

# **La liviandad de la Audiencia Nacional española vs. la sofisticación de las decisiones sobre el dopaje de deportistas.**

## **De certezas a probabilidades y errores**

**Carmen Vázquez  
Universitat de Girona**

I. Introducción II. La incompatibilidad de la exigencia de certezas probatorias con las garantías procesales III. El carácter probabilístico del razonamiento probatorio IV. La falibilidad del conocimiento científico y su impacto en las decisiones judiciales V. ¿Cómo sabemos en un proceso judicial cuál es el conocimiento científico válido y fiable? VI. La distribución del riesgo del error en las decisiones (científicas y jurídicas) sobre dopaje

### **I. Introducción**

El deporte limpio es fundamental en las competiciones deportivas para asegurar, por un lado, una competencia en condiciones igualitarias, pero también la protección de la salud de los deportistas. No hay que olvidar que varios deportistas han sufrido daños severos o hasta la muerte buscando mejorar su rendimiento a través de sustancias exógenas u otros medios artificiales. El dopaje es un tema sumamente serio por sus consecuencias para el deporte y los deportistas.

Aunque el dopaje en sí mismo seguramente ha estado presente desde los inicios de las prácticas deportivas, la prohibición de sustancias es tan reciente como inicios del siglo pasado y, todavía más reciente, el establecimiento de controles para su identificación en el cuerpo de los deportistas. No obstante, al empleo de sustancias hace tiempo que se ha sumado el dopaje por sangre, creándose cada vez nuevos agentes eritropoyéticos y más difíciles de identificar. Y, ahora, de la mano de los avances de la ciencia y la tecnología, se enfrenta igualmente el gran desafío del dopaje genético, es decir, “el uso no terapéutico de células, genes, elementos genéticos o de la modulación de expresión genética”<sup>1</sup> para mejorar el rendimiento deportivo.

---

<sup>1</sup> Así lo identifica el Portal Iberoamericano para la lucha contra el dopaje, dentro del apartado “nuevos desafíos”.

La lucha contra el dopaje es una tarea compleja que ha sido materia de muy diversos acuerdos entre distintos Estados, creándose una Agencia Mundial Antidopaje (WADA, por sus siglas en inglés) y un Código Mundial Antidopaje (en adelante, Código Mundial), lo que ha logrado homogeneizar y estandarizar criterios y prácticas y redoblar esfuerzos en lo que se puede identificar como la “ciencia del dopaje”<sup>2</sup>; es decir, el desarrollo de conocimiento científico que permite conocer cuáles son las sustancias, medicamentos y métodos que pueden ser utilizados para doparse y que, por tanto, deben ser prohibidos y también sometidos a control en las competiciones deportivas.<sup>3</sup> Como dice la Agencia Mundial Antidopaje “los controles de dopaje a menudo evocan la tradicional recolección de muestras y su análisis. No obstante, las tareas *anti-doping* actualmente se basan en mucho más que eso.”<sup>4</sup>

Sin lugar a duda, uno de los grandes avances de la ciencia antidopaje ha sido el llamado “pasaporte biológico del deportista”, “que es la recopilación de los parámetros fisiológicos de un deportista, mediante varios análisis de sangre y orina a lo largo de un periodo de tiempo (normalmente, 5 o 6 analíticas). A estos datos se les aplica un modelo matemático con el que se determina cuál es el perfil biológico del deportista, y se establece un intervalo de valores entre los que deberían moverse los siguientes análisis.”<sup>5</sup> Este enfoque de *detección indirecta* es adecuado porque los marcadores biológicos indirectos que se miden pueden revelar variaciones anormales en la sangre y la orina *potencialmente* inducidas por el dopaje. Nótese pues que el dopaje y los controles para identificarlo ya no se limitan a encontrar directamente rastros de una sustancia prohibida, pues ello

---

<sup>2</sup> Entre ellos, por ejemplo, resalta el trabajo que se lleva a cabo sobre los laboratorios de control de dopaje, habiéndose desarrollado un Estándar Internacional que define los requisitos que los laboratorios deben cumplir para garantizar que los resultados de los análisis y las pruebas presentadas sean válidos; y el Estándar Internacional de Gestión de Resultados, que establece, entre otras cosas, los diversos procedimientos que los expertos deben seguir cuando se está ante los llamados resultados atípicos o anómalos y resultados adversos en las muestras o perfiles de los deportistas.

Como se verá, sobre todo en el último epígrafe, estos documentos contienen información muy detallada sobre el procedimiento probatorio del dopaje de deportistas y, por ello, son relevantes para las decisiones jurídicas sobre esta materia.

<sup>3</sup> También tienen un lugar aquí las ciencias sociales, pues la WADA se han planteado como objetivo intentar entender por qué algunos deportistas deciden doparse y, con ello, buscar mejorar sus estrategias de prevención. Véase <https://www.wada-ama.org/en/data-research/social-science-research>

<sup>4</sup> <https://www.wada-ama.org/en/data-research>

<sup>5</sup> <http://deportelimpio.com/informacion-básica/la-ciencia-en-la-lucha-contra-el-dopaje/>

dejaría sin sancionar el empleo de sustancias de difícil identificación y del uso de métodos prohibidos.<sup>6</sup>

Estamos pues, ante un escenario muy sofisticado tanto en las alternativas que tiene un deportista para doparse, como en las herramientas que se han desarrollado para sancionarlo por la detección de la presencia de una sustancia prohibida en su organismo o por su probable dopaje en atención a los efectos que ciertas sustancias o métodos prohibidos han generado en su organismo.<sup>7</sup>

También desde el punto de vista normativo estamos ante un escenario sofisticado, pues encontramos no solo normativa nacional e internacional detallada sobre el dopaje y la actuación de los diversos implicados en los controles antidopaje y sus resultados, sino que además claramente el conocimiento científico y los expertos juegan aquí un rol fundamental. Es un contexto entonces donde los legos interactúan con los expertos en la toma de decisiones.

¿Cómo es tratado todo ese sofisticado escenario en el que se resuelve una sanción de dopaje a un deportista cuando esta es cuestionada ante los tribunales de justicia? Más allá de dar cuenta de la posible gran diversidad de respuestas que podríamos encontrar, quisiera centrarme en una reciente decisión de la Audiencia Nacional española que, desafortunadamente, muestra una liviandad abrumadora en la resolución de un caso de dopaje. Pero vale la pena dedicar el análisis sobre todo al criterio de su resolución y no en sí mismo al caso concreto que resolvió. Entre otras cosas porque quizá ese criterio podría ser compartido

---

<sup>6</sup> Cada año, la WADA publica la lista de sustancias y métodos que considera prohibidos porque tienen el potencial de mejorar el rendimiento deportivo o son riesgosas para la salud del deportista o vulneran el espíritu del deporte. Esas sustancias y métodos pueden estar prohibidos solo en algunos deportes o en competición o fuera de competición o siempre o en función de la vía de administración o por encima de determinada concentración (conocidas como “sustancias umbral”).

<sup>7</sup> Nótese que algunos supuestos de dopaje aluden a la mera “presencia de cualquier cantidad comunicada de una Sustancia Prohibida, de sus Metabolitos o Marcadores en la Muestra de un Deportista” (art. 2.1.3. del Código Mundial Antidopaje), la infracción a la norma se da cuando se identifican la sustancia, sus metabolitos o marcadores. En ese escenario se habla de una responsabilidad objetiva del deportista. Como se verá, la situación es distinta cuando se trata de dopaje deportivo a través del uso de métodos prohibidos, como el dopaje sanguíneo.

por otros Tribunales o usado en otras resoluciones de la misma Audiencia Nacional con terribles consecuencias para la lucha contra el dopaje.

La liviandad de la Audiencia Nacional radica básicamente en su exigencia de certezas probatorias, algo imposible en el ámbito probatorio en general, incluido el científico, y en el jurídico en particular. Así, la Sala Contencioso-Administrativa, Sección Sexta, de la Audiencia Nacional en su sentencia de apelación 39/2021, ha afirmado lo siguiente:

No es respetuoso con la presunción de inocencia afirmar que «existe una alta probabilidad» en el consumo de sustancias o métodos prohibidos, puesto que *el derecho sancionador no puede operar sobre el terreno de las probabilidades sino sobre el plano de las certezas*. Solo puede ser sancionada la persona, en este caso el deportista, que de manera *indubitada* haya cometido la infracción y la conducta tipificada; no respeta el principio de presunción de inocencia el acuerdo sancionador que descansa en un mayor o menor grado de posibilidad o probabilidad de que una infracción se haya cometido. (F.J. 3º) [Cursivas añadidas].

Las primeras cuatro líneas con las que la Audiencia Nacional española zanja el recurso de apelación 4/2021 interpuesto por la Agencia Mundial Antidopaje,<sup>8</sup> están plagadas de serios malentendidos en materia probatoria y podría decirse que algunos de ellos incluso desconocen jurisprudencia del Tribunal Supremo que reconoce la incertidumbre fáctica. Es dable decir que dicho reconocimiento se ha visto reflejado en la adopción del estándar de prueba denominado “más allá de toda duda razonable”, que sería el más exigente de todos y, por ello, aplicable al ámbito penal. Este estándar de prueba, que debe ser satisfecho para derrotar la presunción de inocencia, tiene la peculiaridad de haber introducido en nuestra cultura jurídica el debate sobre la “duda razonable” respecto la verdad de una hipótesis fáctica, lo que resulta incompatible con la “certeza”<sup>9</sup>.

Para profundizar todo lo anterior, se seguirá la siguiente hoja de ruta. En primer lugar, es indispensable argumentar por qué no debemos esperar certezas

---

<sup>8</sup> Se recurre la sentencia de 20 de octubre de 2020, dictada en procedimiento abreviado 47/2019, seguido ante el Juzgado Central de lo Contencioso-Administrativo nº 7.

<sup>9</sup> Con esto no estoy sugiriendo que el estándar de prueba aplicable en el ámbito administrativo se el “más allá de toda duda razonable”, solo estoy afirmando que en el ámbito donde más exigencias probatorias se hacen, como el penal, se ha incorporado la “duda razonable” que es incompatible con la certeza.

racionales ni de la prueba en el derecho ni de la prueba científica en general. Pondré especial atención en mostrar por qué la declaración de hechos probados en el derecho no puede requerir ningún tipo de certeza (o cualquier estado mental) de los jueces y juezas sobre la ocurrencia de tales hechos. Con ello se podrá observar que la exigencia de certeza por parte de la Audiencia Nacional está vinculada con una concepción de la prueba incompatible con el debido proceso.

Una vez descartada la certeza, psicológica o racional, como requisito para considerar probado un hecho, analizaré la noción de probabilidad, como característica intrínseca del razonamiento probatorio en general y del jurídico en particular. Ahora bien, considerar el razonamiento probatorio como probabilístico es decir todavía muy poco, debido a la existencia de muy distintos tipos de probabilidad. Argumentaré porqué en el derecho solo podemos acudir a la probabilidad lógica o inductiva para la toma de decisiones judiciales sobre la prueba de determinados hechos (irrepetibles y normalmente pasados). Esto de ninguna manera soslaya la importancia que pueda tener la probabilidad estadística cuando la hipótesis del caso también tiene esta naturaleza o la probabilidad bayesiana, por ejemplo, cuando es utilizada en cierto tipo de pruebas periciales.

Dado que el caso en comento trata sobre conocimiento científico que funda una decisión judicial, la Audiencia Nacional no solo erra en su concepción de la prueba, también podría estar partiendo de una concepción equivocada sobre las pruebas científicas al suponer que estas podrían ofrecer algo más que probabilidades. Puede decirse desde ya que no hay nada menos científico que exigirle certezas a la ciencia, pues el conocimiento científico, como veremos es, por naturaleza, falible. La ciencia es un tipo de investigación permanente, siempre abierta, que exige una disposición a revisar incluso las afirmaciones más consolidadas cuando hay nuevas pruebas que así lo requieran; y, por ello, se considera que todas las afirmaciones científicas son potencialmente provisionales. Por eso, en tercer lugar, se analizará el carácter falible del conocimiento científico, pues por más que se trate de conocimiento sólido siempre está abierta la posibilidad de que haya avances teóricos, más

desarrollos tecnológicos o nuevos experimentos que terminen mostrando la falsedad de lo que en un momento determinado se asumía como verdadero o correcto.

Lo anterior de ninguna manera supone asumir una actitud escéptica ante la ciencia o el conocimiento, sino solo conducir nuestra atención a los genuinos problemas del uso del conocimiento científico para la toma de decisiones jurídicas o con efectos jurídicos. Por eso, en cuarto lugar, se tratará la fiabilidad del conocimiento científico, poniendo especial cuidado en cómo los sistemas jurídicos pueden atender a ello. A estos efectos, acudiremos a los conocidos criterios Daubert, recientemente utilizados por la Sala Penal del Tribunal Supremo español, y que nos muestran la necesidad de considerar el trabajo de validación de las teorías, métodos, técnicas, etc., que hacen las comunidades expertas.

Finalmente, en quinto lugar, se abordará el caso concreto del dopaje, para mostrar lo racional que resulta el razonamiento probatorio cuando, en lugar de exigir lo imposible (como las certezas racionales) o denigrar al conocimiento científico valorándolo con criterios netamente subjetivos (como las certezas psicológicas), nos centramos en el análisis de los errores y la distribución de su ineliminable riesgo en la toma de decisiones. Para ello resulta indispensable abordar las presunciones y, sobre todo, los estándares de prueba aplicables al ámbito del dopaje deportivo.

## **II. La incompatibilidad de la exigencia de certezas probatorias con las garantías procesales**

Vale la pena, en primer lugar, dedicar algunas líneas a distinguir posibles significados del término “certeza” para mostrar porqué las exigencias de la Audiencia Nacional no tienen cabida en nuestro sistema constitucional actual.

Según Baron Reed (2008), es posible distinguir al menos tres aproximaciones a la llamada “certeza”:

- a. Certeza como una preferencia respecto de cómo debemos conducirnos en la vida. En ese sentido, se identifica con la certeza moral y regula nuestro comportamiento.<sup>10</sup>
- b. Certeza como el estado mental de una persona ante determinados hechos, que se identifica como certeza psicológica.
- c. Certeza como el más alto nivel de conocimiento o de nuestras creencias sobre el mundo o incluso como algo superior al conocimiento, en este caso se trataría de una certeza racional, propiamente de carácter epistémico.

Para profundizar en estas tres distinciones, vale la pena usar como ejemplos sentencias que parecen presuponer cada una de estas maneras de concebir a la certeza. Por ejemplo, la Sala de lo Penal del Tribunal Supremo, en una antigua sentencia, nos brinda un claro ejemplo de la certeza moral que consideran debería regular el comportamiento de los jueces al valorar la prueba:

...de acuerdo con la reiterada jurisprudencia del Tribunal Supremo «lo declarado probado por las Audiencias deviene indestructible e invulnerable» (Sentencia de 15 de marzo de 1980) y la facultad de apreciación de la prueba en conciencia que concede a los Tribunales el art. 741 de la LECrim., no exige explicación ni razonamiento (Sentencia de 11 de octubre de 1978), de modo que el juzgador, a la hora de apreciar las pruebas practicadas, las alegaciones de las partes y las declaraciones o manifestaciones del acusado, lo hará no ya sin reminiscencias de valoración tasada o predeterminada por la Ley, o siguiendo las reglas de la sana crítica, o de manera simplemente lógica o racional sino de un modo tan libérrimo y omnímodo que no tiene más freno a su soberana facultad valorativa que el de proceder al análisis y a la consecutiva ponderación *con arreglo a su propia conciencia, a los dictados de su razón analítica y a una intención que se presume siempre recta e imparcial* (Sentencia de 10 de febrero de 1978). [Cursivas añadidas].

