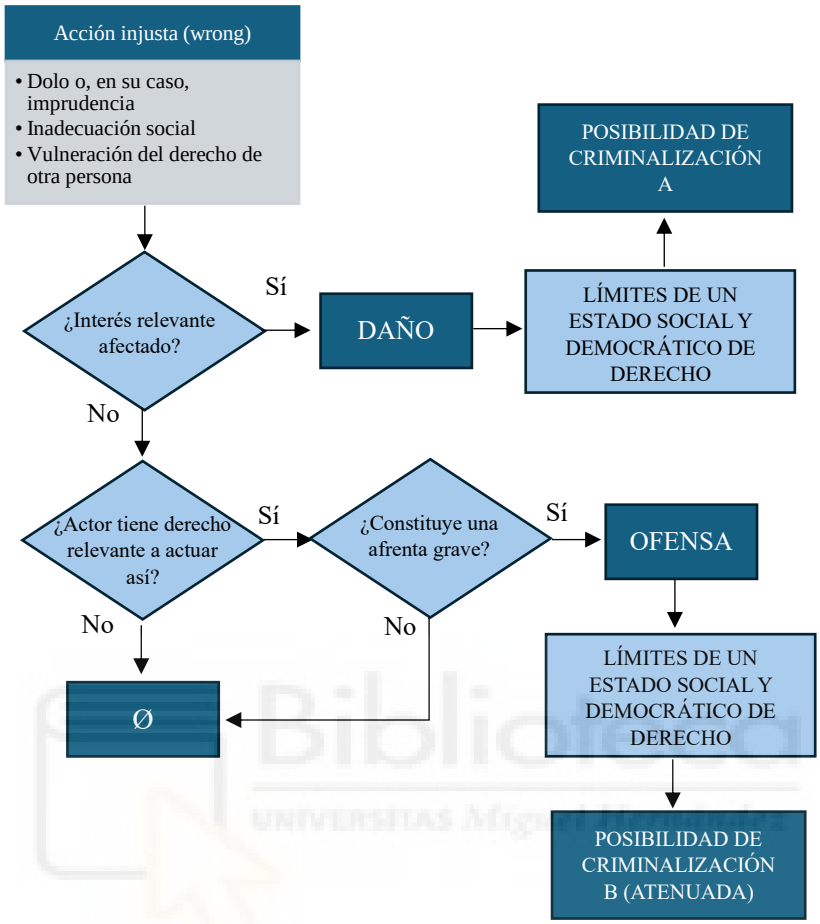
No dejamos de reconocer lo problemático de esta propuesta y necesitará una mayor exploración, matización y complementación. Un problema claro que suscitaría es que podría llevar a la paradoja de que se permitan sanciones administrativas o civiles frente a conductas lesivas de intereses tutelables que no se corresponden a un bien jurídico-penal pero, en cambio, en las ofensas, al carecer de interés tutelable, no permitiría la sanción administrativa pero, en cambio, sí la intervención penal en tanto pueda identificarse de algún modo un bien jurídico-penal. Esto parece contradecir claramente el principio de necesidad y ultima ratio del derecho penal, pues permite el castigo más grave frente a las acciones menos graves, cerrando la puerta a una intervención legal menos intrusiva para con la libertad, mientras que restringe la posibilidad de intervención penal

frente a acciones que lesionen auténticos intereses relevantes cuando estos no lleguen a tener el estatus de bien jurídico-penal.

Esta cuestión, sin duda, debe explorarse para tratar de solventar el problema desde la premisa de que debe priorizarse una reaccionar ante estas conductas por medios no punitivos. Una posible vía sería el entendimiento de que el interés tutelable que se lesiona con el *harm principle* determine la necesidad de ofrecer algún tipo de reacción social institucionalizada —y esta sólo puede ser penal en tanto este interés se pueda reconducir al paraguas de un bien jurídico—mientras que, en el *offense principle,* su propia naturaleza normativa permite el empleo de otros medios menos graves con sustento en argumentos o criterios adicionales, pero no determina la necesidad de dar tal respuesta colectiva, estableciendo que cuando se considere necesaria tal reacción sólo podrá ser penal en tanto exista un bien jurídico. Por ejemplo, una acción ofensiva podrá ser reprochada legalmente si hay buenas razones sociales para ello, pero sólo podrá darse la reacción penal en tanto contradiga directamente los principios y pilares rectores de un Estado social y democrático de Derecho, como podría ser en nuestro caso, la profunda contradicción de los valores superiores del ordenamiento jurídico del art. 1 CE y/o incidan de forma negativa en algún tipo de esfera tutelada como derecho fundamental, como la igualdad (art. 14 CE) o el honor (art. 18 CE).

En cualquier caso, lo que parece más sencillo es, por el momento, optar por la vía seguida por Miró-Llinares (2017), prescindiendo del intento de integrar en el sistema el concepto de bien jurídico crítico en aras de la practicidad y sencillez, remitiendo en el ámbito del concepto del daño al entendimiento del interés como aquellos relacionados directamente con la autonomía personal del individuo que permiten ‘’el desarrollo de la personalidad en dignidad’’ (p.29) y el daño como ‘’una real afectación a una esfera de libertad esencial del individuo’’ (p.29). De este modo, los dos ejes principales del modelo de criminalización que manejaríamos serían los del daño, como acción injusta por la ausencia de justificación moral basada en la adecuación social y que vulnera el derecho de otro —*wrong*— y produce un menoscabo o frustración de un interés relevante relacionado con la autonomía, libertad y dignidad de este otro; y la ofensa, como acción igualmente injusta o ilícita que, pese a ser objetivamente inaceptable socialmente, no constituye un auténtico ataque a un interés individual. Y a estos principales ejes, como se ha visto, deberán añadirse otros principios limitadores del *ius puniendi*, derivados de su fundamento funcional y de los valores propios de un Estado social y democrático de Derecho ya señalados, como el principio de utilidad o el de *ultima ratio*.

FIGURA 4: Esquema básico del modelo de criminalización propuesto



Fuente: elaboración propia a partir de Feinberg (1984, 1985) y Miró-Llinares (2015)

Sea como fuere, y pese a que sea necesario continuar investigando esta cuestión para tratar de perfilar más claramente qué es un interés relevante, este desdoblamiento de la intervención penal ante daños y ofensas conlleva, aun prescindiendo del intento de conjugarlos con el concepto crítico de bien jurídico-penal, dos consecuencias primordiales. La primera de ellas, la más evidente, es que toda propuesta de introducción de nuevos delitos debe pasar la comprobación de que aquella conducta que se pretende penalizar responda al concepto de daño o de ofensa. En segundo lugar, que aquellos bienes jurídicos en sentido dogmático, esto es, previstos como objeto lesionado por el delito ya tipificado; ocultan implícitamente uno de los dos tipos de acción que pueden ser legítimamente castigados. Es decir, los distintos tipos penales vigentes responden a acciones que pueden considerarse bien como daños, bien como ofensas; y debe tratarse de especificar qué naturaleza debe presentar la acción castigada por cada uno de ellos para no producir extralimitaciones en la imposición de castigos, para lo cual volvemos a caer en el problema de ‘’llegar al acuerdo de qué es aquello que daña un interés digno de protección o cuál de estos lo sería’’ (Miró-Llinares, 2017, p.29). Parece claro que algunas de las figuras delictivas existentes en nuestro país, como los delitos contra la vida humana independiente (arts. 138 y ss. CP), las lesiones (arts. 147 y ss. CP), los delitos contra la libertad (arts. 163 y ss. CP) o la libertad e indemnidad sexual (arts.178 y ss. CP), entre

otros, son manifestaciones de delitos que se producen por acciones que constituyen auténticos daños; y que, por otro lado, el discurso de odio (art. 510 CP) o los delitos contra los sentimientos religiosos y el respeto a los difuntos (arts. 523 y ss. CP) son manifestaciones de conductas gravemente ofensivas —pese a que algunos de ellos prevean penas de prisión—, pues aunque puedan suponer conductas graves que hieren sensibilidades sociales relevantes, lo que no puede pretenderse es decir que afectan a una esfera esencial de libertad del individuo que las sufre. Sin embargo, de nuevo caben dudas sobre si los delitos contra el honor (arts. 205 y ss. CP), especialmente las injurias, constituyen delitos que se cometen por medio de daños o de ofensas. Sobre esto se volverá más adelante.

En todo caso, ante estas conclusiones, debe entenderse que, ante la aparición de nuevas realidades, la intervención del Derecho penal debe partir del desentrañamiento de la naturaleza de estas nuevas conductas para determinar si pueden reconducirse al concepto de daño o de ofensa, y esto tanto para la creación de nuevos delitos como para el castigo desde los delitos preexistentes. A pesar de las incógnitas que quedan por resolver para ello, aquí sólo hemos tratado de ofrecer una aproximación teórica a un modelo de criminalización que creemos puede ofrecer algún tipo de utilidad para analizar la cuestión de los *deepfakes* desde este punto de vista para lograr las máximas garantías penales. Y, aunque sea cierto que esto se ha logrado no sin un notable grado de abstracción y poca concreción sobre los contenidos sustantivos de los diversos principios incluidos; creemos que también lo es que pueden tomarse las ideas de daño —y su constituyente esencial distintivo, el interés lesionado— y ofensa, aun en sus sentidos poco concretos con los que tenemos que resignarnos a trabajar, como *idealtypus* weberianos que nos permitan un examen de las conductas a través de sus grados de semejanza para con uno u otro principio de criminalización. Nos habilitaría así, al menos, a realizar un análisis provisional sobre, por ejemplo, si los *deepfakes* sexuales se asemejan más al concepto de daño o el de ofensa, legitimando así su castigo desde unos requisitos específicos y a través de un marco punitivo determinado.

IV. BREVE VIAJE EN EL AUTOBÚS DE LOS DEEPFAKES DESDE EL MODELO DE CRMIMINALIZACIÓN ESBOZADO

Es célebre el ejercicio mental que propuso Feinberg (1985) para analizar los distintos grados de seriedad de las ofensas. Consistía en que el lector se colocase a sí mismo en un autobús que le conducía a una cita importante y en el que el autor plantea un amplio abanico de conductas realizadas por las demás personas presentes en dicho vehículo, yendo desde algunas simplemente molestas a otras profundamente repulsivas. Nosotros hemos tratado de exponer la variedad de formas en que se puede manifestar el fenómeno de los *deepfakes* en la introducción del presente trabajo y en el epígrafe dedicado a la aproximación conceptual a las ultrafalsificaciones. A través de esas páginas se pudo ver que los distintos usos problemáticos de la tecnología de la IA generativa han dado lugar a una casuística de muy diversa índole, ante la cual se suscita la pregunta sobre qué papel debe jugar el Derecho penal en su afrontamiento y cuáles serían los límites de esta posible intervención penal.

Con la confección de los rasgos principales de un modelo de criminalización que hemos tratado de llevar a cabo en las páginas precedentes creemos que hemos dejado sentadas unas bases más o menos sólidas para perfilar los límites generales del uso del *ius puniendi*. Lo que nos resta a continuación será, pues, tratar de realizar una breve

prospección en el terreno práctico desde dichas bases teóricas para buscar algún tipo de orientación sobre la forma en que debe darse, en su caso, la respuesta penal ante los *deepfakes*.

Aunque los deepfakes puedan ser plasmados en soporte físico y desplegar sus efectos directamente en el espacio analógico tradicional como podría hacer cualquier otro tipo de falsificación, la forma en que de forma preeminente aparecen es en el ciberespacio, pues su difusión se realiza predominantemente en el nuevo medio digitalizado basado, primordialmente, en las redes sociales como espacio de interacción social virtual. Por ello, puede ser buena idea tratar de ordenar su análisis desde su concepción como ciberdelitos y su categorización, siguiendo la clasificación de cibercrímenes elaborada por Miró-Llinares (2012), en *deepfakes* económicos, políticos y sociales. Sin embargo, este trabajo, ya demasiado extenso, no puede ocuparse de analizar pormenorizadamente todos estos tipos de ultrafalsificación. Por ello nos limitaremos a realizar una serie de sucintas reflexiones sobre las diferentes clases de *deepfake*, tratando de señalar los principales problemas que plantea cada una de ellas y las implicaciones que presenta su análisis desde el modelo de criminalización esbozado, pero prestando especial atención a las ultrafalsificaciones de carácter sexual, que en la clasificación aquí seguida caería en el ámbito de los *deepfakes* sociales, por parecer ser las que más sensibilidades ha suscitado.

En el ámbito de los *deepfakes* económicos no parece que se planteen especiales complicaciones para la articulación de la respuesta penal en atención a los criterios de legitimación de intervención del Derecho penal. En este campo, se ha visto que las ultrafalsificaciones han servido, principalmente, como elementos instrumentales al servicio de los ciberfraudes. De tal forma pueden concebirse esta clase de *deepfakes* como herramientas mediales empleadas para la realización de ciberataques réplica que trasladan un ataque tradicionalmente realizado en el ámbito analógico al campo de las relaciones sociales en el ciberespacio (Miró-Llinares, 2012). Y es que la cuestión del criterio de criminalización opera precisamente respecto del daño que en sí mismo produce el perjuicio patrimonial, y no en el medio de que se dé uso para lograrlo. El *deepfake*, que en cualquier caso es el medio comunicativo por el que se logra el engaño de la víctima, no presenta un carácter de daño u ofensa en sí mismo, sino que el daño que habilita la intervención penal se encuentra precisamente en el perjuicio económico que por medio de él se produce.

El análisis jurídico penal, entonces, deberá enfocarse en la valoración del grado de realismo del *deepfake* y la capacidad de considerar este desde una perspectiva socialmente objetiva como idóneo para generar este engaño bastante, cuestión que entronca con la propia conceptualización de las ultrafalsificaciones, como más arriba se tuvo la posibilidad de indicar. Por consiguiente, no parece especialmente problemático desde el punto de vista de la legitimidad del ejercicio del *ius puniendi* reconducir las conductas en que se emplee un *deepfake* objetivamente realista y capaz de producir en otros un error sobre la realidad para llevarlos a realizar un acto que les perjudique patrimonialmente en beneficio del actor al ámbito conceptual del delito de estafa previsto en los arts. 248 y ss. CP.

Por la parte de los *deepfakes* políticos, pueden presentar cierta trascendencia en el ámbito del discurso del odio, pudiéndose encarnar en imágenes falsas que presenten a personas de determinados colectivos sociales, económicos o raciales de forma estereotipada o realizando determinados actos deleznables que podrían ahondar en la discriminación y la perpetuación de los prejuicios contra estos grupos de personas. Desde

esta perspectiva cobrará especial trascendencia el papel del *ofence principle* para determinar la gravedad de la conducta desde un paradigma normativo y de adecuación social para, en estos términos, poder valorar si la acción presenta la relevancia jurídico-penal necesaria para poder defender su castigo desde figuras delictivas como la prevista en el art. 510 CP.