Es más, la literalidad de la Ley de Enjuiciamiento Criminal, artículo 741, parece conducirnos todavía hacia este tipo de certeza moral: “El Tribunal, apreciando *según su conciencia*, las pruebas practicadas en el juicio, las razones expuestas por la acusación y la defensa, y lo manifestado por los mismos procesados, ...”.

---

<sup>10</sup> Este tipo de certeza estaba ya presente, por ejemplo, en Descartes (PW 1: 289 n. 2): “la certeza moral es la certeza que resulta suficiente para regular nuestro comportamiento o aquella que está a la altura de la certeza que tenemos sobre cuestiones relativas a nuestras conductas cotidianas sobre las que normalmente nunca dudamos, aunque sabemos que es posible, claramente, que sean falsas”.

No obstante, entendida así la certeza, no parece haber un espacio ni siquiera para la prueba o el conocimiento de los hechos, dado que se trata de una suerte de preferencia sobre cómo conducirnos en la vida, una preferencia que no es necesariamente siquiera susceptible de verdad o falsedad. Es más, la propia sentencia citada lo reconoce al decir que la declaración de hechos probados, realizada con certeza moral, “*no exige explicación ni razonamiento*”<sup>11</sup>. Resulta claro, sin embargo, que esta manera de entender la certeza debe ser completamente descartada, dada su evidente incompatibilidad con las garantías procesales de un estado Constitucional.<sup>12</sup>

Por otro lado, la certeza psicológica ha estado muy presente en el ámbito judicial, representada como el estado mental que debe tener un juzgador para la declaración de hechos probados. El famoso caso estadounidense *In re Winship*<sup>13</sup> es un claro ejemplo de ello:

El estándar de prueba es un intento de instruir al Juzgador de los hechos acerca del grado de confianza que nuestra sociedad piensa que *él debe tener* en la corrección de las conclusiones sobre los hechos para tomar decisiones.

En nuestro país, el Tribunal Constitucional ofrece un claro ejemplo de la vinculación entre que un hecho esté jurídicamente probado y la certeza psicológica del juzgador en la STC 124/1983 (FJ 1.º):

conviene también señalar aquí que el razonamiento del juez de instrucción de Noya contenido con el considerando primero de su sentencia es impecable desde el punto de vista constitucional, esto es, bajo el enfoque del artículo 24.2 CE, pues si a él, bajo el sistema de libre apreciación de la prueba del artículo 741 LECrim, las pruebas practicadas no *le convencieron* de la culpabilidad de los encartados, actuó perfectamente al absolverlos, porque bajo el imperativo del artículo 24.2 CE es evidente que nadie puede ser condenado solo porque existan indicios que apunten hacia la posible participación del o de los acusados en los hechos delictivos. Sucede, sin embargo, que

---

<sup>11</sup> Otro ejemplo lo brinda la Fiscalía del Tribunal Supremo en su Memoria de 1983: “para el descubrimiento de la verdad, no debe sujetarse el criterio judicial a reglas científicas, ni a moldes preconcebidos y determinados por la ley, sino más bien debe fiarse al sentido íntimo e innato que guía a todo hombre en los actos importantes de su vida”.

<sup>12</sup> La certeza así entendida tiene un origen religioso relacionado con el buen comportamiento. Así, Aliste Santos (2021: 103 y 104) dice que: “[e]l expediente de la certeza moral, cuya originalidad ha sido justamente afirmada por el procesalismo canónico, es el resultado de una intensa construcción conceptual de los procesalistas eclesiásticos. Su actual positivización es heredera de la época de codificación del Derecho de la Iglesia, aunque hunde sus raíces normativas en el viejo Derecho canónico, y su mención en las obras de diversos canonistas anteriores a la codificación evidencia una elaboración doctrinal de larga tradición”.

<sup>13</sup> *In re Winship*, 397 U.S. 358 (1970).



si con los mismos elementos probatorios otro tribunal, el de apelación, apreciando según su conciencia las pruebas practicadas en el juicio (art. 741 LECrm) llega un resultado contradictorio, y en este caso, al *convencimiento razonable y razonado* de la culpabilidad de los antes absueltos, no por ello puede afirmarse violación alguna contra la presunción de inocencia, siempre que las pruebas practicadas en el juicio sean las que le proporcionen fundamento para su convicción. Así las cosas no puede hablarse de vulneración de la presunción de inocencia, sino de una discrepancia en la valoración de la prueba hecha por dos órganos judiciales igualmente libres para valorar en conciencia, con el resultado de que entre ambas valoraciones ha de imponerse la del Tribunal de apelación. [Cursivas añadidas].

El párrafo anterior nos indica, en primer lugar, que un hecho estaría probado si, y solo si, el juzgador alcanzó la convicción de que ese hecho ocurrió. Ahora bien, hay una diferencia importante entre la STS antes citada y esta sentencia del Tribunal Constitucional, pues la última sí otorga cierto rol a las pruebas: “siempre que las pruebas proporcionen el fundamento de la convicción”.<sup>14</sup> No obstante, hay un serio problema para satisfacer dicha exigencia: la convicción o la creencia es un estado mental que nos sucede, esto quiere decir que *no podemos decidir* tenerlo en función de las pruebas o con independencia a ellas. Sólo podríamos encontrar explicaciones sobre qué fue lo que *causó* nuestra creencia, pero no podemos decidir creer.<sup>15</sup> Por eso, por ejemplo, no tiene sentido la manera en que Murphy (1980: 101) define el resultado de la institución probatoria: “[e]s la medida o el grado de certeza ... que las pruebas *deben* generar en la mente del juzgador de los hechos”. Y, en cambio, es totalmente compatible con lo que en España ha dicho De La Oliva (2002: 514): “no parece razonable pedir que se exprese lo que pertenece a los internos procesos psicológicos de convicción, muchas veces parcialmente objetivables, sí, pero también parcialmente pertenecientes al ámbito de lo inefable”.

Así pues, las circunstancias que conducen causalmente a adquirir la certeza psicológica son solo muy superficialmente explicables. Además, la situación se complica porque un sujeto puede creer que *q* es verdadera y, no obstante, *q* ser falsa. El hecho de que se tenga una certeza psicológica respecto un hecho no permite inferencia alguna sobre la verdad de ese hecho ni del grado de

---

<sup>14</sup> Vale la pena aclarar aquí que, desde el punto de vista de la epistemología, las creencias, la convicción, la certeza psicológica, son todas ellas un mismo tipo de “actitud proposicional”. Es decir, el estado mental o psicológico que una persona tiene sobre una proposición *q*. Y, atendiendo a ello, es que se consideran sinónimos.

<sup>15</sup> Aunque, desde luego, podríamos decidir hacer cosas para intentar cambiar nuestras creencias.

corroboración que las pruebas permitan razonablemente concluir sobre tal hecho (Ferrer Beltrán, 2021: 32)<sup>16</sup>. O, como afirma, Taruffo (2010: 104):

La intensidad y la profundidad del convencimiento no garantizan de ningún modo la verdad de su objeto. Muchas personas están intensamente persuadidas de muchas cosas respecto de ellas mismas, de los otros, de la edad del mundo, de la existencia del diablo y de mundos ultraterrenos, pero es claro que la fuerza de su persuasión no vuelve verdadero aquello en lo que creen.

En este punto, ya que se habla de la intensidad del convencimiento, vale la pena traer a colación la llamada certeza epistémica, que está relacionada con ciertas características del conocimiento o de nuestras creencias que las harían superiores a otras. Así se dice que no cualquier creencia alcanza la certeza, sino solo algunas que serían indubitables o infalibles o inderrotables. Ahora bien, como dice Reed (2008), con la certeza como indubitabilidad se enfrenta un serio dilema: o se tienen buenas razones para no-dudar o no se tienen. Si no se tienen buenas razones, entonces solo estamos ante una certeza psicológica, sin ningún carácter epistémico relevante; en cambio, si se tienen buenas razones para no dudar, el trabajo lo hacen las razones y no el hecho de que la creencia sea indubitable. Algo similar sucedería con las creencias infalibles, que serían aquellas creencias del sujeto que tienen una garantía (*guaranteed*) de ser verdaderas. Ahora bien, la creencia estaría garantizada como verdadera por los fundamentos que tiene el sujeto. Pero, si la verdad de la creencia está garantizada por los fundamentos para sostenerla, entonces la certeza de la creencia depende de esos fundamentos. Es decir, la creencia sería cierta, no en virtud del hecho de que se garantice que es verdadera, sino en virtud de los fundamentos que harían posible esa garantía.<sup>17</sup>

El párrafo anterior nos muestra que el debate a final de cuentas se centra en *la justificación* (fundamentos o razones) de las creencias, dado que las creencias

---

<sup>16</sup> “No hay vinculación conceptual ni empírica demostrada entre el grado de confianza en una creencia de un sujeto cualquiera y el valor de verdad del contenido proposicional de la creencia ni tampoco con el grado de apoyo que lo creído tenga en las pruebas existentes” (Ferrer Beltrán, 2021: 33).

<sup>17</sup> Por lo que hace a la certeza como creencias inderrotables debería sostenerse que son aquellas inmunes a la duda dentro de un sistema de creencias. Esto supone cierta interacción de las creencias, en las que unas impactan sobre las otras. Pero en un sistema de creencias podemos tener creencias contradictorias, además de *creencias falsas*.

en sí mismas no justifican.<sup>18</sup> Y ello es así en cualquier ámbito de conocimiento empírico, es decir, siempre que queramos saber algo sobre el mundo, sea en el proceso judicial o en la ciencia. En esa línea, por ejemplo, ya lo advertía Popper (1962: 56):

una experiencia subjetiva, o un sentimiento de convicción, nunca pueden justificar un enunciado científico; semejantes experiencias y convicciones no pueden desempeñar en la ciencia otro papel que el de objeto de una indagación empírica (psicológica). Por tanto, que sea un sentimiento de convicción nunca podrá justificar un enunciado. Por tanto, puedo estar absolutamente convencido de la verdad de un enunciado, seguro de la evidencia de mis percepciones, abrumado por la intensidad de mi experiencia, puede parecerme absurda toda duda. Pero, ¿aporta, acaso, todo ello la más leve razón a la ciencia para aceptar mis enunciados? ¿Puede justificarse ningún enunciado por el hecho de que [yo] esté absolutamente convencido de su verdad? La única respuesta posible es que no...

Nuevamente, lo que importa son las razones que justificarían nuestras proposiciones sobre determinados hechos, no nuestras creencias o certezas.<sup>19</sup> Desde el punto de vista de la epistemología, “verdad” y “certeza” no deben confundirse, la certeza *no puede* ocupar el lugar de la verdad. Pero es que, además, en los procesos judiciales actuales suele haber muchas limitaciones respecto a la información que tienen los jueces y magistrados para formarse sus creencias: las partes suelen tener una incidencia importante en el ofrecimiento de pruebas, hay un tiempo limitado para la conformación del conjunto de pruebas, hay reglas jurídicas que exigen la exclusión de información relevante (que podría, además, ser fiable);<sup>20</sup> etc. Por lo que la corrección de la decisión fáctica dependerá del conjunto de pruebas disponibles.

---

<sup>18</sup> Ferrer Beltrán (2021: 186) se refiere a esta cuestión en el contexto de las decisiones judiciales: “El debate en la literatura epistemológica se centra precisamente en el punto de la justificación de las creencias. Por el contrario, la asunción mayoritaria en materia probatoria jurídica parece ser la absurda tesis de que las creencias el juzgador justifican sus decisiones.”

<sup>19</sup> Precisamente por ello, algunas veces encontramos que las alusiones a la “certeza” en el discurso probatorio solo son retóricas; como cuando Aliste Santos (2021: 108) dice “la certeza moral exigida al juez ... para la formación y dictado de su sentencia implica *la máxima probabilidad*” [cursivas añadidas].

<sup>20</sup> En palabras de Medina Alcoz (2016: 115): “... el conocimiento procesal de los hechos es imperfecto o relativo y ello por dos razones. La primera es de carácter procesal de esta modalidad del conocimiento empírico. A diferencia de la investigación histórica o detectivesca, la procesal tiene una finalidad práctica: la resolución de conflictos mediante la aplicación de normas. A su vez, el ordenamiento canaliza en el proceso la realización de otros fines, lo que no ocurre en otras formas de conocimiento empírico. Todo ello singulariza el conocimiento procesal como un conocimiento institucional que persigue la averiguación de la verdad, pero condicionado por reglas no epistemológicas que responden a otros objetivos [:] ... los plazos procesales; la previsión de que la prueba documental y pericial acompaña la demanda civil; la obligación de acreditar el pago del precio estipulado únicamente mediante confesión o documento acreditativo

En definitiva, la certeza psicológica no es una buena compañera para la declaración judicial de hechos probados si consideramos que al menos una parte de ella es inefable<sup>21</sup> y que el hecho de que el juez se forme la creencia sobre un hecho no es justificación alguna sobre la ocurrencia de ese hecho. Y a esto podríamos agregar una cuestión fundamental: si el derecho pretende regular el comportamiento de los ciudadanos, es indispensable asumir que la institución probatoria tiene un compromiso con la verdad de los hechos<sup>22</sup>. Podemos además añadir el hecho de que algunas veces el propio derecho establece que el juez tiene que decidir *con independencia* a sus creencias, por ejemplo, cuando se está en presencia de una presunción *iuris tantum* que exige tener como probado un hecho presumido para el que no se ofrecen pruebas más allá de las exigidas sobre el hecho base; o cuando hay una presunción *iure et de iure* que ni siquiera pruebas en contra del hecho presumido son admisibles, etc.

Desafortunadamente, aunque han pasado 40 años desde la referida sentencia 124/1983 del Tribunal Constitucional, la concepción de la prueba que supone sigue permeando ampliamente en los juzgados y tribunales españoles.<sup>23</sup> Una

---

en el juicio de desahucio... . La relatividad del conocimiento procesal deriva también del hecho en sí de consistir una modalidad del conocimiento humano. El conocimiento empírico no permite acceder a la verdad absoluta. Ello implica que es, no imposible, sino imperfecto; no está radicalmente desconectado de la realidad objetiva ni abocado al irracionalismo.”

Al respecto, puede verse Ferrer (2007) y Haack (2020), con un título en sí mismo sugerente del problema “¿Diferencias irreconciliables? El problemático matrimonio entre la ciencia y el derecho”.