Por otro lado, pueden presentar una radical potencialidad a la de profundizar en el fenómeno de la desinformación y pudiendo alterar el debate político e influir en los procesos electorales. En este sentido podría darse su uso como herramienta de ciberterrorismo, difundiendo mensajes tendentes a la radicalización; o bien como un medio para la intromisión en el buen funcionamiento de las instituciones y el procedimiento democrático. Para estos casos, de forma similar a lo indicado respecto de los *deepfakes* económicos, las ultrafalsificaciones no constituirían un daño u ofensa en sí mismo, sino que se erigirían como instrumento idóneo para la causación de resultados que sí pueden constituir un daño, u ofensa, si se quiere. Sin embargo, su posible incidencia en el campo de la desinformación, aunque sea una cuestión que ciertamente puede presentar cierto carácter transversal y no exclusivamente político, sí suscitaría dudas no sólo sobre si existen verdaderos intereses individuales involucrados que pueden quedar afectados por conductas dañosas o si, por el contrario, se circunscribiría al espacio de la ofensa; sino también en torno al debate sobre la intervención pública en la definición de la verdad y la limitación de la libertad de expresión. Lo que está claro es que si este tipo de tecnología alcanza un perfeccionamiento que llegase a ser indetectable supondría una gran amenaza para la seguridad cognitiva sobre la realidad de lo que se ve y se escucha. No podría alcanzarse grado de certeza alguno acerca de qué es real, y ello conllevaría importantes consecuencias para con la confianza en las instituciones, el debate político y el buen funcionamiento de los procesos democráticos; pero también podría suponer un perjuicio directo a los individuos, que se verían desprovistos de cualquier tipo de seguridad a la hora de desplegar su vida libre en sociedad, sometiéndolos a un estado de incertidumbre y desorientación completa y perpetua.

Esta plenitud de la posverdad podría determinar la necesidad de concebir esta capacidad de discernimiento de la realidad como un auténtico interés individual susceptible de ser lesionado por conductas verdaderamente dañosas frente a las cuales el Derecho penal estaría legitimado a intervenir, si bien es cierto que debiendo proceder de forma extremadamente cautelosa por el grado de intromisión y regresión en las libertades individuales que podría suponer un mecanismo de definición y preservación de una suerte de verdad oficial. Esta cuestión del grado de afectación a las capacidades epistémicas del ser humano que podrían desencadenar las ultrafalsificaciones está siendo objeto de un interesante debate, del que se pueden destacar, entre otras, las aportaciones de Rini (2020), Raymond (2021), Cohen y Rini (2022), o de Habgood-Coote (2023), de carácter general —aunque algunas de sus reflexiones se refieran específicamente al asunto de los *deepfakes* sexuales— o la de Twomey et al. (2023), con una aplicación de este problema epistémico al campo político, concretamente en lo relativo a las ultrafalsificaciones en relación con la invasión de Ucrania por parte de Rusia. Sin duda, al rol del Derecho penal ante el, correctamente o no, llamada ‘’epistemic apocalypse’’ (Habgood-Coote, 2023, p.1) y su articulación desde un modelo de criminalización determinado deberán dedicarse ulteriores investigaciones.

El nuevo —o ya no tan nuevo—ciberespacio en el que a diario nos relacionamos es un medio en que se dan todo tipo de interrelaciones y comunicaciones sociales, lo que ha permitido que ciertas formas de comportamiento socialmente inaceptable migren a este nuevo entorno. Así, ‘’amenazas, coacciones, injurias, calumnias y otras agresiones al

honor o a la libertad pueden realizarse como siempre, pero a través del ciberespacio'' (Miró-Llinares, 2012). De este modo aparecen los *deepfakes* sociales como una forma de intercomunicación social virtual, que tiene lugar en el ciberespacio y que puede presentar, de igual modo, la significación penal que otras conductas realizadas en el entorno social analógico de carácter no estrictamente económico o político. Esto hace que no se erijan en meros medios para la comisión de una conducta que, separada del instrumento empleado, posee una significación jurídico-penalmente relevante, sino que esta acción de generación y difusión de la ultrafalsificación debe presentar en sí misma la entidad necesaria para que pueda considerarse como relevante penalmente. Así, los *deepfakes* pueden aparecer como manifestaciones del concepto general de ciberacoso ''entendido como una macrocategoría englobadora de todas las conductas en las que se aprovecha el uso de distintos instrumentos de comunicación […] para realizar el atentado contra la libertad de otra persona'' (Miró-Llinares, 2012, p.84). Sin embargo, como ya se ha adelantado, nuestro interés en el campo de las ultrafalsificaciones que podrían englobarse en la ciberdelincuencia social lo recibe, principalmente las ultrafalsificaciones de carácter sexual.

Respecto de aquellas por las que se producen representaciones sexuales de menores de edad no se suscitan especiales problemas teóricos para reconducir la conducta a la figura delictiva de pornografía infantil, máxime si se tiene en cuenta que nuestro legislador ha concebido tal clase de pornografía, entre otros, como ''todo material que represente de forma visual a una persona que parezca ser un menor participando en una conducta sexualmente explícita, real o simulada, o cualquier representación de los órganos sexuales de una persona que parezca ser un menor, con fines principalmente sexuales, salvo que la persona que parezca ser un menor resulte tener en realidad dieciocho años o más en el momento de obtenerse las imágenes'' (art. 189.1.c) CP) o bien como ''imágenes realistas de un menor participando en una conducta sexualmente explícita o imágenes realistas de los órganos sexuales de un menor, con fines principalmente sexuales'' (art. 189.1.d) CP). Siempre que el grado de realismo sea suficiente, no suscita grandes problemas tratar de calificar este tipo de ultrafalsificaciones como pornografía infantil conforme a su definición legal. El problema esencial radica en la proporcionalidad o no, en atención a la consideración de si este tipo de representaciones realistas pero ficticias constituyen un daño o una ofensa, de las penas previstas, que ya son, en su modalidad típica básica, penas de prisión de 1 a 5 años (art. 189.1 CP), un marco punitivo excesivo en caso de que se concluyere que se tratan de acciones que no generan auténticos daños sino, tan solo, ofensas.

Dejando de lado las ultrafalsificaciones que tienen por objeto a menores de edad, la cuestión no es menos controvertida, pues debe tratarse de discernir si este tipo de conductas constituye un daño o una ofensa y si lo es respecto del bien jurídico en sentido dogmático de la intimidad o del honor. A su vez, debe responderse a si esta intimidad u honor, depende de por cuál se opte, constituirían intereses individuales susceptibles de sufrir auténticos daños. De tal modo, sólo podrá castigarse desde el delito preexistente si se da un solapamiento entre la naturaleza ofensiva o dañosa de la conducta y el carácter, por así decirlo, «ofendible» o dañable del bien jurídico dogmático desde el que se pretenda responder a esta realidad. En caso de que no se dé tal alineamiento de ambos elementos podrá crearse un nuevo delito, pero siempre teniendo en cuenta el marco penológico permisible según se concluya si este tipo de acciones constituyen daños u ofensas.

Personalmente, no creemos que deba enfocarse la cuestión desde la consideración de que el bien jurídico que se afecta sea el de la intimidad o incluso el de la libertad sexual,

procediendo a su reconceptualización, tal y como proponen, respectivamente, Simó (2023) o Rodrigues (2023). Creemos, en cambio, que la cuestión se alinea con la idea básica del honor en un sentido amplio, en tanto produce una representación falsa de otra persona dando la imagen frente a terceros de una conducta o apariencia que no es real y que puede menoscabar la reputación y apreciación social de la persona que la sufre. Ahora bien, si este honor es un interés individual o no queda sin responder y es necesaria una más profunda reflexión, pues de ello dependerá la posibilidad de castigar estas acciones como daños u ofensas, lo cual es esencial para no causar una extralimitación punitiva. Por ello, de forma provisional puede ser razonable defender que sólo se responda desde el marco punitivo común a ambas ramas del modelo de criminalización: las penas no privativas de libertad, lo cual sería compatible con ambos principios de criminalización, es decir, con las penas permisibles tanto respecto de las conductas dañinas como las ofensivas.

Parece que, hasta cierto punto, esta perspectiva ha sido la que se ha dado por el partido político Sumar, que realizó una propuesta para la modificación del Código Penal genialmente analizada por Santisteban (2024), y en la cual se pretendía incluir de forma expresa la producción de *deepfakes* con fines difamatorios en la definición típica de las injurias, con las penas ya previstas para este delito en el actual Código Penal, que son en todo caso de multa (art. 209 CP). Si bien este modo de proceder parecería coherente con lo expuesto, no deja de ser cierto, como apunta Santisteban (2024) que el delito de injuria no prevé un medio de comisión específico, de tal suerte que la inclusión expresa de las ultrafalsificaciones en su definición típica no deja de ser ‘’una aclaración bienvenida, pero prescindible’’ (p.17). Sin embargo, ha señalado también el propio Santisteban (2024) que esta propuesta, por este marco penológico, sería contraria a la propuesta de Directiva realizada por el Parlamento Europeo y el Consejo en materia de criminalización de *deepfakes*, que, a nuestro juicio de forma injustificada, exigiría para este tipo de ultrafalsificaciones penas de prisión de al menos un año.

No puede tampoco soslayarse el otro de los elementos de análisis necesario, y es si estas ultrafalsificaciones entrañan en sí mismas la entidad propia de un *harm* o de una *ofense*, la cual deberá estar alineada en su caso con el tipo existente desde el que se quiera castigar o influir de forma determinante en la respuesta penológica que se pueda establecer en un delito creado *ad hoc.*

Al margen del hipotético daño epistémico que con carácter general algunos han tratado de atribuir a las ultrafalsificaciones, como ya se ha mencionado, resulta aquí especialmente relevante la reflexión realizada por Cohen y Rini (2022) sobre cómo los *deepfakes* sexuales, a su parecer, producen *harms*. En sus páginas se pone de manifiesto cómo ciertas formas de pornografía constituyen actos de cosificación del cuerpo de la mujer, reduciéndola a un mero objeto de deseo sexual que debe ser puesto al servicio de la voluntad del hombre, entendiendo que, en este ámbito concreto, ‘’there is something painfully literal in the sort of objectification at work in deepfake porn’’ (Cohen y Rini, 2022, p.146) y, de tal modo, ‘’deepfaked frankenporn, then, is virtual domination, an extreme expression of sexual objectification aimed against specific women […] Frankenporn turns real people into digital toys. Even those unpersuaded by feminist objections to traditional pornography ought to recognize the moral wrong here’’ (Cohen y Rini, 2022, p.147).

A nuestro parecer, no cabe duda acerca de lo acertado de esta reflexión. Sin embargo, como no dudan en indicar, lo que realizan mediante ella es justificar, creemos que debidamente y con suficiente solvencia, que este tipo de conducta constituye un

wrong, una acción socialmente —o, si se prefiere, moralmente—inaceptable, pues no puede haber exigencia más básica de un orden social basado en el respeto de la dignidad humana el que las personas sean tratadas en tanto personas, y no como meros objetos. Pero el *wrongdoing*, como se ha tenido oportunidad de ver, es elemento común tanto de la conducta que genera daños como de la meramente ofensiva.

Se precisa, entonces, de forma adicional, dilucidar si este injusto constituye un menoscabo de un interés individual. La cosificación de la persona, no cabe duda alguna, supone la negación del trato digno de que toda persona es acreedora en tanto tal. Pero, de nuevo, esta cuestión del trato respetuoso y digno de una persona entronca con el respeto a su honor, pues la naturaleza de la injuria es la de ser ‘’una acción o expresión que lesionan la dignidad de otra persona, menoscabando su fama o atentando contra su propia estimación’’ (art. 208 CP), sin perjuicio de que en casos especialmente graves o de reiteración este tipo de actuaciones pueda comenzar a introducirse en el campo de la degradación mediante actos lesivos de la integridad moral del sujeto pasivo, si bien tampoco puede decirse que el concepto de integridad moral, habida cuenta de la diversidad de supuestos de hecho que se aglutinan en el art. 173 CP, nos ofrezca especial contenido explicativo de la naturaleza del acto perpetrado. En cualquier caso, esto nos lleva, de nuevo, al problema de determinar cuál es el bien jurídico dogmático que quedaría afectado —tanto si este es preexistente o de nueva creación— como paso previo para afirmar el carácter de daño u ofensa de la conducta analizada, y a la necesidad de contestar a la pregunta de si este bien jurídico es manifestación de un auténtico interés o no. Lamentablemente, como ya se ha dicho, no tenemos tal respuesta.

Estas incógnitas deben ser, desde luego, respondidas si se quiere garantizar la existencia de un Derecho penal legítimo digno de todo Estado social y democrático de Derecho. Pero tampoco puede pretenderse que el legislador permanezca inactivo hasta su respuesta, cuya consecución es harto compleja. Pero la simple constatación de que existen dudas y que se requiere una más profunda reflexión en torno al honor o, en su caso, a la integridad moral como interés y, por tanto, su lesión como daño, ya constituye una duda razonable que exigirá la extrema cautela a la hora de criminalizar estos actos. Ello nos lleva a insistir en la solución provisional ya señalada. En estos casos de duda será en todo caso mejor optar por un marco penológico atenuado, que trate de evitar en la medida de lo posible las penas privativas de libertad. Si se imponen penas de prisión se corre el riesgo de castigar como daño algo que no estamos seguros de que constituya verdaderamente un daño, y dado que el daño puede castigarse con penas de prisión, pero permite también castigos más leves; pero la ofensa no permitirá nunca la prisión, se revela más racional imponer penas no privativas de libertad, compatibles con ambas esferas de intervención penal legítima.

Debe tenerse en cuenta, además, que la constatación de que existe un injusto grave sólo permite legitimar el castigo, de una u otra forma en función de si existe un auténtico daño o una mera ofensa, pero no determina la necesidad de establecerlo. Por ello, deben tenerse en consideración argumentos adicionales sobre la oportunidad y deseabilidad de tal respuesta penal, así como, por supuesto, la influencia de ulteriores criterios y principios. En este sentido es especialmente trascendente, a la hora de tomar la decisión sobre la creación de nuevos delitos, el principio de utilidad de la pena. Y es que no pueden perderse de vista las aportaciones empíricas realizadas por autores como Gómez-Bellvís y Castro (2022) —sin desconocer que se refieren al campo de la comunicación de contenidos políticos—, que indican que el castigo de ciertas formas de comunicación, y la creación de *deepfakes* no puede entenderse sino como un acto comunicativo, puede resultar contraproducente al poder causar un efecto desafío que lleve, precisamente a

aquellas personas a las que se destina especialmente la conminación penal, a realizar estas conductas como un acto de rebeldía contra un sistema social que sienten trata de desplazarles; mientras que serían más proclives a producir en aquellos sujetos que sienten cierta afección hacia los valores que motivan la criminalización de estas conductas la caída en el efecto desaliento y la autocensura. Cabe entonces ser, si cabe, todavía más cautos a la hora de tomar la decisión sobre la criminalización de ciertas conductas relacionadas con las ultrafalsificaciones, por más deleznables que, con razón, puedan parecernos.

CONCLUSIONES

Estamos de acuerdo con Franks y Waldman (2019) cuando afirman que no hacer nada contra toda expresión dañina en la era digital no siempre tiene que significar alinearse con los principios de un Derecho penal liberal, sino que puede ser una opción normativa de mantenimiento del statu quo y la injusticia que, entienden, lleva aparejada. Puede ser cierto que el atrincheramiento en el marco y las convicciones existentes, en la doctrina penal y los delitos establecidos, pueda suponer en alguna ocasión un medio para perpetuar ciertas estructuras y dinámicas sociales que, quizá, fuere conveniente y de justicia subvertir. Pero, esta afirmación pasa necesariamente por una valoración que hacen las autoras de dichas expresiones en la era digital y que va más allá de su crítica: son expresiones verdaderamente dañinas.