<sup>21</sup> Como lo advierte Ferrajoli (1997: 139), las razones íntimas en conciencia han dado lugar a “una de las páginas políticamente más amargas e intelectualmente más deprimentes de la historia de las instituciones penales”.

El éxito de la institución probatoria se produce cuando las proposiciones sobre los hechos que se declaran probadas son verdaderas, no cuando se logra convencer o persuadir a los jueces.

<sup>22</sup> Pues si el ciudadano calcula que la aplicación de las normas es indiferente a si ha observado o no lo que estas ordenan, no tendrá ningún incentivo para adecuar su comportamiento al derecho. Es decir, si al ciudadano se le sanciona con independencia a que se haya comportado conforme las normas jurídicas, pues no tendrá razones para limitar su conducta.

<sup>23</sup> Como afirma Medina Alcoz (2016: 109-110), “la dogmática jurídica, centrada en la construcción de teorías explicativas del sistema jurídico y en la interpretación de normas, ha prestado tradicionalmente escasa atención a los hechos, como si no fueran derecho. Partía de que el juicio al hecho debe abandonarse irremediabilmente a la subjetividad del juzgador o, a la inversa, daba por sentado que ese juicio es tan objetivo que no plantea problemas jurídicos. La confianza ciega en el conocimiento empírico heredada de la ilustración condujo a considerar que si un hecho es cierto siempre podrá ser probado, por lo que los juristas deben emplearse en la tarea que es propiamente suya: la elaboración de conceptos y la interpretación de las normas. ... Menos todavía en el contencioso-administrativo histórico como proceso objetivo destinado a controlar la legalidad para expulsar el acto administrativo e ilegal; configurado como segunda

concepción que asume que el objetivo de la institución probatoria es lograr *cierto estado mental en el juzgador*: su certeza, su convicción, su creencia o su convencimiento. No obstante, tal como ha dicho el Magistrado Ramírez Ortiz (2023), “[u]n modelo constitucional de proceso... que prohíbe la arbitrariedad de los poderes públicos (artículo 9.3 CE) garantiza la tutela judicial efectiva (artículo 24.1 CE) y la presunción de inocencia (artículo 24.2 CE) e impone el deber de motivar las decisiones judiciales (artículo 120.3 CE) apunta en otra dirección: el razonamiento probatorio en materia de hechos tiene que ser intersubjetivamente controlable...”.

La verdad de enunciado que afirma que un hecho ocurrió no depende de las creencias de quien lo afirma, sino del mundo. Y que esa ocurrencia esté probada no depende tampoco de las creencias de nadie, sino de las pruebas disponibles. En otras palabras, la justificación o “el grado de confirmación (o apoyo inductivo o corroboración, como se prefiera) [que den las pruebas a los hechos] es objetivo e intersubjetivamente controlable” (Ferrer Beltrán, 2021: 87). Ahora bien, ese conjunto de pruebas solo permitirá atribuir cierto grado de confirmación o de probabilidad de que un hecho (o, mejor dicho, una proposición fáctica o enunciado sobre los hechos) es verdadero<sup>24</sup>.

Hasta aquí podemos concluir que la Audiencia Nacional al afirmar que “el derecho sancionador no puede operar sobre el terreno de las probabilidades sino sobre el plano de las certezas”, está dando cuenta, simple y llanamente, de su propia concepción subjetivista de la prueba, sin ni siquiera entrar a la valoración del conocimiento experto implicado en el caso. Una concepción errónea, tanto desde el punto de vista epistemológico como jurídico; y, por supuesto, muy alejada de la realidad, pues ignora que el razonamiento probatorio es de carácter probabilístico.<sup>25</sup> Ahora bien, ¿qué exactamente quiere decir esto?

---

instancia o recurso de casación, entendía que los hechos eran solo problema de la administración activa.”

<sup>24</sup> En realidad, no son los hechos los que se prueban, sino nuestras afirmaciones sobre esos hechos.

<sup>25</sup> Desafortunadamente, esta decisión de la Audiencia Nacional desmiente lo afirmado ya en 1994 por Twining (1994: 97): “[p]or supuesto, la retórica aspiración y suposiciones irreflexivas en el discurso ordinario de legos y juristas puede proporcionar numerosos ejemplos de palabrería suelta sobre la “certeza”, tanto en lo que respecta a las normas como, quizá menos comúnmente, en lo que respecta a la determinación de los hechos. Pero, cabe preguntarse, ... ¿quién cree en

### III. El carácter probabilístico del razonamiento probatorio

La probabilidad es un tema complejo y muy serio, al contrario de lo que parece sugerir la Audiencia Nacional al asimilarlo con “meras posibilidades”; debe quedar claro que no es suficiente la mera posibilidad de que un hecho haya acontecido, sino que se exige que sea probable que dicho hecho ocurriera. El Tribunal Supremo, nos ofrece una decisión que justamente plantea esta distinción, se trata de la STS 353/2022 de 6 de abril de 2022, en donde la Sala de lo Penal afirma que:

De ahí, que la suficiencia de la prueba para fijar la verdad procesal se funde no tanto en la regla de la certeza, entendida como reproducción exacta, sino en la *correspondencia aproximativa*: esto es, que el hecho declarado probado se ajuste, desde la lógica de lo razonable, en términos altísimamente prevalentes, a la manera en que debió producirse el hecho histórico, convirtiendo, correlativamente, a las otras hipótesis fácticas en liza, en manifiestamente improbables, reduciéndolas a un grado de mera posibilidad fenomenológica escasa o irrelevante.

Entonces, dejando de lado las meras posibilidades, “cualquiera que aspire a la racionalidad debe guiarse ante la incertidumbre por probabilidades” (Black, 1984: 88). En el proceso judicial enfrentamos un escenario de incertidumbre cuando se trata de reconstruir hechos irrepetibles,<sup>26</sup> la mayoría de las veces pasados, y se debe hacer en función de las pruebas admitidas y practicadas<sup>27</sup>. Por ello, puede decirse que un hecho se deberá declarar probado si tiene *suficiente* apoyo en los elementos de juicio disponibles. Estamos, entonces, en la justificación epistémica (no en las creencias o convicciones o certezas) sobre las afirmaciones fácticas realizadas por las partes, que es probabilística.

---

la certeza sobre los hechos? ¿Merece la pena atacar el ‘mito de la certeza’? En lo que respecta a la determinación de los hechos, es sin duda una caricatura equiparar ‘el mito de la certeza’ con ‘el punto de vista de las instituciones’. Se reconoce amplia y abiertamente que las cuestiones probatorias deben concebirse en términos de probabilidades relativas y no de certezas.”

<sup>26</sup> Pese a la concepción de la Audiencia Nacional, el derecho reconoce la incertidumbre fáctica. Tenemos reglas de decisión en contextos de incertidumbre, como la carga de la prueba, las presunciones o los estándares de prueba. Es más, los recursos y la apelación son ya un reconocimiento respecto la posibilidad de errores en el proceso judicial (Twining, 1984: 97).

<sup>27</sup> De hecho, si en algo se caracteriza la prueba en el derecho es que no siempre se toman decisiones sobre los hechos probados con todos los elementos de juicio relevantes disponibles, pues hay reglas de exclusión de información que no puede ser considerada debido a otros objetivos del derecho (como la protección de ciertas garantías contra la obtención ilícita de pruebas). Esta caracterización de la prueba jurídica es realizada por Ferrer Beltrán (2007: 68ss); sobre el impacto de las reglas de exclusión en el ámbito probatorio, véase Laudan (2006).

Ahora bien, hay muchas nociones de probabilidad y distintas clasificaciones de los conceptos de probabilidad<sup>28</sup>. Siguiendo a Ferrer Beltrán (2007: 94) podemos distinguir la probabilidad de eventos o sucesos y la probabilidad de proposiciones. La primera, llamada probabilidad frecuentista o estadística, mide “la frecuencia con la que un evento se produce en una sucesión dada de acontecimientos” (Ferrer Beltrán, 2007: 94). Es una noción objetiva, no subjetiva, de probabilidad, ya que mide el número de posibilidades de que un suceso ocurra respecto el número de posibilidades de que no ocurra. La segunda, en cambio, mide nuestro grado de conocimiento sobre el mundo y, por ello, es de carácter epistémico. Se llama probabilidad de proposiciones, entonces, porque “gradúa las posibilidades de que una determinada proposición sea verdadera”.

La probabilidad estadística solo tiene un rol en el proceso judicial cuando la hipótesis que se plantea en un caso es estadística; en cambio, cuando se habla de hechos individuales, como en la inmensa mayoría de procesos judiciales, al no ser informativa sobre ellos, tal probabilidad no tiene un espacio. Pensemos, por ejemplo, en los casos de discriminación indirecta contra ciertos grupos o clases: así, por ejemplo, dado que “la labor del hogar es realizada preponderantemente por mujeres”, una norma que excluya a quien se dedique a las labores del hogar de la seguridad social podría ser discriminatoria contra las mujeres (dado su impacto desproporcionado en relación con el que tendría sobre los hombres). Como bien dice Giles (2021: 173):

[!]o que llamamos ‘impacto discriminatorio’ se conforma, generalmente, con enunciados que suponen un enfoque de los hechos de tipo frecuencial o estadístico y no uno individualizado (...). En ellos no se hace pie en la individualidad de los eventos, en lo que los hace únicos e irrepetibles, sino en lo que comparten con otros eventos del mismo tipo. Los consideran instancias positivas o negativas de cierto atributo, susceptibles de ser contadas (o, de modo más elocuente, tratadas como un número).<sup>29</sup>

Por otro lado, respecto de la probabilidad de proposiciones podemos encontrar dos modos principales de concebirla: una, la probabilidad inductiva o lógica; otra,

---

<sup>28</sup> Al respecto, puede verse Dei Vecchi (2020).

<sup>29</sup> El Tribunal Europeo de Derechos Humanos, por ejemplo, decidió el caso D.H. y otros (2007) con información estadística que mostraba que el 56% de todos los estudiantes asignados a escuelas especiales en Ostarva eran gitanos. Es decir, datos estadísticos que comparaban la proporción de estudiantes con una característica y sin ella.

una probabilidad subjetiva que se traduce en el grado de creencia racional que una persona tiene en la verdad de la proposición dada cierta prueba. La probabilidad subjetiva se apoya en una fórmula matemática conocida como el Teorema de Bayes, que tiene la siguiente forma:

$$P(H/E) = P(E/H) \times P(H) / P(E/\text{no-H})$$

Esta fórmula dice que la probabilidad condicional de que sea verdadera la hipótesis H dada una evidencia E es igual a la probabilidad de que se dé E si es verdadera H multiplicado por la probabilidad de H (sin tomar en cuenta E), dividido por la probabilidad de que se dé E si no es verdadera H.

Como dice Ferrer Beltrán (2007: 110), “lo que el método de cálculo nos dice es qué grado de creencia es racional tener en la hipótesis dado un elemento de juicio si antes de conocer el elemento de juicio teníamos un grado de creencia x en ella y dadas las probabilidades inversas de que esté presente el elemento de juicio si la hipótesis es verdadera y si no lo es.” El punto más problemático para aplicar esta concepción de la probabilidad al ámbito jurídico es la identificación de las llamadas probabilidades *a priori*.<sup>30</sup> En algunos contextos es posible obtener esas probabilidades de forma estadística;<sup>31</sup> pero no en todos los contextos esto es posible, por ejemplo, cuando se trata de decidir judicialmente si un hecho individual ha tenido lugar, pues esas probabilidades *a priori* solo

---

<sup>30</sup> En el ámbito del razonamiento jurídico probatorio, no obstante, hay varios bayesianistas. Puede verse, por ejemplo, Aitken, Taroni, y Bozza (2022) y, en general, el número de debate al trabajo de Mikel Pardo y Ronald Allen, “Relative plausibility and its critics” en *The International Journal of Evidence and Proof*, vol. 23, no. 1 y 2.

<sup>31</sup> En algunas pruebas provenientes de la ciencia forense, por ejemplo, se habla de la *likelihood ratio* (LR, por sus siglas en inglés) o relación de verosimilitudes: “es la relación -división- entre las probabilidades de dos hipótesis alternativas: si el sospechoso es el autor de la muestra dubitada, en caso de encontrarse el match entre ambos perfiles la probabilidad será igual a 1 ..., respecto -dividido por- a la probabilidad de que se match provenga de cualquier otro individuo dentro de la población de referencia. ... esa valoración probabilística se calcula a partir de bases de datos poblacionales conocidas, eliminando la subjetividad de la persona experta a la hora de valorar la fuerza -o peso- de la evidencia. Eso no quita que la persona experta tengo que tomar decisiones, por ejemplo, sobre qué población de referencia usar, o cómo muestrear una población de referencia para hacer las estimaciones probabilísticas necesarias, pero de hecho las decisiones se documentan y se pueden discutir y mejorar, y, en caso de cambio en los datos de referencia, se puede recalculan el peso de la evidencia aportada.” (González Rodríguez, 2022).



podrían depender de las creencias o la certeza de cada juez<sup>32</sup>. No solo enfrentaríamos el problema de la conformación de esas creencias de cada juez y su posible variación ante el mismo caso, sino que, además, hay que atribuirle un valor mayor de 0 y menor de 1, es decir, traducirlo numéricamente. Lo que ya en sí mismo resulta incompatible con la presunción de inocencia, pues cualquier número mayor a 0 supondría que el juez *a priori* de las pruebas estaría atribuyendo a la hipótesis de la culpabilidad una probabilidad positiva.

En todo caso, que la probabilidad estadística y la probabilidad subjetiva no sean útiles para la declaración de hechos probados en el ámbito jurídico, por supuesto, nada dice sobre su utilidad y éxito no solo en otros ámbitos, sino incluso en alguna de las pruebas específicas que se presentan en el proceso judicial que están basadas en su aplicación.

En el propio ámbito del dopaje, se identifica como un avance el empleo de redes bayesianas para determinar la probabilidad del resultado de una prueba en función de varias variables y/o factores. Así, por ejemplo, Sottas, et.al. (2008: 193):

[e]l problema de la interpretación de los parámetros sanguíneos se aborda desde la perspectiva de la estadística bayesiana... Se desarrolla una red bayesiana jerárquica empírica (BN) centrada en la detección del dopaje sanguíneo. En primer lugar, la red bayesiana jerárquica se utiliza para determinar la probabilidad de que el resultado de un análisis dependa de varias variables o factores. En segundo lugar, demostramos cómo la misma BN puede ampliarse de forma natural para analizar y clasificar una secuencia completa de parámetros sanguíneos individuales. En tercer lugar, mostramos cómo puede estimarse la prevalencia del dopaje sanguíneo para una población razonablemente grande cuando se introduce en la BN un nodo que modela el efecto del dopaje sanguíneo.

Finalmente, por lo que hace a la probabilidad lógica o inductiva, siguiendo a Cohen (1989), está vinculada con el método de inducción eliminativo, que no admite cálculo matemático. Como dice Ferrer Beltrán (2007: 122):

el hecho de que la probabilidad inductiva de tipo baconiano no admita el cálculo matemático no implica que no se pueda graduar y comparar el nivel de soporte inductivo

---

<sup>32</sup> Evidentemente, esto nos devolvería a los problemas que se plantearon en el epígrafe anterior. Pero siempre hay quien insiste en ello, como De Finetti (1970: 14): “[I]o que al final es fundamental es tener como punto de partida las probabilidades iniciales de los eventos reales, sin que importe cómo han sido evaluadas.”

con que cuenta cada hipótesis. El grado de corroboración o de soporte inductivo de las distintas hipótesis en conflicto *puede compararse*, lo que permite hacer una ordenación de las hipótesis, pero no se puede cuantificar numéricamente la probabilidad de cada una de ellas (cosa bastante intuitiva para un jurista). ... [L]o que se pretende medir no es una frecuencia relativa ni tampoco la fortaleza de nuestras creencias subjetivas en la verdad o la falsedad de una proposición, si no la fiabilidad de la inferencia que va de una proposición a otra en contextos donde no la puede garantizar la lógica deductiva.

La probabilidad como grado de apoyo inductivo permite “tan solo” una comparación entre hipótesis en competencia. Pero no es en absoluto una minucia, más bien lo contrario, si podemos identificar cuál de las explicaciones integra mejor todos los datos disponibles o la mejor explicación entre varias en competencia es algo muy valioso.<sup>33</sup> De hecho, así lo reconoce la Sala Penal del Tribunal Supremo en la STS 4466/2022 de 17 de noviembre de 2022:

Como consecuencia, y de contrario, surge la obligación de declarar no acreditada la hipótesis acusatoria cuando la prueba practicada arroja un resultado abierto. Lo que se dará cuando la hipótesis defensiva singular o la hipótesis presuntiva general de no participación que garantiza, de partida, el principio de presunción de inocencia, como regla de juicio, aparezcan, desde criterios racionales de valoración, también como probables, aun cuando lo sean en un grado menor que la tesis acusatoria.

Insistimos, el problema se centra en el diálogo entre dos hipótesis, una acusatoria y otra defensiva, pero que no parten, ni mucho menos, de las mismas exigencias de acreditación. La primera, reclama un fundamento probatorio que arroje resultados que en términos fenomenológicos resulten altísimamente concluyentes. La segunda hipótesis, la defensiva, no.

Este doble estándar responde a las diferentes funciones que cumplen. La primera, la acusatoria, está llamada a servir de fundamento a la condena y, con ella, a la privación de libertad o de derechos de una persona. Por tanto, está sometida al principio constitucional de la presunción de inocencia como regla epistémica de juicio, por lo que debe quedar acreditada más allá de toda duda razonable.

La función de la segunda, la hipótesis defensiva, es muy diferente: es la de debilitar, en su caso la conclusividad de la primera. No, de forma necesaria, excluirla.

El Tribunal Supremo identifica pues las dos hipótesis que se comparan (la acusatoria y la defensiva), acentúa el análisis del fundamento probatorio de cada una de ellas y, luego, el umbral de suficiencia a la que cada una de ellas está sometida. Esto último, desde luego, recalca que las premisas confieren más o menos apoyo o corroboración a la conclusión.<sup>34</sup> Así, veamos algunas cuestiones

---

<sup>33</sup> Como bien dice Letelier (2018: 211): “Todo incumplimiento es un evento pasado con referencia al momento en que se realiza el juicio de desaprobación sancionatorio. Siempre, entonces, es necesaria la reconstrucción de los hechos que constituyen la infracción para que de ello pueda seguirse un proceso de imputación institucional de la correspondiente infracción. La reconstrucción es, así, un proceso de comprobación. Mediante diversas formas hacemos plausible una historia que narra cómo los hechos –que acreditan o desacreditan la infracción– efectivamente ocurrieron.”

<sup>34</sup> Por eso, uno de los significados que se le pueden atribuir a la palabra “prueba” es la actividad de corroborar la veracidad de un enunciado fáctico en un proceso judicial.

sobre los temas que se deslizan de la STS antes citada que tendrán impacto en el aumento o disminución de la corroboración o apoyo a las hipótesis que se comparan. Siguiendo a González Lagier (2022: 384-395)<sup>35</sup>:

- a. Sobre las hipótesis a comparar habría que tomar en consideración la *capacidad explicativa* que tienen al integrar los elementos de juicio con que se cuentan; su *simplicidad*, que se traduce en preferir hipótesis “que nos exigen asumir menos presupuestos no probados que aquellas cuya corrección requiere un número mayor de datos con los no se cuenta”; su *coherencia* narrativa; que *no haya sido refutada*, es decir, que no se haya mostrado un hecho directa o indirectamente incompatible con la hipótesis; que se hayan confirmado las *predicciones* que se deriven de la hipótesis principal; y, desde luego, que se hayan *descartado* otras hipótesis alternativas sobre los mismos hechos. El mayor o menor cumplimiento de estos criterios aumentará o disminuirá la confirmación de la hipótesis.
- b. Respecto el fundamento probatorio o las diversas pruebas que hayan sido admitidas y practicadas, la corroboración exige atender a su *cantidad*, pues un mayor número de pruebas será más favorable; su *variedad*, no basta con tener muchas pruebas, sino que una situación ideal es cuando hay distintos tipos de pruebas; y, por supuesto, es indispensable tomar en consideración su relevancia y su *fiabilidad*; finalmente, el número de pasos inferenciales que separan la hipótesis de las diversas pruebas. Nuevamente, una hipótesis estará más confirmada cuánto más satisfagan estos criterios las pruebas que para ello se tienen.

La relevancia y la fiabilidad de los elementos de juicio tienen una gran importancia: el primero para poder discernir qué información sobre el caso puede

---

<sup>35</sup> Haack (2020: 301) al hablar sobre el aval de una conclusión, identifica más o menos los mismos elementos relevantes sobre la justificación epistémica “(o, como lo pondríamos de manera más idiomática, cuán probable hacen las pruebas que la conclusión sea verdadera) depende de tres factores: i) cuán fuerte es la conexión entre las pruebas y la conclusión: grado de apoyo; ii) cuán sólidas son las pruebas en sí mismas, independientemente de la conclusión: seguridad independiente; iii) Cuántas de las pruebas relevantes están incluidas: comprehensividad.”

Gascón Abellán (2004: 178-187) alude a: a. La fiabilidad de las generalizaciones o máximas de experiencia empleadas; b. La fiabilidad de las pruebas o información utilizadas; c. La cantidad y variedad de información; d. El número de pasos inferenciales que separan la hipótesis de las pruebas; e. El número de hipótesis alternativas posibles que expliquen los mismos datos y su grado de probabilidad.

ser considerada como prueba; el segundo, para atender a la calidad de la información disponible. Dado que los factores a considerar para valorar la calidad de las pruebas dependen del tipo de información en juego, dejaré esta cuestión para otro epígrafe, donde abordaré la fiabilidad del conocimiento experto.<sup>36</sup>

Por otro lado, aunque en el párrafo citado de la STS 4466/2022 no se haga mención explícita a las conexiones o enlaces empleados cuando relacionamos las proposiciones en juego, es indispensable entrar en este punto. Seguramente, será fácil ver su importancia si acudimos a la idea de “la sana crítica”<sup>37</sup>, introducida en la Ley de Enjuiciamiento Civil española en el 2000 y luego desarrollada por la jurisprudencia como “compuesta por las máximas de la experiencia, el conocimiento científico y las reglas de la lógica”<sup>38</sup>. Tanto las máximas de la experiencia como el conocimiento proposicional desarrollado por la ciencia son *generalizaciones* que parten de casos particulares para llegar a una conclusión mediante un razonamiento inductivo. El punto relevante para que una generalización sirva como enlace entre dos proposiciones es su fundamento cognoscitivo, que dependerá: de la cantidad de ejemplos, casos u observaciones a partir de los cuales se hace la generalización; de cuál es la fiabilidad de estos ejemplos; de si son o no una muestra representativa del tema objeto de la generalización; y de si hay contraejemplos (González Lagier, 2022: 389).

Como puede verse, decir que el razonamiento que concluye con la determinación judicial que da por probados ciertos hechos es de carácter probabilístico y que eso, además, involucra una probabilidad lógica o inductiva, exige una tarea de valoración del conjunto de pruebas que está lejos de ser sencilla. Sin embargo, contra lo que sostiene la Audiencia Nacional en la sentencia que da pie a este trabajo, asumir el carácter probabilístico inductivo del razonamiento probatorio y

---

<sup>36</sup> Sobre la relevancia de la prueba como criterio de admisión de todo tipo de elemento de juicio, véase Vázquez (2022b: 152 ss).

<sup>37</sup> La sana crítica, se dice, constituye “el camino del discurrir humano que ha de seguirse para valorar sin voluntarismos ni arbitrariedades los datos suministrados por las pruebas” (SAP Guipúzcoa, de 15 de mayo de 2006). En esa misma línea, Montero Aroca (2006: 434), ha considerado que “frente a otros ordenamientos en los que, como reacción a la prueba legal, se pone el acento en la libertad del juzgador, el español pone el énfasis en la racionalidad que ha de estar en la base de la valoración.

<sup>38</sup> Sobre las máximas de la experiencia y su fundamento cognoscitivo, puede verse Limardo (2021) y Taruffo (2023).

adoptar una concepción racional de la prueba (que parte precisamente de esa asunción) es la única posición compatible con la presunción de inocencia y la exigencia de motivación de las decisiones judiciales entendida como *justificación* de estas<sup>39</sup>. La certeza psicológica de los juzgadores es un elemento subjetivo que varía de juez en juez, nada dice sobre la verdad de los enunciados sobre los hechos y tampoco sobre cuán corroborados estén por las pruebas presentadas. La certeza racional sobre los hechos, en cambio, es inalcanzable, puesto que cualquier conclusión probatoria resulta falible y ningún conjunto de pruebas nos permite justificar conclusiones ciertas. Queda, pues, asumir el carácter gradual, probabilístico, del apoyo inductivo que un conjunto de pruebas puede dar a una conclusión sobre los hechos y preguntarse cuán fuerte debe ser esa corroboración para poder considerar probados los hechos. Ahora bien, aunque se haga todo este trabajo justificativo y se tenga un conjunto rico de pruebas relevantes y fiables, siempre hay un espacio para el error. Como también reconoce la Sala Penal del Tribunal Supremo en la STS 4466/2022 de 17 de noviembre de 2022:

Toda reconstrucción probatoria arroja sombras de dudas, espacios fácticos que resultan de imposible reproducción. Pero la cuestión esencial reside en determinar si dichas incertezas impiden a los jueces justificar de forma cognitiva la hipótesis acusatoria, ya sea por ausencia de prueba sobre elementos fácticos esenciales sobre los que aquella se apoya, porque los medios utilizados para ello vienen afectados de un racional déficit de habilidad reconstructiva, porque se acredite que lo relatado es subjetivamente inverosímil, porque, a la luz de las otras pruebas, resulta fenomenológicamente imposible o poco probable o porque susciten una duda razonable.

Por supuesto, después de todo lo que se ha visto, la duda no puede interpretarse como un estado psicológico del juzgador y tampoco como la otra cara de la certeza racional sobre los hechos del mundo. Si fuera así, dado que la certeza racional no puede ser alcanzada, la duda tendría siempre su espacio, lo que conllevaría necesariamente la absolucón y la inutilidad del derecho sustantivo de que se trate, porque nunca podría ser aplicada la sanción prevista (Ferrer Beltrán, 2021: 147). Como dice Dei Vecchi (2020):

---

<sup>39</sup> En el ámbito de la motivación, es indispensable distinguir entre “explicar”, que responde a la pregunta, ¿por qué x?, y “justificar”, que respondería a ¿es correcto x? Así, explicar una decisión supone indicar que fue lo que causó esa decisión (circunstancias, motivos, etc.); mientras que justificar una decisión, en cambio, exige dar cuenta de las razones de esta (González Lagier, 2022: 360 y 361). Sobre este tema, véase también Ferrer Beltrán (2021: 187ss).

Es por esto que a menudo se dice que la justificación epistémica es falible en todo ámbito en las ciencias, en el diagnóstico clínico, en la historia, en la investigación de crímenes y, naturalmente, en el marco de las decisiones judiciales. Por más que contemos con una robusta justificación epistémica, con innumerables observaciones y con máximas de experiencia bien sustentadas en ellas, la proposición aseverada, en favor de la cual esa justificación se alega, puede ser falsa.

En el siguiente epígrafe se hablará sobre la falibilidad de nuestro conocimiento, haciendo referencia concretamente al conocimiento científico, ahora vale la pena contextualizar la fortaleza exigible sobre la inducción o, como refiere la STS 4466/2022, el estándar, los resultados altísimamente concluyentes o la acreditación más allá de toda duda razonable. Como veíamos líneas arriba, el derecho pretende asegurar que, en al menos algunas decisiones, como la de condenar en el ámbito penal, las hipótesis tengan un grado de apoyo especialmente alto. Cuán elevado debe ser ese apoyo depende de un juicio político-moral relativo a la importancia de los valores en juego.<sup>40</sup> Para este objetivo se tienen *los estándares de prueba*, que son las reglas que establecen el umbral de exigencia probatoria, con las que se busca distribuir entre las partes el riesgo de error (por ejemplo, falsas condenas y falsas absoluciones)<sup>41</sup>. Así, un estándar de prueba muy exigente hace que el riesgo de condenas falsas sea bajo, porque es difícil que se disponga de pruebas tan contundentes que hagan que la hipótesis de la culpabilidad de un inocente supere el estándar; y, en cambio, también hace que el riesgo de absolver a culpables sea alto, dado que también puede ser muchas veces complicado reunir pruebas de la culpabilidad de quien sí cometió los hechos que permitan satisfacer ese estándar exigente. Por el contrario, un estándar de prueba menos exigente disminuirá el riesgo de las absoluciones de personas culpables, pero aumentando el riesgo de condenas falsas. Dado que decidir el umbral de exigencia probatoria en el que se sitúa el estándar supone decidir qué reparto del riesgo del error se estima aceptable, esta es claramente una decisión político-moral.

---

<sup>40</sup> Aun cuando es una cuestión de carácter epistemológico el cómo se construye ese tipo de reglas de decisión.

<sup>41</sup> Los estándares de prueba sirven también como criterios de justificación y como garantía para las partes, al respecto véase Ferrer Beltrán (2021).

En materia penal, ahora es común decir que el estándar de prueba aplicable para la decisión final en sentencia es el más allá de toda duda razonable y considerar que se trata de un estándar muy exigente<sup>42</sup>; entonces, no basta con que la hipótesis de la culpabilidad en el proceso penal sea más probable que la hipótesis de la inocencia, sino que debe contar con un grado de corroboración más elevado, que suele identificarse con la corroboración más allá de toda duda razonable. En cambio, al menos en el ámbito anglosajón, en materia civil se suele aludir al estándar de la prueba prevaleciente, que se considera menos exigente.