Por ello, también creemos que el progreso de la sociedad y el enfrentamiento contra las conductas injustas no puede realizarse con el coste exorbitado de transformar nuestras comunidades en una suerte de ciudad de Omelas (Le Guin, 1973). Para evitarlo, la reacción frente a las nuevas conductas problemáticas que aparecen no puede darse desde la relativización, atenuación o ya ignorancia de los principios, límites y garantías penales que históricamente se concibieron como conquistas de derechos frente a un poder potencialmente tiránico. Si, precisamente, tratamos de ofrecer una respuesta penal frente a ciertos actos que atentan contra las bases de la estructura social y política en la que nos encontramos no podemos en modo alguno rendir esta empresa a la demolición de aquellas bases que pretendemos proteger. Por el contrario, la articulación de esta respuesta institucionalizada debe quedar siempre comprendida en los márgenes de aquellos límites que se imponen, como garantía, a esta posible respuesta; y ello se traduce en el deber de partir necesariamente de la definición de los criterios de criminalización que entendamos deben regir nuestro Estado social y democrático de Derecho para proceder a analizar si aquello que pretendemos es debidamente respetuoso con aquellos. Si, como afirma Agustina (2021), ‘’el delito es, ante el nuevo paradigma relacional que plantea el ciberespacio, sobre todo y más que nunca, comunicación’’ (p.726); debemos entender necesaria una especial profundización en el examen de la naturaleza de las conductas que pretendemos castigar y argumentar la existencia en ellas o no de los elementos que las hacen merecedoras de reproche penal, de la comunicación de significado que es la pena; y no circunscribirnos a aludir, como hace el propio autor, a ‘’la alarma social que genera este tipo de delitos’’ (Agustina, 2021, p.727). De lo contrario podríamos encontrarnos movilizando el poder punitivo contra acciones que simplemente presentan una mera apariencia delictiva o, en el mejor de los casos, reaccionando penalmente de forma desproporcionada para satisfacer las pulsiones punitivistas.

Creemos entonces que ha quedado patente a lo largo de este trabajo la necesidad de analizar la cuestión de los *deepfakes*, tanto para su castigo por medio de una figura

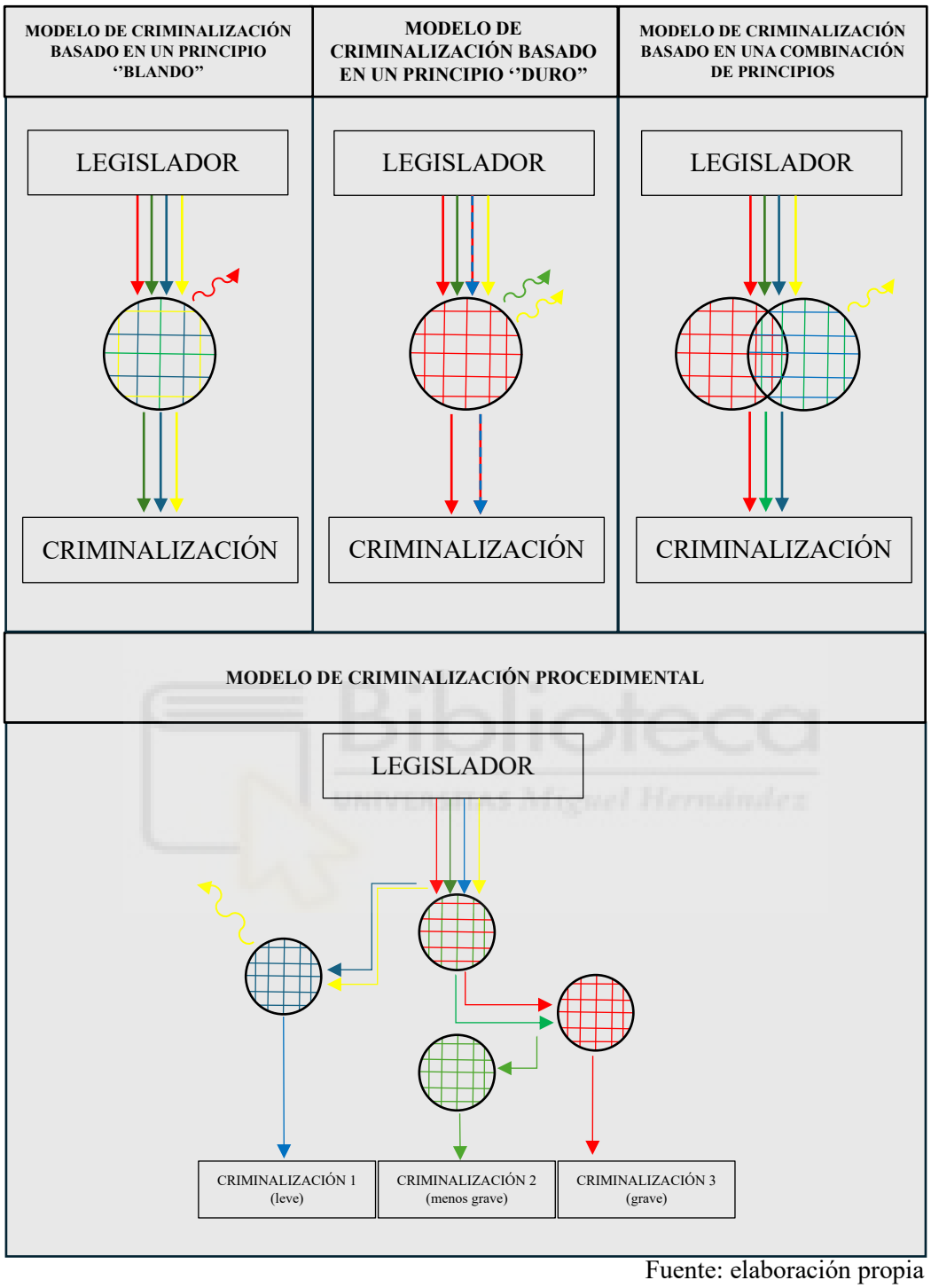
delictiva preexistente como para la creación de nuevos delitos, desde la perspectiva de un modelo de criminalización coherente con los principios fundacionales de nuestros Estados, así como que limitarnos a exponer lo indudablemente trágico de los supuestos de hecho que han tenido a lugar no basta para evitar tal necesidad.

Ahora bien, en nuestro intento de profundizar un análisis de este tipo hemos observado, en primer lugar, que la propia conceptualización de las mismas no es pacífica pese a que en la actualidad podamos encontrar los primeros intentos de definición legal de los *deepfakes* para construir su marco regulatorio. Sea como fuere, parece que podemos concluir, sin miedo a equivocarnos, que se entiende por *deepfake* todo tipo de contenidos audiovisuales generados por medio o con asistencia de sistemas de IA en que aparecen personas, objetos o lugares de una forma que no se corresponde con la realidad o en situaciones que no tuvieron lugar, pero que para poder considerar en un sentido jurídico estricto una ultrafalsificación como tal esta debe introducirse un criterio normativo basado en el grado de realismo que presentan, debiendo ser objetivamente idóneas para generar e quien las observa la percepción de realidad de tal contenido.