La probabilidad es, por definición, gradual, de modo que no basta con decir “que haya una probabilidad” de que sucedió determinado hecho, sino que el punto crucial es determinar, precisamente, cuál es la probabilidad *suficiente* (Ferrer Beltrán, 2022: 432). El carácter político-moral de la determinación del umbral de suficiencia probatoria, i.e, del estándar de prueba, supone que las razones para fijarlo en un punto u otro de la escala probabilística serán también de carácter político-moral. Para ello, Jordi Ferrer (2021: 150 ss) ha propuesto atender a una serie de cuestiones, sin que puedan entenderse como una lista cerrada:

- Cuán grave nos parezca como sociedad el error en caso de condena falsa
- Cuán grave nos parezca como sociedad el error en caso de absolución falsa
- Las dificultades probatorias de los tipos de casos
- El impacto en la distribución del riesgo del error de otras reglas (como las presunciones o la carga de la prueba)
- La distribución de poder entre las partes
- La etapa del proceso en que debe adoptarse la decisión para la que se regula el estándar de prueba.

Si esto es así, resulta claro que nada dice que el estándar de prueba aplicable en una jurisdicción y en otra deba ser el mismo, o en un tipo de casos y en otro, o en una decisión que finalice un proceso y en otra intermedia, Los bienes

---

<sup>42</sup> Sobre las dudas que plantea este estándar, véase Laudan (2006: 59 ss).

afectados y la gravedad de su afectación en caso de decisiones erróneas serán distintos y, con ello, también es razonable que lo sean los estándares de prueba aplicables. Así pues, dado que la determinación del estándar de prueba es variable, dependiente de la jurisdicción de que se trate o incluso de un tipo de casos en la misma jurisdicción, dejaremos para el último epígrafe de este informe un análisis más detallado sobre la distribución del riesgo del error en materia de dopaje.

Llegados a este punto, puede fácilmente concluirse que la Audiencia Nacional desconoce la sofisticación teórica y práctica que supone afirmar que determinados hechos considerados probados son lo *suficientemente probables*. Ello no es baladí, pues una correcta motivación de su decisión exigiría justificar adecuadamente qué es exactamente lo que confiere más o menos apoyo a cada una de las dos hipótesis que en el caso se comparaban (el deportista se dopó vs. el deportista no se dopó). La motivación de la decisión sobre los hechos en un proceso judicial exige argumentar la fiabilidad de las pruebas empleadas en el caso, justificar el grado de corroboración de las distintas hipótesis que esas pruebas aportan en su conjunto y justificar que ello permite satisfacer el estándar de prueba aplicable. En cambio, desconociendo el carácter probabilístico del razonamiento probatorio y la naturaleza falible del conocimiento científico, la Audiencia Nacional se limitó a señalar que no se le estaba ofreciendo algo imposible: certezas científicas.

#### **IV. La falibilidad del conocimiento científico y su impacto en las decisiones judiciales**

En los últimos años se ha incrementado el uso del conocimiento experto en los procesos judiciales, tanto en la cantidad de pruebas periciales que se ofrecen como en las aplicaciones forenses de teorías o métodos surgidos en otros contextos, como el clínico. Quizá por ese incremento, también en los últimos años hay mayor preocupación por la calidad del conocimiento experto que llega hasta los procesos judiciales. En esa línea se ha planteado, por ejemplo, la



necesidad de distinguir las “pruebas periciales” de las “pruebas científicas”<sup>43</sup> o de usar el carácter “científico” como criterio de calidad de esta prueba<sup>44</sup>. La idea de fondo parece ser que lo científico es sinónimo de calidad, de fiabilidad, de certeza, mientras no lo sería otro tipo de conocimiento o de investigación. Se desconoce con ello que el conocimiento científico también es *fallible*.

La falibilidad del conocimiento científico está directamente relacionada con las diversas limitaciones epistémicas que se enfrentan, tanto respecto lo que sabemos sobre cómo funciona el mundo, como sobre aquello que desconocemos<sup>45</sup>. Así, las proposiciones científicas que ahora se consideran verdaderas, mañana podrían demostrarse falsas a la luz de experimentos posteriores, de avances teóricos o de innovaciones tecnológicas. Por ello, la investigación científica supone (y exige) una disposición para cuestionar y una apertura para reformular lo que se tiene como verdadero.

Cuando el conocimiento científico, además, se lleva al proceso judicial para la resolución de un caso las cuestiones se complican<sup>46</sup>. Más allá de las tensiones

---

<sup>43</sup> El Anteproyecto de la Ley de Enjuiciamiento Criminal, por ejemplo, distingue las pruebas periciales de las pruebas científicas. En la exposición de motivos (p. 53) se dice: “Una de ellas es el régimen singularizado de las pericias que pretenden aportar conocimientos que, por estar basados en métodos empíricos dotados de una especial fiabilidad en el seno de la comunidad científica, aportan al proceso conclusiones de especial valor. A pesar de que el concepto mismo de ciencia, especialmente cuando se aplica a la elaboración de una prueba, resulta discutible, se suele aludir a este tipo de diligencia como informe pericial científico o, más genéricamente, como prueba científica.”

<sup>44</sup> A nivel jurisprudencial, la Audiencia Provincial de Barcelona, Sección 14a, en la muy citada sentencia 181/2006 del 24 de marzo, hace referencia al “nuevo paradigma de la prueba pericial”, afirma que éste se caracteriza, entre otras notas, por el hecho de que el papel del juez se ha de centrar en *el control de la científicidad del proceso de investigación* llevado a cabo por el perito. Por otro lado, entre la dogmática española, Pérez Gil (2010:46), afirma que “[e]n el caso del proceso civil [la jurisprudencia española] viene invariablemente eligiendo el término ‘prueba científica’ cuando una actividad probatoria, casi siempre pericial, se ha desarrollado empleando tecnologías que arrojan resultados ampliamente fiables”. O Corda (2013: 116) sostiene que “[al juez] no se le requiere –sería imposible– transformarse en un científico, sino más bien dotarse de los instrumentos metodológicos adecuados a los fines de ponerse en situación de valorar la tasa de científicidad en sentido estricto de la técnica probatoria de que se trate”.

<sup>45</sup> Incluido aquello que desconocemos que no sabemos.

<sup>46</sup> Haack (2020: 158-159) nos dice que: “hay profundas tensiones entre los objetivos y los valores de la empresa científica y la cultura jurídica...: o entre el carácter investigador de la ciencia en la cultura adversarial de nuestro sistema jurídico; entre el objetivo de la investigación científica de cuestiones generales y el objetivo jurídico de resolver casos particulares; entre el omnipresente falibilismo de las ciencias -su apertura la revisión a la luz de nuevas pruebas- y la preocupación del derecho por resoluciones céleres y definitivas; entre el impulso científico por la innovación y la preocupación del sistema jurídico por el precedente; entre el pragmatismo informal, orientado a la solución de problemas, de la investigación científica y la dependencia del sistema jurídico a

que pueden darse debido a las características tan diversas de la investigación científica y la dinámica del proceso judicial, no siempre se puede allegar lo mejor del conocimiento científico (a veces, por ejemplo, incluso por falta de recursos) y, no obstante, a veces pareciera asumirse lo contrario, cuando se exige a las pruebas científicas certeza o se les dota de un halo de infalibilidad. Pero la realidad es más tozuda que nuestras expectativas y, por ello, vale la pena acudir a un conjunto de casos reales, identificados como los “bebés sacudidos”, pues ellos nos ejemplifican de manera nítida la falibilidad del conocimiento experto y el impacto de ello en los procesos judiciales.<sup>47</sup> Una de las peligrosas consecuencias de asumir que ciertas proposiciones científicas son incuestionables es, quizá, que se deja de lado la preocupación sobre si el sistema jurídico está funcionando adecuadamente o no (Tuerkheimer, 2014: xii).

Los casos de los “bebés sacudidos” tuvieron lugar fundamentalmente en el Reino Unido y en Estados Unidos,<sup>48</sup> en los que cientos de ciudadanos fueron sancionados civil o penalmente por considerarse que infligieron daños a niños menores de tres años, en algunos casos incluso causándole la muerte<sup>49</sup>. En muchos de estos procesos judiciales solo había pruebas periciales de carácter médico,<sup>50</sup> a las que subyacía una teoría: si un menor de tres años presenta cierto tipo de daños, con tres concretos síntomas neurológicos, la única explicación causal posible sería que fue sometido a una sacudida violenta.<sup>51</sup> Se le llamó el Síndrome del Bebé Sacudido (SBS, por sus siglas en inglés) y los tres síntomas que en teoría lo componían eran encefalopatía, hemorragia subdural y hemorragias retinianas, conocidos como “la triada”. Además, según lo que

---

las reglas y procedimientos formales; y entre las aspiraciones y esencialmente teóricas de la ciencia y la inevitable orientación del sistema jurídico hacia las políticas públicas”.

<sup>47</sup> Son casos significativos, además, porque se basaron en conocimiento e investigación de carácter médico, como la que tiene lugar en el ámbito del dopaje.

<sup>48</sup> También se conocen casos en otros países, como Colombia. Véase la SP1344-2022, Radicación No. 51710, resuelto por la Sala de Casación Penal de la Corte Suprema de Justicia de Colombia.

<sup>49</sup> Se tiene registro también de su incidencia en casos sobre custodia de los hijos en materia familiar.

<sup>50</sup> No necesariamente el testimonio experto de *un* solo perito, sino, por ejemplo, de seis médicos con diferentes especialidades. Así, un internista pediátrico, un neurólogo pediátrico, un oftalmólogo pediátrico y pediatras -cfr. *State v. McClary*, 541 A.2d 96, 102 (Conn. 1988)-, incentivando la idea de que el diagnóstico de la sacudida violenta gozada de la aceptación general de los médicos.

<sup>51</sup> A no ser que hubiese sido víctima de un aparatoso accidente vehicular o que hubiera sufrido una caída de un edificio.

entonces se sabía, era relativamente fácil identificar al perpetrador porque el niño perdía la conciencia de forma inmediatamente posterior a la sacudida violenta.

La hipótesis sobre la relación entre la triada y el SBS surgió en 1971, planteada por el neurocirujano pediátrico A. Norman Guthkelch al estudiar 13 casos de niños con hemorragia subdural y ver que 5 de ellos no tenían signos de violencia. Luego, en Estados Unidos se publicaron un par de trabajos por un pediatra radiólogo que había analizado 27 casos en los que alertaba que ciertos juegos comunes entre las personas con los infantes *podrían* causarles hemorragias subdurales y, por ello, hacía un llamado a alertar de esto a la población. Así, en la década de 1970 y 1980 se fue gestando en la medicina el SBS, hasta que, en 1993, la Academia estadounidense de pediatría estableció una suerte de presunción médica afirmando que “los datos que se tenían sobre la naturaleza y la frecuencia del trauma cerebral la apoyaban”.

No obstante, el conocimiento sobre el SBS cambió radicalmente cuando se sometió a un escrutinio severo los estudios publicados y la literatura médica sobre el síndrome del bebé sacudido. Según Donohoe (1998), se encontraron lagunas de información (v.gr. una base de datos muy pequeña), errores lógicos (por ejemplo, se identificaron como casos del SBS la presencia de hemorragias retinales y subdurales, asumiendo lo que en realidad debían mostrar) e inconsistencias en la definición de los casos (entre otras cosas, algunos estudios consideraron la confesión del supuesto agresor para establecer que un caso determinado era SBS).

La medicina en la época en que surgió la teoría del SBS funcionaba de una manera muy distinta respecto como puede funcionar actualmente, luego de que a finales de los años 90's se difundiera la llamada “medicina basada en la evidencia” (EBM, por sus siglas en inglés)<sup>52</sup>. La EBM defiende fundamentalmente que el ejercicio de la medicina no puede basarse exclusiva o

---

<sup>52</sup> Con esa idea, se han desarrollado no solo diferentes escalas que señalan el valor de la evidencia, sino también han ido surgiendo cierto tipo de documentos de consenso que son cada vez más utilizados en la práctica médica. Dada esa gran diversidad, hay varias complicaciones para identificar adecuadamente las mejores pruebas disponibles.

prioritariamente en la *expertise* de cada médico ni en el mero consenso de grupos médicos, sino en pruebas provenientes de estudios empíricos a partir de los cuales se pueda inferir la eficacia y seguridad de un tratamiento, etc.<sup>53</sup> Y, por supuesto, conviene recordar que la experiencia clínica no es investigación científica ni desde un punto de vista experimental y tampoco metodológico (Roberts, 2013: 172). La experiencia clínica de un médico es una fuente de conocimiento muy limitada,<sup>54</sup> tanto por el número de pacientes que puede ver como la forma subjetiva de evaluar el éxito de sus diagnósticos o tratamientos<sup>55</sup>. Roberts (2013: 167), por ejemplo, cita el caso *Henderson* en el Reino Unido, donde el argumento fundamental del perito fue que en sus veinticinco años ejerciendo la oftalmología pediátrica solo había visto pliegues retinianos en casos de sacudidas de infantes o de sacudidas más golpes. Uno se debería preguntar ante una afirmación así: ¿cómo puede saber el perito (o cualquier otra persona) que sus observaciones de esos 25 años fueron correctas y que esas lesiones no tenían una explicación alternativa que no fue refutada porque ni siquiera se consideró?<sup>56</sup>

---

<sup>53</sup> En 1997, Sackett definió la *Evidence-Based Medicine* como el “*conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*”. Y, desde entonces, el debate en el área ha sido intenso, en buena medida sobre la idea de “*best evidence*”, muchas veces argumentándose que la medicina *siempre* se ha basado en las mejores pruebas disponibles. ¿Dónde estaría entonces la supuesta diferencia? Fundamentalmente en que para los seguidores de la EBM la práctica del diagnóstico y la intervención médica posterior debería estar guiada por *investigación biomédica sistemática* (Martini, 2020). Y, como es bien sabido, en cambio, la práctica de la medicina, incluso tradicionalmente considerada como *un arte*, ha dependido en mayor medida de la intuición, la experiencia, la tradición, la autoridad y el buen juicio del galeno.

<sup>54</sup> Según Howick (2011: 4), “pronto los proponentes de la EBM mostraron que muchas terapias empleadas extensamente y adoptadas solo con base en las formas de prueba últimas en la jerarquía, se demostraron inútiles o dañinas cuando fueron sometidas a ensayos controlados”.

<sup>55</sup> El debate, además, se vio enfatizado por las investigaciones sobre los sesgos cognitivos en el razonamiento humano, realizadas fuera del ámbito médico, pero con influencia en él. Lo que es sumamente relevante si consideramos, como dice González Garay (2014: 43), que “en la práctica médica existe una tendencia a aplicar de manera automática el conocimiento adquirido con gran reticencia a utilizar intervenciones basadas en nuevos conocimientos”. Precisamente por ello se resalta que “la MBE [o EBM] permite reemplazar las intervenciones médicas aceptadas por otras más potentes, más exactas, más eficaces y más seguras” (Cfr. Villanueva y Lezana, 2017: 153).

<sup>56</sup> A diferencia del juicio de un experto, los ensayos controlados a los que invita la EBM reducen los sesgos del experimentador, por un lado, al hacer la división de los participantes de manera aleatoria y, por otro lado, al organizar el experimento de manera ciega para todos los participantes de este, incluidos por supuesto los investigadores principales y demás colaboradores. Precisamente, los ensayos controlados ciegos son considerados el estándar de oro de los estudios empíricos clínicos; sin embargo, no están exentos de aspectos problemáticos como el tamaño de la muestra analizada, la interpretación de los datos que se obtienen y, por supuesto, diversas cuestiones éticas relacionadas con la investigación en humanos, el tiempo y los recursos que suponen.