Por otro lado, a la hora de decidir qué tipo de modelo de criminalización debería emplearse, para poder dar cuenta de la variedad de nuevas conductas que se presentan en nuestro contexto al tiempo que presente cierto grado de concreción en los principios que lo integren y no tenga nula capacidad de filtrado de conductas legítimamente criminalizables frente a las que deben permanecer impunes; hemos concluido que quizá el medio más idóneo para nuestros fines optar por un modelo de criminalización de corte más o menos procedimental pero compuesto por una serie de criterios o principios llamados a tener algún tipo de contenido sustantivo determinado, de entre los cuales pueden ser especialmente útiles los principios de daño y ofensa manejados tradicionalmente en el contexto jurídico anglosajón y que aquí proponemos como agregación a los ya asentados en nuestra doctrina límites del Derecho penal de un Estado social y democrático de Derecho; y que desdoble la posible intervención penal legítima en dos niveles distintos, uno que dé la posibilidad de penas de prisión, y otro atenuado en que esta posibilidad quede negada.

Sin embargo, a la hora de dotar de tal contenido material a los distintos criterios, así como de aplicarlos a casos concretos de ultrafalsificaciones, la labor se ha revelado como un trabajo verdaderamente sisífico. No obstante, abrazamos aquí la opinión que para Camus (1942) merecen este tipo de proyectos y señalamos la necesidad de continuar en el avance de estas labores. A este respecto, hay que imaginarse al académico penal dichoso. Pero es que, pese a este notable grado de incertidumbre, que impone la obligación de seguir investigando y reflexionando, sí hemos podido sacar una conclusión clara: existen buenos motivos para abordar la toma decisiones jurídico-penales sobre el asunto con una extrema cautela, una cautela que aconseja tratar de reducir al máximo la intervención penal en favor de otras vías menos represivas y restringir las penas a imponer, en caso de que el Derecho penal sea el instrumento más eficaz para la prevención de las conductas más graves, al ámbito de las penas no privativas de libertad.

ANEXO I



Fuente: elaboración propia

En el presente Anexo se muestra de modo gráfico el funcionamiento de los distintos tipos de teorías de criminalización expuestas más arriba de modo que las propuestas legislativas de criminalización de distintas conductas (flechas) tratan de ''atravesar'' los distintos filtros que suponen los principios y criterios que integran las distintas teorías, representados como círculos rayados. Estos principios tan solo permiten la criminalización de aquellas conductas cuya naturaleza (color) concuerda con el contenido sustantivo (color) del principio con el que se examina.

Puede observarse así cómo la naturaleza abstracta y poco precisa de los modelos basados en un principio blando o flexible llevan a entender que su contenido sustantivo es respetado por conductas de distinta naturaleza, permitiendo criminalización de todas, en principio, desde un mismo marco penológico. Un problema similar plantean los modelos basados en la mera combinación simultánea de dos o más principios.

Por su parte, los modelos de criminalización basados en un principio sustantivo duro o rígido son más taxativos al estar más determinado su contenido sustantivo, por lo que tan solo permitiría el castigo de cierto tipo de conductas. Sin embargo, este sistema podría llevar a que el legislador trate de forzar la naturaleza y carácter sustantivo de la conducta que plantea criminalizar para adecuarla al principio que debe superar, lo que llevaría al castigo desde un mismo marco penológico de conductas de naturaleza diversa.

Por último, el modelo procedimental, pese a que quizá desde una perspectiva meramente cuantitativa permita el castigo del mismo número de conductas igual a una teoría basada en un único principio blando, lo hace diferenciando claramente el carácter y gravedad de las distintas clases de conducta, permitiendo su castigo y evitando la táctica legislativa de mutar el sentido de los principios adoptados pero limitando este castigo mediante una escala en el marco penológico tal que limite ciertos tipos de pena a determinadas categorías específicas. Por ejemplo, clasificando en función del tipo de conducta y sus particularidades en los tres tipos de delitos existentes en España en la actualidad: delitos leves, menos graves y graves.

BIBLIOGRAFÍA

Agustina, J.R. (2021) Nuevos retos dogmáticos ante la cibercriminalidad ¿Es necesaria una dogmática del ciberdelito ante un nuevo paradigma? *Estudios penales y criminológicos,* vol. 41, pp.705-*777*. https://doi.org/10.15304/epc.41.7433

Ajder, H.; Patrini, G.; Cavalli, F. y Cullen, L. (2019) THE STATE OF DEEPFAKES LANDSCAPE, THREATS, AND IMPACT. *DeepTrace.* https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Beccaria, C. (1764) *De los delitos y de las penas*. Alianza editorial (reimpresión del 2020 de la edición de 1968).

Blázquez, R. (2023) Deepfakes en el procedimiento probatorio. *Revista vasca de derecho procesal y arbitraje = Zuzenbide prozesala ta arbitraia euskal aldizkaria, ISSN 0214-7246, Vol. 35, Nº. 3, 2023, págs. 223-256*.

Boheemen, P.; Das, D.; Fatun, M.; Gerritsen, J.; Huijstee, M.; Jahnel, J.; Karaboga, M.; Kool, L.; Nierling,L. (2021) Tackling deepfakes in European policy. *European Parliamentary Research Service. Panel for the Future of Science and Technology*. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039

Borges, J.L. (1949) El Aleph. Disponible en https://www.ucm.es/data/cont/docs/119-2014-02-11-Borges.El%20Aleph76.pdf

Camus, A. (1942) *El mito de Sísifo*. Random House.

Canal Cruzcampo TV (2021) Cruzcampo | Con Mucho Acento [Archivo de vídeo] https://www.youtube.com/watch?v=Yewm6TfLZ3Q&ab_channel=MalagaWargames

Canal FaceToFake (2019) EL EQUIPO E, con E de España [DeepFake]. [Archivo de vídeo] https://www.youtube.com/watch?v=dj5M4s-cdAw&ab_channel=FaceToFake

Canal Nikolay Valiev (2023) TOOTHBRUSH GANGSTA'S PARADISE | AI COVER | [Archivo de vídeo] https://www.youtube.com/watch?v=2GRnrFfURYo&ab_channel=nikolayvaliev

Cerdán-Martínez V., García-Guardia M. L. y Padilla-Castillo G. (2020). Alfabetización moral digital para la detección de deepfakes y fakes audiovisuales. CIC. *Cuadernos de Información y Comunicación, 25, pp. 165-181*. https://doi.org/10.5209/ciyc.68762

Chen, H. y Magramo, K. (4 de febrero de 2024). Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. *CNN.* Recuperado de: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

Cianciardo, J. (2000) *El conflictivismo en los Derechos fundamentales.* EUNSA.

Circular 2/2015, de 19 de junio, sobre los delitos de pornografía infantil tras la reforma operada por la Ley Orgánica 1/2015 [Fiscalía General del Estado]. https://www.boe.es/buscar/doc.php?id=FIS-C-2015-00002

Citron, D. y Chesney, R. (2019) Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107 1753.* https://scholarship.law.bu.edu/faculty_scholarship/640

Código Penal [CP] Ley Orgánica 10/1995, de 23 de noviembre (España). https://www.boe.es/buscar/act.php?id=BOE-A-1995-25444

Cohen, L. y Rini, R.(2022) Deepfakes, Deep Harms. *Journal of Ethics and Social Philosophy* Vol 22 No 2: Volume XXII, Issue 2. https://doi.org/10.26556/jesp.v22i2.1628

Comisión Europea (2020) *Libro Blanco sobre la inteligencia artificial. Un enfoque un enfoque europeo orientado a la excelencia y la confianza.* https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_es?filename=commission-white-paper-artificial-intelligence-feb2020_es.pdf

Constant (1819) *La libertad de los modernos*. Alianza Editorial (edición 2019).

Constitución Española [CE] BOE núm.311, de 29 de diciembre de 1978 https://www.boe.es/buscar/act.php?id=BOE-A-1978-31229

Corte Suprema de los Estados Unidos (10 de noviembre de 1919) Abrams v. United States, 250 U.S.616. https://supreme.justia.com/cases/federal/us/250/616/

Duff, R.A. (2018) *The Realm of Criminal Law.* Oxford University Press. https://doi.org/10.1093/oso/9780199570195.001.0001

Durães, D., Freitas, P.M., Novais, P. (2024). The Relevance of Deepfakes in the Administration of Criminal Justice. En Sousa Antunes, H., Freitas, P.M., Oliveira, A.L., Martins Pereira, C., Vaz de Sequeira, E., Barreto Xavier, L. (eds) *Multidisciplinary Perspectives on Artificial Intelligence and the Law. Law, Governance and Technology Series* (vol 58, pp. 351-369). Springer, Cham. https://doi.org/10.1007/978-3-031-41264-6_19

Feinberg, J. (1984) *The Moral Limits of the Criminal Law. Volume 1. Harm to Others.* Oxford University Press.

Feinberg, J. (1985) *The Moral Limits of the Criminal Law. Volume 2. Offense to Others Oxford University Press.*
Ferrajoli, L. (1995) *Derecho y razón. Teoría del garantismo penal*, Editorial Trotta
Fioravanti, M. (2014) *Constitucionalismo. Experiencias históricas y tendencias actuales*. Editorial Trotta
Franks, M.A. y Waldman, A.E. (2019) Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions, *Maryland Law Review*, vol.78, issue 4. Pp. 892-898. https://digitalcommons.law.umaryland.edu/mlr/vol78/iss4/6/
FundéuRAE (19 de diciembre de 2022) *inteligencia artificial es la expresión del 2022 para la FundéuRAE.* Recuperado de: https://www.fundeu.es/recomendacion/inteligencia-artificial-es-la-expresion-del-2022-para-la-fundeurae/
FundéuRAE (22 de diciembre de 2023) *Candidatas a palabra del año 2023 de la FundéuRAE.* Recuperado de: https://www.fundeu.es/recomendacion/candidatas-a-palabra-del-ano-2023-de-la-fundeurae/
García de la Torre, F. (2021). Crisis del principio penal de ultima ratio.¿Debemos retomar la orientación constitucional del derecho penal? *Anales de la Cátedra Francisco Suárez. Protocolo I*, pp. 131-154. https://revistaseug.ugr.es/index.php/acfs/article/view/16747
García-Ull, F. J. (2021). Deepfakes: el próximo reto en la detección de noticias falsas. *Anàlisi: Quaderns de Comunicació i Cultura*, 64, 103-120. DOI: https://doi.org/10.5565/rev/analisi.3378
Gil, R.; Virgili-Gomà, J.; López-Gil, J.M. y García, R. (2023) Deepfakes: evolution and trends. *Soft Comput* 27, 11295–11318 (2023). https://doi.org/10.1007/s00500-023-08605-y
Goite, M. y Medina, A. (2018) La inseguridad ciudadana y los excesos en la utilización del poder punitivo. En Suarez, J.M. et al. (dir.) *Estudios jurídico penales y criminológicos.* (Volumen II, pp. 2057-2086). Dykinson.
Gómez-Bellvís, A.B. y Castro, F.J. (2022) Los delitos de expresión en redes sociales desde los efectos de la sanción penal: ¿Efecto disuasorio o efecto desafío? *Revista Chilena De Derecho Y Tecnología*, 11(1), 323–358. https://doi.org/10.5354/0719-2584.2022.66547
Gozzi, L. (2024). Giorgia Meloni: Italian PM seeks damages over deepfake porn videos. *BBC*. Recuperado de: https://www.bbc.com/news/world-europe-68615474
Habgood-Coote, J. (2023) Deepfakes and the epistemic apocalypse. *Synthese* 201, 103 (2023). https://doi.org/10.1007/s11229-023-04097-3
Habod-Coote, J. (2023) Deepfakes and the epistemic apocalypse. *Synthese 201, 103 (2023). https://doi.org/10.1007/s11229-023-04097-3*
Harris, K.R. (2021) Video on demand: what deepfakes do and how they harm. *Synthese 199,* 13373–13391 (2021). https://doi.org/10.1007/s11229-021-03379-y
Hefendehl (2007) El bien jurídico como eje material de la norma penal. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 173-190). Marcial Pons (edición de 2016).
Hefendehl, R.; Hirsch, A. y Wohlers, W. (2007.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* Marcial Pons (edición de 2016)
Hervada, J. (2000) *Lecciones propedéuticas de Filosofía del Derecho.* EUNSA.
Hirsch, A. (2007) El concepto de bien jurídico y el «principio del daño» en Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 33-48). Marcial Pons (edición de 2016).
Hörnle, T. (2007) La protección de sentimientos en el StGB. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 375-390). Marcial Pons (edición de 2016).
Hörnle, T. (2016) Theories of Criminalization. *Criminal Law and Philosophy 10,* pp. 301–314. https://doi.org/10.1007/s11572-014-9307-4
Hörnle, T. (2019) One Masterprinciple of Criminalization – Or Several Principles? *Law, Ethics and Philosophy*, Vol. 7, pp. 208-220. https://doi.org/10.31009/LEAP.2019.V7.12.
Husak, D. (2008) *Sobrecriminalización. Los límites del Derecho penal.* Marcial Pons (Edición del 2013)
Jakobs, G. (1991) *Derecho penal. Parte general. Fundamentos y teoría de la imputación.* Marcial Pons (2ª edición).
Jakobs, G. (1992) *El principio de culpabilidad*. BOE. https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-P-1992-30105101084
Kahlo, M. (2007) Sobre la relación entre el concepto de bien jurídico y la imputación objetiva en Derecho penal. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 49-64). Marcial Pons (edición de 2016).

Lascuraín, J.A. (1998) La proporcionalidad de la norma penal. *Cuadernos de Derecho público*, nº5, 1998, pp. 159-190. https://dialnet.unirioja.es/servlet/articulo?codigo=194671&orden=1&info=link

Lavanda, M. (2022) Deepfake: Cuando la inteligencia artificial amenaza el Derecho y la Democracia. *Lawgic Tec - Revista de Derecho y Tecnología Nº.02, julio 2022.* https://lawgictec.org/wp-content/uploads/Lawgic-Tec-Revista-de-Derecho-y-Tecnologia-No_2-Julio-2022.pdf

Le Guin, U.K. (1973) *Los que se alejan de Omelas*. Disponible en: https://primaduroverales.wordpress.com/wp-content/uploads/2020/06/los-que-abandona-omelas.pdf

Malacarne, A. (2023) "Profundamente falso" y "profundamente incierto" el deepfake como automated evidence en el proceso penal. Consideraciones generales. *Revista General de Derecho Procesal, ISSN-e 1696-9642, Nº.60,2023* https://www.iustel.com/v2/revistas/detalle_revista.asp?id_noticia=426109

Mania, K. (2024). Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study. *Trauma, Violence, & Abuse, 25(1), 117-129*. https://doi.org/10.1177/15248380221143772

MidJourney (2023) The Pope Drip [Publicación en foro online]. Reddit. https://www.reddit.com/r/midjourney/comments/120vhdc/the_pope_drip/?rdt=62336

Mill, J.S. (1859) *Sobre la libertad*. Edaf (versión de 1869, 12ª edición de 2020).

Mir, S. (1976) *Introducción a las bases del Derecho penal*. (Reimpresión de la 2ª edición) Bdef.

Mir, S. (1984) *Derecho penal. Parte general.* Editorial Repertor (4ª reimpresión del 2018 de la 10ª edición).