Como dice Haack (2021: 16):

[t]oda investigación empírica depende de la experiencia; pero uno de los modos en que la investigación científica es a menudo 'algo más' reside en que la experiencia en la que ella depende no es una carencia de auxilios, si no una experiencia habilitada y mediada por instrumentos de todo tipo; no carece de corroboración, sino que está abierta al escrutinio de otros en la comunidad, y no está librada al azar o a la mera ventura, sino que es deliberada, articulada, controlada.

Los instrumentos de observación son una de esas grandes ayudas y, por supuesto, en el ámbito médico las imágenes que se pueden tener del cuerpo humano para el diagnóstico han sido pieza clave. De hecho, la aparición y evolución de la resonancia magnética contribuyó a mostrar que la triada podía manifestarse como consecuencia de lesiones accidentales o desórdenes médicos (Tuerkheimer, 2014: 19). Es decir, el desarrollo tecnológico permitió mostrar que las observaciones del perito del caso *Henderson* eran altamente falibles debido a que no había considerado, y mucho menos descartado, esas explicaciones también plausibles.<sup>57</sup>

Los avances en las técnicas y métodos que se emplean han tenido también un impacto jurídico procesal importantísimo, al punto que ahora se habla de las “exoneraciones del ADN” para aludir a aquellos casos en los que se ha identificado una condena errónea, por ejemplo, mediante un reanálisis de los vestigios encontrados en la escena del delito con una tecnología moderna que es más sensible y más fiable que la usada cuando se tomó la decisión. El caso de Rafael Ricardi en España muestra precisamente esto, pues inicialmente fue condenado con un informe de ADN en 1995 y posteriormente, liberado como consecuencia del avance tecnológico que permitió medir más marcadores genéticos de los que se habían medido. Este reanálisis resultó en el descarte del perfil genético de Ricardi y, por tanto, en su absolución.<sup>58</sup>

---

<sup>57</sup> Como se dijo en la decisión del caso *Henderson*: “El descubrimiento de un caso de pliegues retinianos debido a leucemia y, por tanto, sin causa traumática alguna, enfatiza la importancia de reconocer los límites del conocimiento médico en un tiempo determinado y la necesidad de apreciar que aquello que no ha sido previamente contemplado puede, no obstante, ocurrir. Debemos reconocer los límites de la ciencia médica y en particular que puede haber eventos, muertes o síntomas que no tienen explicación...”

<sup>58</sup> La sentencia condenatoria fue dictada por la Sección Quinta de la Audiencia Provincial de Cádiz el 15 de octubre de 1996 (Rollo nº 27/1995) y la Sentencia del Tribunal Supremo que da lugar al recurso de revisión es la 792/2009 de 16 de Julio de 2009.

La falibilidad de la ciencia no solo radica en el desarrollo tecnológico que potencia o limita sus investigaciones, también en que la imagen del mundo que tenemos no es completa, es decir, hay todavía cuestiones relevantes sobre las que habría que teorizar o factores a considerar. El caso *Henderson* nos vuelve a ser útil aquí, pues la Corte inglesa reconoce esta incertidumbre científica y, con ella, las limitaciones epistémicas de las pruebas científicas que lejos están de suministrar respuestas concluyentes:

enfaticamos que dentro del ámbito del conocimiento médico actual, nos estamos enfrentando a causas naturales de muerte, porque en esta apelación se nos recuerda claramente que en ningún momento en un área como esta se puede considerar que el conocimiento es completo y comprensivo. Hay límites para el alcance del conocimiento y no se puede obtener ninguna conclusión sin reconocer la posibilidad de que surja una causa desconocida a la luz de la percepción médica y que la mera exclusión de toda posible causa conocida no prueba que se haya infligido violencia en forma deliberada.

En esa línea, muy recientemente ha vuelto a acaparar la atención pública el caso de Kathleen Foldigg, una australiana condenada por matar a sus cuatro bebés, por lo que ha estado 20 años en prisión.<sup>59</sup> Gracias a los avances científicos,<sup>60</sup> se *sabe ahora* que hay una posible explicación alternativa de la muerte de los bebés: pudieron morir por causas naturales debido a una mutación genética que se ha identificado recientemente en las nuevas investigaciones y que se encontró en dos de sus hijos. Esa mutación les podría haber provocado una muerte súbita mientras dormían y no haber sido privados de la vida por su madre. No tendremos certeza de que la Sra. Foldigg es inocente, pero ha sido claro que no hay pruebas suficientes en su contra cuando se han presentado pruebas periciales que pueden explicar sólidamente la muerte de al menos de dos de sus hijos. Precisamente por ello, Kathleen Foldigg ha sido indultada el 5 de junio de este año.

---

<sup>59</sup> La prueba "más sólida" en contra de la Sra. Foldigg fue una prueba pericial en la que un médico declaró: (i) que la probabilidad de que fueran muertes naturales era de 1 entre 73 millones; y, en la que se enfatizó, además que (ii) una muerte súbita es una tragedia, dos son sospechosas y tres son asesinato. La afirmación (i) recoge un dato estadístico que el médico nunca fundamentó en estudios, básicamente sacó los números de la chistera; y la (ii) es claramente una opinión netamente subjetiva y que, además, sobrepasa las funciones periciales.

<sup>60</sup> Cuya autoría principal es de la inmunóloga española Carola García de Vinuesa.

La ciencia es, pues, falible y, por tanto, de ninguna manera podemos exigirle certezas, pero sí la validez y fiabilidad de sus métodos, técnicas, teorías, hipótesis, etc. La experiencia comparada nos ha mostrado también cuál es el camino que debemos transitar para aumentar nuestras probabilidades de éxito con el uso de conocimiento experto justificado y fiable, como analizaremos en el siguiente apartado. Pero todo ello, así como su *aplicación fiable al caso concreto*, debe ser valorado por el juez, pues, como nos dicen Gascón, Lucena y González (2010: 5 y 6):

si los datos [obtenidos mediante pruebas científicas] dijese directamente lo que el juez necesita saber ... y el grado de probabilidad con que lo afirman fuese tan elevado que pudiera considerarse infalible, no tendría ningún sentido deferir al juez la valoración de dicho enunciado, pues esta puede ser adelantada ya por el perito de manera categórica.

Con lo dicho en este punto sobre la falibilidad del conocimiento científico, se muestra el grave error de las exigencias que la Audiencia Nacional impone como criterio en la decisión que se comenta. Por un lado, exigir certezas al conocimiento experto elimina toda posibilidad de control sobre su calidad, pues si se afirma que “no hay espacio para el error” o “que es 100% fiable”, ¿qué más habría que controlar? Y, por otro lado, debido a unas expectativas infundadas se puede estar dejando de lado conocimiento científico altamente fiable que justificaría tomar una decisión con base en él.

## **V. ¿Cómo sabemos en un proceso judicial cuál es el conocimiento científico fiable?**

La falibilidad del conocimiento científico alude a las limitaciones que enfrenta, sus potenciales cambios ante nuevos experimentos o desarrollos teóricos o avances tecnológicos; su fiabilidad, en cambio, nos da cuenta de la información que tenemos sobre su correcto funcionamiento dentro de un contexto e, incluso, en un momento determinado.<sup>61</sup> Si los jueces y juezas deben decidir, resulta

---

<sup>61</sup> Por facilidad expositiva, se emplea aquí “fiabilidad” como sinónimo de la calidad del conocimiento experto y porque los Tribunales normalmente suelen aludir a él en esos términos. No obstante, técnicamente, es una palabra que puede tener significados variados y que puede distinguirse de otros tecnicismos relevantes en el conocimiento experto. Por ejemplo, se puede distinguir entre validez y fiabilidad: la validez de un método o técnica o teoría indicaría si mide lo que pretende medir y su fiabilidad, en cambio, nos informaría sobre cuán bien lo mide.

entonces importante que de alguna manera cuenten con información sobre la fiabilidad del conocimiento proposicional usado por los peritos, sus instrumentos, etc.

La manera tradicional en que este tipo de información llega a los jueces en el proceso judicial es mediante las pruebas solicitadas por él u ofrecidas por las partes. Desafortunadamente, no todos los sistemas jurídicos se han preocupado por solicitar o incentivar que en el proceso judicial se presente información sobre la fiabilidad de los métodos, técnicas, teorías, etc., que subyacen a una prueba pericial. Una de las excepciones es el sistema estadounidense,<sup>62</sup> que es además un hito en la valoración judicial de la prueba científica a efectos de su admisión.

El caso *Daubert*<sup>63</sup> y los cuatro criterios que la *Supreme Court* estadounidense estableció en su análisis son un gran referente en el debate sobre la calidad de la prueba pericial<sup>64</sup>. Esos cuatro criterios son:

---

<sup>62</sup> Otro ejemplo al que puede aludirse es Italia, concretamente el caso *Cozzini*, Cass., sez. IV, de 17 de septiembre de 2010, no. 43786. En esa decisión, se dice que los jueces tienen que realizar una función de “garante de la científicidad del conocimiento factual expresado en el proceso” y “desplegar una penetrante función crítica, convirtiéndose en guardián del método científico”. Para cumplir esta función, la Corte hizo una lista de criterios de valoración de la prueba pericial, algunos referentes al experto y otros al contenido del informe pericial. Así: autoridad indiscutida del sujeto que realiza la investigación, independencia del sujeto que realiza la investigación, finalidades por las que se mueve el sujeto que realiza la investigación, amplitud de la investigación, rigurosidad de la investigación, la objetividad de la investigación, grado de apoyo que los hechos ofrecen a la tesis, intensidad de la discusión crítica que ha acompañado la elaboración del estudio, amplitud explicativa de la elaboración teórica y consenso que la tesis suscita en la comunidad científica. Un análisis sobre este caso y sus criterios puede verse en Igartua Salaverria (2021: 197 ss).

El Tribunal de Apelación de Nueva Zelanda, por su parte, ha dicho sobre las pruebas científicas que es indispensable que el objeto de la opinión del perito sea una rama de la ciencia suficientemente reconocida en el momento en que presta su testimonio y que la opinión de un experto no es admisible si no se tiene publicaciones de apoyo u otra verificación del ‘pedigrí’ de la opinión. (Makoare [2001] 1NZLR 318, 324).

<sup>63</sup> Se trata de un caso por responsabilidad civil contra una farmacéutica, Merrell Dow, por el fármaco Bendectin y sus alegados efectos teratógenos. El caso *Daubert* fue uno entre miles de casos que se plantearon en Estados Unidos por la misma causa. En España este tipo de casos también existió y fueron conocidos como los casos de la talidomida. En ambas latitudes se trató de un medicamento que usaban las mujeres embarazadas para paliar los mareos típicos de los primeros meses del embarazo.

No pretendo hacer un análisis exhaustivo del caso y mucho menos de la literatura que lo ha analizado en todos estos años, pues, como afirma Dwyer (2008: 342) el tratamiento teórico del caso es tan extenso que bien podría haber una carrera llamada “*Daubert studies*”.

<sup>64</sup> *Daubert v. Merrell Dow Pharms., Inc. (Daubert III)*, 509 U.S. 579 (1993).



1. Si la teoría o técnica puede ser (y ha sido) sometida a prueba, lo que constituiría un criterio que comúnmente distinguiría a la ciencia de otro tipo de actividades humanas.
2. Si la teoría o técnica empleada ha sido publicada o sujeta a la revisión por pares.
3. El rango de error conocido o posible, si se trata de una técnica científica, así como la existencia de estándares de calidad y su cumplimiento durante su práctica.
4. Y, finalmente, si la teoría o técnica cuenta con una amplia aceptación de la comunidad científica relevante.

Estos criterios, y parte del debate que tuvo la Corte estadounidense, han sido recientemente traídos a España por el Tribunal Supremo en la STS 232/2022 de 14 de marzo de 2022<sup>65</sup>:

Así, se previenen tres simples y elásticos criterios de selección: a) que la conclusión científica tenga fundamento fáctico; b) que se hayan utilizado principios y metodología fiables c) que la conclusión sea aplicable a lo sucedido de manera verificablemente correcta.

El desarrollo de estas reglas básicas permite precisar las siguientes reglas específicas: para ser calificada una determinada aserción o inferencia como conocimiento técnico-científico, debe haberse elaborado de conformidad al método técnico-científico; como presupuesto básico de la fiabilidad, *una conclusión científica ha de poder someterse a test*. El estatus científico de una teoría viene determinado por su sometimiento a procesos de refutabilidad y de control; la evaluación de la fiabilidad exige también *la explícita identificación de una comunidad científica relevante* y una expresa definición de un particular grado de aceptación interna en la misma. La aceptación difusa puede ser un factor importante para establecer la admisibilidad de una particular prueba. Sin embargo, una técnica conocida pero que disponga de un soporte mínimo en la comunidad científica puede ser vista con escepticismo; en el caso de una particular técnica, los tribunales deben considerar *la tasa conocida o potencial de error* y resistencia, y ordenar la aplicación de estándares de control de la eficacia de la técnica; ser conscientes de que las conclusiones científicas aportadas por los expertos mediante la prueba pericial adquieren, en la mayoría de los casos, un peso especial para la decisión, pero que también pueden provocar confusión y despiste debido a las dificultades para su evaluación. Por ello, el juez debe ejercitar un control mucho mayor que respecto a otros medios probatorios.

Las anteriores reglas de conformación/corroboración constituyen buenos instrumentos para que el juez pueda realizar su labor de custodia de tal manera que solo lleguen a/

---

<sup>65</sup> También han formado parte de la jurisprudencia de México o de Perú. En México la Suprema Corte de Justicia introdujo los criterios Daubert en la contradicción de tesis 154/2005-PS. Así, se dijo: “Para que un órgano jurisdiccional pueda apoyarse válidamente en una opinión de algún experto en una rama de la ciencia, es necesario que esa opinión tenga las siguientes características: 1. Que la evidencia científica sea relevante para el caso concreto de estudio, es decir, que a través de la misma pueda efectivamente conocerse la verdad de los hechos sujetos a prueba; y 2. Que la evidencia científica sea fidedigna, esto es, que se haya arribado a ella a través del método científico, para lo cual se requiere, generalmente, que la teoría o técnica científica de que se trate:

- a. Haya sido sujeta a pruebas empíricas, o sea, que la misma haya sido sujeta a pruebas de refutabilidad;
- b. Haya sido sujeta a la opinión, revisión y aceptación de la comunidad científica;
- c. Se conozca su margen de error potencial, y
- d. Existan estándares que controlen su aplicación”.