Mir, S. (2005) Límites del normativismo en Derecho penal. *Revista Electrónica de Ciencia Penal y Criminología* (en línea). 2005, núm. 07-18, p.18:1-18:24. Disponible en internet: http://criminet.ugr.es/recpc/07/recpc07-18.pdf ISSN 1695-0194 [RECPC 07-18 (2005), 23 dic]

Miró-Llinares (2006) Persona o enemigo; vigencia real o postulada de las normas; Estado de Derecho perfecto u óptimo en la práctica. Al hilo de la segunda edición del libro Derecho penal del enemigo, de Günther Jakobs y Manuel Cancio Meliá. *Revista de la Facultad de Ciencias Sociales y Jurídicas de Elche.* Vol.1; Núm. 1; pp. 133-163. https://dialnet.unirioja.es/servlet/articulo?codigo=2152760&orden=91271&info=link

Miró-Llinares, F. (2012) *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*. Marcial Pons.

Miró-Llinares, F. (2015) La criminalización de conductas "ofensivas" A propósito del debate anglosajón sobre los "límites morales" del derecho penal. *Revista Electrónica de Ciencia Penal y Criminología (en línea). 2015, núm. 17-23, pp. 1-65*. http://criminet.ugr.es/recpc/17/recpc17-23.pdf

Miró-Llinares, F. (2017) Derecho penal y 140 caracteres. Hacia una exégesis restrictiva de los delitos de expresión. En Miró-Llinares, F. (dir.) *Cometer delitos en 140 caracteres. El Derecho penal ante el odio y la radicalización en internet.* Marcial Pons.

Miró-Llinares, F. (2024) AI and criminal law reform: notes on the inadequacy of a criminalization model based on a substantive principle. (En prensa).

Morales, A.M. (2021) Inteligencia artificial y derecho penal: primeras aproximaciones. *Revista jurídica de Castilla y León.* Nº 53, pp. 177-202. https://www.jcyl.es/web/jcyl/AdministracionPublica/es/Plantilla100Detalle/1131978346397/Publicacion/1285024121501/Redaccion

OCDE (OECD/LEGAL/0449) Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OpenAI (2024) Sora. Creating video from text. https://openai.com/sora

Ozafrain, L. (2017) *El principio de ultima ratio. Fundamentos en el Derecho Internacional de los Derechos Humanos para una política criminal minimalista*. Universidad Nacional de la Plata. https://doi.org/10.35537/10915/68145

Pawelec, M. (2022) Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *DISO 1, 19 (2022)*. https://doi.org/10.1007/s44206-022-00010-6

Pawlik, M. (2023) Presupuestos y límites del derecho penal del ciudadano. *Derecho Penal y Criminología*. 44, 117 (jun. 2023), 11–30. DOI: https://doi.org/10.18601/01210483.v44n117.02

Propuesta de la Comisión Europea COM/2021/206 de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206

Rana, S.; Soaliman, M.; Gudla, C.; Sohan, F. (2024) Deepfakes – Reality under threat? *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, *Las Vegas, NV, USA, 2024, pp. 0721-0727*, doi: 10.1109/CCWC60891.2024.10427659

Raymond, K. (2021) Video on demand: what deepfakes do and how they harm. *Synthese*, Volume 199, pages 13373–13391. https://doi.org/10.1007/s11229-021-03379-y

Resolución legislativa del Parlamento Europeo P9_TA(2024)0138, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.html

Rini, R. (2020) Deepfakes and the epistemic backstop. *Philosophers' Imprint* 20 (24):1-16. https://philarchive.org/rec/RINDAT

Rini, R. (2020) Deepfakes and the Epistemic Backstop. *Philosopher's Imprint*, 20 (24):1-16. https://philarchive.org/rec/RINDAT

Rodrigo, J.F. (2020) Tragic Realism: How to regulate Deepfakes in Colombia?. *Latin American Law Review, no. 08 (2022): 125-145*, doi: https://doi.org/10.29263/lar08.2022.08

Rodrigues, P. (2023) Deepfakes pornográficas não-consensuais: a busca por um modelo de criminalização Non-consensual deepfake porn: searching for a model of criminalization. *Revista Brasileira de Ciências Criminais*. *Vol. 199, pp. 277-311*. https://doi.org/10.5281/zenodo.8380977

Rouhiainen, L. (2018) *Inteligencia Artificial*. *101 cosas que debes saber hoy sobre nuestro futur*o. Editorial Planeta.

Rousseau, J.-J. (1762) *El contrato social*. Akal (edición de 2017).

Roxin, C. (2007) ¿Es la protección de bienes jurídicos una finalidad del Derecho penal? En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 433-448). Marcial Pons (edición de 2016).

RTVE (18 de septiembre de 2023) *Investigan la difusión de imágenes de menores desnudas creadas con inteligencia artificial*. Recuperado de: https://www.rtve.es/noticias/20230918/difusion-imagenes-menores-desnudas-inteligencia-artificial/2456241.shtml

Santisteban, M. (2024) La criminalización de las ultrafalsificaciones (con especial atención a las implicaciones de la normativa europea de Servicios Digitales e Inteligencia Artificial). (En prensa)

Schwab, K. (2020). La Cuarta Revolución Industrial. *Futuro Hoy*, 1(1), 06–10. https://doi.org/10.52749/fh.v1i1.1

Seelman, K. (2007) El concepto de bien jurídico, el harm principle y el modelo del reconocimiento como criterios de merecimiento de pena. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 365-374). Marcial Pons (edición de 2016).

Seher, G. (2007) La legitimación de normas penales basada en principios y el concepto de bien jurídico. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 65-87). Marcial Pons (edición de 2016).

Silva, J.M. (2001) *La expansión del Derecho penal. Aspectos de la Política criminal en las sociedades postindustriales* (Reimpresión de la 2ª edición). Bdef. Recuperado de: https://edisciplinas.usp.br/mod/resource/view.php?id=2854945

Simó, E. (2023) Retos jurídicos derivados de la Inteligencia Artificial Generativa. Deepfakes y violencia contra las mujeres como supuesto de hecho. *InDret 2.2023* ,pp. 493-515. https://indret.com/retos-juridicos-derivados-de-la-inteligencia-artificial-generativa/

STC 117/1994 (Sala 2ª), de 25 de abril. Recurso de amparo 2016-1990. https://hj.tribunalconstitucional.es/es-ES/Resolucion/Show/2634#complete_resolucion

Stratenwerth, G. (2007) La criminalización en los delitos contra bienes jurídicos colectivos. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 357-364). Marcial Pons (edición de 2016).

STS 13505/1988 (Sala 2ª, de lo Penal), de 28 de noviembre de 1988. Nº Recurso 10/1985. https://www.poderjudicial.es/search/indexAN.jsp

STS 2037/2024 (Sala 2ª, de lo Penal), de 11 de abril de 2024. Nº Recurso 11066/2023. https://www.poderjudicial.es/search/AN/openDocument/ece23c77da7836fda0a8778d75e36f0d/20240426

Twomey J, Ching D, Aylett MP, Quayle M, Linehan C, Murphy G (2023) Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS ONE* 18(10): e0291668. https://doi.org/10.1371/journal.pone.0291668

United Unknown (2023) Iglesias y Díaz. United Unknown Guerrilla Visual. https://unitedunknown.com/revista-papel/

Van der Sloot, B., Wagensveld, Y. (2022). Deepfakes: Regulatory challenges for the synthetic society. *Computer Law & Security Review*, 46, 1-15. Article 105716. https://doi.org/10.1016/j.clsr.2022.105716

Wohlers, W. (2007) Las jornadas desde la perspectiva de un escéptico del bien jurídico. En Hefendehl, R.; von Hirsch, A. y Wohlers, W. (Eds.) *La teoría del bien jurídico ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmáticos?* (pp. 393-398). Marcial Pons (edición de 2016).