*proceso opiniones dotadas de suficientes fundamentos teóricos para producir resultados correctos y, en consecuencia, pueda excluir del cuadro probatorio aquellas opiniones científicas o técnicas basadas en conjeturas probablemente erradas, en los términos utilizados por el juez Blackburn en su voto concurrente en la sentencia Daubert. Con ello no se puede asegurar, sin embargo, que la conclusión pericial resulte irrefutable. El papel del tribunal no es controvertir ni negar el hecho irremediable de la incertidumbre técnico-científica sino manejar ese déficit epistémico de una manera institucionalmente aceptable.*<sup>66</sup> [Cursivas añadidas]

Y vale la pena también decir que estos criterios jugaron un rol importante en el actual Anteproyecto de la Ley de Enjuiciamiento Criminal, particularmente en el artículo 483 que establece las “especialidades del informe pericial científico”:

1. De ser precisos conocimientos científicos para acreditar o apreciar los hechos relevantes para la causa, el informe pericial, además de los contenidos expresados en el artículo anterior, habrá de incluir:
  - a) La acreditación del perito, la homologación del laboratorio conforme a la normativa de calidad correspondiente, así como los controles periódicos a los que se someten el laboratorio y la metodología empleada.
  - b) La identificación clara y precisa del ámbito de conocimiento sobre el que se asientan las teorías, técnicas y metodología empleadas, así como su grado de aceptación por parte de la comunidad científica de referencia.
  - c) La descripción, en lenguaje comprensible, del método aplicado para la práctica de la pericia, así como de las evaluaciones e instrumentos existentes para la comprobación de su correcta realización.
  - d) En su caso, la ratio de error o grado de falibilidad de la técnica o método aplicados, así como de los instrumentos y aparatos empleados para su ejecución.
  - e) Las conclusiones que se formulen, que deberán expresarse siempre que sea posible en términos de probabilidad.
2. Si sobre las cuestiones tratadas hubiera diversidad de opiniones en el ámbito del conocimiento científico de que se trate, se explicarán las diferencias existentes.
3. El informe no reflejará conclusiones distintas de las que se deriven directamente de los resultados obtenidos mediante la aplicación del método científico en que se base la pericia. Cuando se trate de la atribución de muestras o vestigios a una determinada persona, las conclusiones se limitarán a establecer si existe o no correspondencia y no se expresarán en términos de culpabilidad o inocencia.<sup>67</sup>

Pero, más allá del impacto de los criterios Daubert en los países de nuestro entorno cultural y, más recientemente, en España, vale la pena atender al giro que suponen. Para atribuir valor probatorio a una prueba científica, tales criterios nos exigen ir mucho más allá de cuestiones meramente relacionadas con quién

---

<sup>66</sup> Antes que la jurisprudencia, la doctrina ya aludía a ellos: véase, por ejemplo, Nieva (2010: 295) que hace referencia a la “aceptación de la comunidad científica *internacional*”. De la misma manera, la doctrina italiana hace amplia referencia a ellos: véase, por todos, Carlizzi (2019).

<sup>67</sup> En otros países también Daubert ha pasado a formar parte de su legislación. En Brasil el artículo 473 del Código Procesal Civil al hablar sobre el contenido del informe pericial alude a algunos de los criterios Daubert, como la “*indicação do método utilizado, esclarecendo-o e demonstrando ser predominantemente aceito pelos especialistas da área do conhecimento da qual se originou*”. En Colombia, el Código procesal penal, Ley 906 de 2004, en su artículo 422, al hablar de la admisibilidad de publicaciones científicas y de prueba novel, traduce los criterios Daubert.

es el perito<sup>68</sup>: hay que considerar diversos aspectos de la dinámica social de la ciencia que sirven para *la validación del conocimiento* que la conforma.

En primer lugar, vale la pena distinguir entre cuán sólidas son las pruebas que fundamentan, por ejemplo, una teoría y, por otro lado, el resultado de la presentación de esa teoría en la interacción entre expertos o científicos. Una teoría puede ser concebida, desarrollada, contrastada, refinada y modificada en el seno de un grupo de investigación o incluso en el ámbito de un único científico que considera tener pruebas sólidas sobre su corrección; pero, no basta con ello: dichas pruebas se escrutan (i.e. critican y revisan) en la interacción de las comunidades científicas con el objetivo de lograr *entre todos* resultados bien sustentados. Por eso se dice que lo explica el éxito de la ciencia y su progreso no es un método de investigación (como el popularmente identificado como “método científico”), sino las dinámicas de la investigación científica como la institución social que es.<sup>69</sup>

Si consideramos ese carácter social, ya podemos vislumbrar la importancia de algunos de los criterios Daubert, como la aceptación general de la comunidad científica o la revisión por pares y la publicación, siendo ambos una parte importante de las dinámicas (falibles) de validación del conocimiento científico. Veamos algunas cuestiones sobre estos dos criterios.

La primera vez que en el ámbito jurídico estadounidense se hace referencia a la comunidad experta es en 1923 en el caso Frye, donde el acusado ofreció como prueba pericial un entonces novedoso detector de mentiras. Esa prueba le fue excluída a Frye quien, luego de ser condenado, apeló basándose, en parte, en la impugnación de esa decisión de inadmisión. En apelación, el análisis del tribunal correspondiente se centró en el fundamento teórico subyacente al citado detector de mentiras: esto es, que las afirmaciones verdaderas serían espontáneas mientras que las mentiras requerirían un esfuerzo consciente que

---

<sup>68</sup> Por ejemplo, si es un perito oficial o un perito de parte, la experiencia que tiene un perito, etc.

<sup>69</sup> Cfr. Haack, 2021.

se reflejaría de forma particular en un aumento de la presión sistólica.<sup>70</sup> Así, en diciembre de 1923, la *Court of Appeals* del *District of Columbia* resolviendo la apelación del caso *Frye v. United States* consideró que todo esto constituía un “testimonio científico novel” y dictaminó que:

Es muy difícil detectar el momento preciso en el que un principio o descubrimiento científico cruza la línea que hay entre su etapa experimental y aquella en la que es demostrable. En algún lugar de esa zona de penumbra, el valor de la evidencia existente a su favor debe ser reconocido y, mientras que los tribunales recorren un largo camino para admitir testimonios expertos derivados de principios científicos o descubrimientos bien reconocidos, aquello de lo que estas pruebas se deducen *debe estar lo suficientemente fundado para tener la aceptación general en el área relevante*.<sup>71</sup> [Cursivas añadidas].

Es importante recalcar el hecho de que se trataba de ciencia novel y que, por tanto, de ninguna manera iba a tener la aceptación general de un área de conocimiento. Casi por definición, si es novel, no puede todavía gozar de esa consolidación. Esta exigencia se aminoró en *Daubert*, pues no solo se habla de una *amplia* aceptación (no general)<sup>72</sup>, sino que podría no gozar de dicha aceptación de la comunidad, pero *sí* haber sido sometido a revisión por pares y publicaciones.

En todo caso, como puede fácilmente advertirse, el criterio de la “aceptación general en el área relevante” constituye un giro importantísimo en los criterios de valoración de la calidad de las pruebas periciales al ir más allá del experto en sí mismo (sus credenciales, su personalidad o su actuación en el juicio)<sup>73</sup>. Esto permitiría resolver el argumento circular denunciado por Redmayne (2001: 95) consistente en que el test de la *expertise* pareciera ser la propia *expertise*, pues

---

<sup>70</sup> Según la sentencia del caso, la defensa afirmaba que: “El incremento que se produce es fácilmente detectado y distinguible del aumento producido por el simple miedo al análisis mismo. En el primer caso, la presión aumenta más que en el segundo y es más pronunciada mientras el análisis se lleva a cabo; en el primer caso, en cambio, si el sujeto está diciendo la verdad, la presión registra un aumento al inicio del análisis, pero luego baja gradualmente durante éste”.

<sup>71</sup> Véase, *Frye v. United States*, 54 App.D.C.46, 293F. 1013, 34 A.L.R. (145) (1923).

<sup>72</sup> Un nivel muy superior parece haberlo previsto el Tribunal Europeo de Derechos Humanos en *Sheffield and Horsham v. UK* (1999), un caso donde los aplicantes solicitaban al gobierno que hiciera lo necesario para que su reciente cambio de sexo (se habían sometido a la operación respectiva) se reflejara en sus diversos documentos oficiales. El Tribunal, en rechazo de la petición, argumentó que la investigación médica sobre la etiología transexual no tenía “el soporte o confirmación universal (“universal support”) de una comunidad científica”.

<sup>73</sup> Como ya puede verse, aunque los criterios *Daubert* fueron dictados para decidir la admisión o exclusión de una prueba pericial, ya en la etapa de admisión hay *cierta* valoración de la calidad de la prueba pericial.

en lugar de preguntarse si las afirmaciones del perito son verdaderas o falsas, la preocupación suele girar en torno a si el sujeto puede ser considerado “realmente” como un experto, si tiene la formación o habilidades correspondientes, si tiene los conocimientos suficientes, etc.<sup>74</sup> Y, por supuesto, también tiene un impacto importante al sacar del dominio de las certezas psicológicas a la prueba pericial, pues al menos es indispensable considerar qué dicen las comunidades expertas.

Vale la pena advertir que el criterio Frye admite al menos dos versiones, cada una con implicaciones sustantivas bien distintas. Una versión epistemológica nos exigiría enfatizar el *fundamento* de la aceptación general en el área relevante; otra versión sociológica, en cambio, pondría el énfasis en la “aceptación general” *en sí misma* en el área relevante, sin hacer necesariamente ninguna referencia a las razones que la fundamentan. Esta segunda, por ejemplo, se encuentra en Denti (1972: 280):

los métodos de investigación se consideran correctos *sólo porque* son aceptados por la generalidad de los estudiosos en un momento histórico determinado, [aunque] sin excluir que dichos métodos puedan ser considerados como erróneos en un momento sucesivo.<sup>75</sup>

No obstante, tal concepción da lugar a equívocos pues no se trata de que un “método de investigación (para usar la expresión de Denti) sea correcto por el mero hecho de que es aceptado por x; más bien, es aceptado porque se tienen pruebas de su corrección. Son, entonces esas evidencias las que producirían el conocimiento y, por ello, no puede ponerse el énfasis en el mero hecho de la aceptación, hay que preguntarse por qué la comunidad experta lo acepta. Aquí cobra relevancia aludir al informe estadounidense conocido como PCAST Report (2016)<sup>76</sup>,

---

<sup>74</sup> También llama la atención sobre este punto Edmond (2012: 37) en una crítica a la “*Law Commission’s Report “Expert Evidence and Criminal Proceedings in England and Wales”* porque, en su opinión, ésta confunde la “cualificación para hacer algo” con la fiabilidad.

<sup>75</sup> O, más recientemente, Rivera Morales (2011: 271) equipara “verdadera ciencia” con “la considerada como tal por el consenso general en la comunidad científica”.

<sup>76</sup> El PCAST es un informe presentado al presidente de los Estados Unidos de América, Barack Obama, en el año 2016, por el Consejo de Asesores del Presidente en Ciencia y Tecnología (PCAST, por sus siglas en inglés). En él se analiza la validez y fiabilidad de las pruebas periciales que se basan en la comparación de ciertas características de personas u objetos que permitirían hacer una identificación relevante en un proceso judicial. Por ello, el título del

Queremos resaltar, finalmente, que ni la experiencia, ni el buen juicio, ni las buenas prácticas profesionales (tales como programas de certificación y acreditación, protocolos estandarizados, pruebas de aptitud técnica y códigos éticos) pueden sustituir las pruebas genuinas sobre la validez de sus fundamentos y su fiabilidad. La frecuencia con la que un patrón o un conjunto de características particulares se observa en diferentes muestras, que es un elemento esencial para inferir las conclusiones, no es un asunto de "juicio". Es un hecho empírico para el cual solo es relevante la evidencia empírica. De forma similar, la expresión de confianza utilizada por un analista, basada en su experiencia profesional, o las expresiones sobre un consenso entre analistas respecto a la precisión a la que llegan en su campo de trabajo, no pueden sustituir a las tasas de error estimadas a partir de estudios relevantes. Para los métodos forenses de comparación de características, el establecimiento de la validez de sus fundamentos a partir de evidencia empírica es entonces una condición sine qua non. Nada puede sustituirla.

Como puede inferirse de lo anterior, el giro estadounidense no solo exige prestar atención a lo que dicen las comunidades expertas, sino preguntar por la evidencia empírica que fundaría cierta aceptación de una comunidad. Y esa evidencia no se refiere a documentos de consenso sobre buenas prácticas o códigos éticos. Nuevamente, al respecto un ejemplo del propio PCAST Report:

La validez de los fundamentos requiere que un método haya estado sujeto a estudios empíricos realizados por múltiples grupos, en condiciones apropiadas para el uso pretendido. Los estudios deben (a) demostrar que el método es repetible y reproducible; y (b) proporcionar estimaciones válidas sobre la precisión del método (es decir, cuántas veces el método alcanza una conclusión incorrecta) que indiquen que es apropiado para la aplicación pretendida.

En definitiva, se trata de la investigación empírica que puede realizar una comunidad experta sobre los métodos o teorías o técnicas, etc., que utilizan. Un complejo escenario en el que se aboga en primer lugar por una cultura de la investigación. Una cultura en la que, según Mnookin, et. al. (2011: 740):

las siguientes preguntas son primordiales: ¿Qué sabemos? ¿Cómo lo sabemos? ¿Qué seguridad tenemos de ello? ... [U]na cultura en la que estas preguntas se responden con datos, estudios publicados y materiales accesibles al público, en lugar de referirse principalmente a la experiencia o el conocimiento artesanal, o donde simplemente se asumen como verdades porque hace tiempo que tratan como si fueran verdaderas.

En este punto vale la pena aludir al criterio de la revisión por pares previsto en Daubert, distinguiendo primero un sentido amplio y un sentido estricto de dicha revisión (Haack 2020: 239). En sentido amplio hace referencia a cualquier

---

informe es titulado "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature – Comparisons Methods".

discusión de las ideas presentadas por un experto en una colectividad científica específica, por ejemplo, en congresos, seminarios, coloquios o incluso entre los compañeros de laboratorio o despacho. Mientras que, en sentido estricto, la revisión por pares hace referencia a una evaluación *formal* de trabajos escritos para determinar, bajo determinados parámetros, si es publicada o no en alguna casa editorial o revista científica. Por supuesto, en términos genéricos y teóricamente, el sentido estricto es más exigente que el sentido amplio y también tiene mayores efectos controlables.

El mero hecho de publicar es un paso en la dirección correcta, pues se hacen *públicas* las ideas, las pruebas, las investigaciones, etc. Es innegable que la publicación sirve como medio de comunicación *para* el área de conocimiento y los especialistas (público) a los que van directa o indirectamente dirigidas. Pero ¿por qué sería un criterio adecuado para que los juristas conozcan la calidad de las pruebas periciales? Una posible respuesta es que podría servir precisamente como indicador de dicha calidad: así, en lugar de valorar las generalizaciones usadas por el experto, bastaría con saber si han sido objeto de revisión por pares y publicadas. No obstante, para ello habría que saber algo más sobre *dónde* se publicó y si fue objeto de una evaluación adecuada por pares.

Ahora bien, ¿cuán relevante sería haber superado la evaluación por pares para la publicación? Será más o menos relevante en la medida en que revele errores o defectos metodológicos e, incluso, permita corregirlos; es decir en función de cuán bien funcione la revisión en sí misma. Sin embargo, muy probablemente lo más relevante es lo sucedido *después* de la publicación y no el mero hecho de la publicación:<sup>77</sup> ¿qué críticas ha recibido la publicación en cuestión por diversos expertos?, ¿existen otras publicaciones confirmando, por ejemplo, los experimentos realizados?, etc.<sup>78</sup>

---

<sup>77</sup> Con el tiempo y la dinámica científica, sólo una ínfima parte de los millones de publicaciones actualmente existentes serán libros o artículos de referencia *del futuro* (Bauer, 1992: 45). Y, por cierto, esa información de referencia tal vez nunca llegue a entrar en los procesos judiciales precisamente por su grado de probabilidad o quizá llegará a entrar en algún proceso judicial pero como un “hecho notorio” y no como una prueba pericial.

<sup>78</sup> Curiosamente el artículo 422 del Código de Procedimiento Penal Colombiano hace más exigente su versión de este factor Daubert: “que la teoría o técnica subyacente haya sido publicada y haya recibido la crítica de la comunidad científica”.

Para recapitular, podríamos decir que desde el punto de vista jurídico cuando los jueces tienen que decidir con conocimiento científico deben considerar su fiabilidad, presuponiendo ya su falibilidad. Y, para esa tarea, los criterios Daubert recientemente traídos a España por el Tribunal Supremo ofrecen criterios guía que dan cuenta de los mecanismos de validación del conocimiento científico que subyace a una prueba pericial: la aceptación de una comunidad científica y las publicaciones que han pasado la revisión por pares.

En los apartados anteriores se han desarrollado argumentos sobre el carácter probabilístico del razonamiento probatorio y la falibilidad del conocimiento científico para mostrar porque la exigencia de certezas hecha por la Audiencia Nacional española no tiene sentido ni en el derecho ni en la ciencia. Ello se ha hecho sin aludir al ámbito del dopaje, pues se trata de argumentos generales independientes al tipo de caso que se trate. No pasa lo mismo con la fiabilidad, puesto que su valoración dependerá del tipo de conocimiento científico en juego. Por ello, en este punto hay que centrarse en el dopaje y, para ello, vale la pena traer a colación la ley Orgánica 11/2021, de 28 de diciembre, de lucha contra el dopaje en el deporte, en su artículo 39 titulado “alegaciones y medios de prueba”, establece que:

- d) Se presume, salvo prueba en contrario, la validez científica de los métodos analíticos y de los límites de decisión que apliquen los laboratorios de control antidopaje debidamente autorizados.
- c) Se presume que los laboratorios de control de dopaje que se encuentren acreditados o aprobados por la Agencia Mundial Antidopaje, realizan los análisis de muestra y aplican los procedimientos de custodia conforme a la normativa aplicable.<sup>79</sup>

A primera vista resulta muy llamativo que una norma jurídica establezca una presunción sobre la validez científica de cierto tipo de conocimiento experto.

---

<sup>79</sup> Y sigue: “No obstante, el deportista u otra persona pueden demostrar que el laboratorio ha contravenido la regulación aplicable y que esta circunstancia podría razonablemente haber causado el resultado analítico adverso, en cuyo caso el órgano competente tendrá la carga de demostrar que esa contravención de la normativa aplicable no ha dado lugar al mismo. En estos casos la Agencia Estatal Comisión Española para la Lucha Antidopaje en el Deporte pondrá esta circunstancia en conocimiento de la Agencia Mundial Antidopaje, a los efectos de la interposición de un posible recurso contra la decisión que se adopte. Los laboratorios de control de dopaje deberán estar siempre en disposición de exhibir los documentos en los que conste la vigencia de la acreditación concedida por la Agencia Mundial Antidopaje.”



Como antes se había dicho, en los procesos judiciales lo más habitual es que ese tipo de información sea parte del debate entre las partes o sea una consideración realizada por los jueces sobre todo en sede de valoración probatoria. Quizá el único caso en el que la calidad del conocimiento experto se da por sentada es la fiabilidad de la genética forense; entiéndase bien, las partes pueden discutir si la *aplicación* de la genética forense a un caso concreto se hizo adecuadamente o no, lo que se da por sentado no es su correcta aplicación en el caso, sino la calidad del conocimiento científico subyacente.<sup>80</sup>

¿Tiene sentido que se prevea en una norma la fiabilidad del conocimiento experto? Este tipo de normativa, en mi opinión, solo es racional si refleja la realidad, es decir, si *realmente* se ha hecho el trabajo necesario para validar los métodos o técnicas correspondientes. No tendría ningún sentido que las normas atribuyan validez científica a métodos que en realidad no la tuvieran.

En el ámbito del dopaje, hay que considerar el trabajo que se ha hecho por la WADA, quien no solo prevé un presupuesto considerable para la investigación continua de los métodos que se utilizan para la identificación del dopaje en el deporte<sup>81</sup>, sino que además el Código Mundial Antidopaje, en su artículo 3.2.1 prevé:

Los métodos analíticos o los estándares para la decisión (*Decision Limits*) aprobados por la WADA previa consulta a la comunidad científica pertinente o que hayan sido objeto de una revisión por pares se presumen científicamente válidos.

Podemos encontrar mayores detalles sobre las exigencias sobre la validación del conocimiento experto por parte de la WADA, por ejemplo, en el artículo 5.4.4,

---

<sup>80</sup> McCormick (1954: 363) considera que la aceptación general de la comunidad científica sería un criterio adecuado para que los tribunales dejen fuera del ámbito probatorio un hecho por considerarlo *notorio*. Siguiendo su razonamiento, entonces, la calidad de la genética forense no sería susceptible de prueba por las partes porque se podría asumir que es así.

En otro lugar, he distinguido los hechos de notoriedad conocida y de notoriedad fácil y fiablemente cognoscible (Vázquez, 2022): los primeros serían aquellos hechos de dominio público, asumidos como verdaderos por la mayoría de las personas en una sociedad determinada; los segundos, en cambio, son aquellos cuya verdad puede conocerse fácilmente si se consultan las fuentes de acceso fácil y que son fiables. La información que, por ejemplo, la propia WADA en el ámbito antidopaje ofrece sobre los métodos que emplea, las publicaciones que al respecto se tienen, etc., permitiría debatir si los jueces pudieran introducirla como hechos de notoriedad cognoscible de forma fácil y fiable.

<sup>81</sup> Puede verse en su página web: <https://www.wada-ama.org/en/data-research/scientific-research>

titulado “Métodos de controles y validación de los métodos”, previsto en el Estándar Internacional para Laboratorios. Es una normativa que hace referencia a las sustancias prohibidas y que, por citar un ejemplo, sobre las identificadas como “sustancias umbral” (cuya prohibición se sitúa por encima de determinada concentración), se exige al laboratorio (artículo 5.4.4.2.2) que:

Los métodos de confirmación de Sustancias Umbral deberán ser validados. Los factores que se investigarán para demostrar que el método es apto para el propósito incluyen, entre otros:

- Especificidad. Se determinará y se documentará la capacidad del ensayo para detectar solo la sustancia de interés. El ensayo será capaz de discriminar entre compuestos de estructuras estrechamente relacionadas;
- Precisión Intermedia. El método permitirá la repetición fiable de los resultados en diferentes momentos y con diferentes operadores realizando el ensayo.
- Se registrará la Precisión Intermedia en el umbral;
- Solidez. Se determinará el método para producir los resultados similares con respecto a pequeñas variaciones en las condiciones analíticas. Serán controladas las condiciones que son esenciales para obtener resultados reproducibles;
- Remanente. Se determinarán e implementarán las condiciones requeridas para eliminar el remanente de la sustancia de interés de una *Muestra* a otra durante el procesamiento o análisis instrumental;
- Interferencias de matriz. El método limitará la interferencia en la medición de la cantidad de *Sustancias Prohibidas* o sus *Metabolitos* o *Marcadores* por componentes de la matriz de la *Muestra*;
- Estándares. Se utilizarán Materiales de Referencia para la cuantificación, si están disponibles.
- Límite de cuantificación (LOQ). El Laboratorio demostrará que un método del umbral tiene un LOQ establecido de no más que el 50 % del valor umbral o de acuerdo con los valores de LOQ requeridos en los Documentos Técnicos o Directrices pertinentes;
- Se documentará la linealidad entre el 50 % y el 200 % del valor umbral, menos que se estipule lo contrario en un Documento Técnico o Directrices.

También está previsto seguimiento del trabajo desarrollado por los laboratorios acreditados a través del plan de evaluación externa de la calidad (EQAS en sus siglas en inglés) que tiene como objetivo garantizar la calidad de los resultados analíticos.

Uno de los problemas que se enfrenta sistemáticamente en los procesos judiciales para conocer el trabajo de validación realizado por una comunidad experta es precisamente la identificación de dicha comunidad, que no siempre está bien delimitada o que a veces solo se tienen meros “grupos de expertos” dispuestos a defender sus intereses. Pero, en el ámbito antidopaje, no estamos en esa situación, por el contrario, se tiene una organización institucional a nivel internacional, por lo que es más fácil identificar el trabajo de validación de lo que se ha constituido como la “comunidad científica” en la materia. Una comunidad

científica también particular debido a las diversas áreas de conocimiento o especialidades que la conforman: hematólogos, médicos deportivos, científicos forenses, etc.<sup>82</sup>

En el trasfondo de la regla que presume la validez de los métodos analíticos hay pues una institución que se encarga de todo el proceso de validación. Esta situación es semejante a la que han sugerido para las ciencias forenses Champod y Vuille (2010: 54), una institución que emita solo recomendaciones no-vinculantes y que actuara fundamentalmente para brindar información sobre la validación de nuevas técnicas forenses o la fiabilidad de nuevas formas de prueba o cuando todo indicara la necesidad de abandonar un tipo de prueba pericial que ha quedado obsoleta.

Un paso más allá que el comentado anteriormente, lo dio la Law Commission, en su conocido informe *The Admissibility of Expert Evidence in Criminal Proceedings in England and Wales* (2009), pues sugirió la posibilidad de que, en casos complejos, el juez pudiera ser auxiliado para determinar la fiabilidad del testimonio del experto. En el mismo sentido, en Estados Unidos, el PCAST Report recomienda que sea el National Institute of Standards and Technology (NIST, por sus siglas en inglés) quien se encargue de evaluar regularmente la validez fundacional del tipo de pruebas periciales que estaban analizando (es decir, los métodos y técnicas forenses de comparación de rasgos).

---

<sup>82</sup> El Estándar internacional de gestión de resultados, en su sección de términos alude a diversas especialidades que pueden constituir los paneles de expertos que valoran el pasaporte del deportista y que nos da una idea de las diversas disciplinas implicadas:

“Los expertos y/o el panel de expertos, con conocimiento en el campo en cuestión, elegidos por la Organización Antidopaje y/o la Unidad de Gestión del Pasaporte del Deportista, son responsables de proporcionar una evaluación del Pasaporte. El experto debe ser externo a la Organización Antidopaje.

Para el módulo hematológico, el panel de expertos debe estar formado por al menos tres (3) expertos que tengan calificaciones en uno o más de los campos de hematología clínica y de laboratorio, medicina deportiva o fisiología del ejercicio, según se apliquen al dopaje sanguíneo. Para el módulo esteroideo, el panel de expertos debe estar compuesto por al menos tres (3) individuos con calificaciones en los campos de análisis de esteroides en laboratorio, dopaje con esteroides y metabolismo y/o endocrinología clínica. Para ambos módulos, un panel de expertos debe estar compuesto por expertos con conocimientos complementarios de modo que todos los campos relevantes estén representados. El panel de expertos puede incluir un grupo de al menos tres (3) expertos designados y cualquier experto (s) ad hoc adicional que pueda ser requerido a solicitud de cualquiera de los expertos designados o por la Unidad de Gestión del Pasaporte del Deportista de la Organización Antidopaje.”

Entonces, no ha sido ajeno al ámbito jurídico probatorio el importante rol que pueden jugar ciertas instituciones de validación del conocimiento científico para que los jueces y juezas puedan tener información relevante al respecto y valorar racionalmente las pruebas periciales. Una tarea que debe atender al mayor o menor grado de fiabilidad del conocimiento experto.

Como puede advertirse de todo lo anterior, la información que debe considerarse para la atribución de valor probatorio a una prueba pericial es mucho más compleja y, a la vez, más sofisticada que cualquier alusión a una pretendida certeza (psicológica). La Audiencia Nacional española no se pronunció sobre la calidad de “la ciencia del dopaje”, pero al exigirle certezas simplemente desconoció todo el trabajo de validación que la respalda.<sup>83</sup>

## **VI. La distribución del riesgo de error en las decisiones (científicas y jurídicas) sobre dopaje**

Como dice el Prólogo de la Ley Orgánica 11/2021 de 28 de diciembre, de lucha contra el dopaje en el deporte: “[f]rente a las prácticas de dopaje deportivo, cada vez más sofisticadas y difíciles de detectar, se requiere adoptar acciones especialmente rigurosas que las prevengan mediante la intensificación de los controles y, en su caso, de las medidas de reacción en caso de producirse.” En este sentido, la misma ley, en su artículo 20, prevé como infracciones en materia de dopaje:

- a) La presencia de una sustancia prohibida, o de sus metabolitos o marcadores, en las muestras biológicas del deportista.
- b) La utilización, uso o consumo de sustancias o métodos prohibidos en el deporte.

Tradicionalmente, en la lucha contra el dopaje las medidas disciplinarias se habían establecido únicamente cuando en la muestra del deportista se encontraba una sustancia prohibida. No obstante, desde hace algún tiempo el Código Mundial Antidopaje, en su artículo 2.2., también ha previsto como

---

<sup>83</sup> Obviamente se podría debatir la fiabilidad de la aplicación de métodos y técnicas al caso concreto, ya que la valoración judicial de la prueba pericial no puede no hacerlo. No obstante, como se ha dicho desde el inicio, este trabajo se centra en el criterio emitido por la Audiencia Nacional y no en análisis de los hechos del caso concreto.

infracción el uso o intento de uso por parte de un deportista de una sustancia prohibida o de un método prohibido. Esto último, desde luego, ha supuesto un importante giro, puesto que las infracciones disciplinarias ahora se pueden acreditar sin que se encuentre una sustancia en una muestra y esto ha sido así porque las técnicas de dopaje han cambiado radicalmente; ahora, por ejemplo, se usan métodos de manipulación genética y de manipulación en los componentes de la sangre. En términos probatorios, esto ha supuesto que no siempre se tenga prueba directa del dopaje de un deportista (i.e. la identificación de una sustancia) y que la información disponible constituya prueba indirecta.

Para ejemplificar lo anterior, consideremos uno de los métodos de dopaje como es la manipulación de la sangre. Desde el punto de vista científico, se ha avanzado en el desarrollo de pruebas fiables que permitan mostrar dicha manipulación a través de *marcadores indirectos* de dopaje. Es muy interesante advertir que se ha pasado de excluir de competiciones a los deportistas que mostraban valores (estadísticamente) anormales en *una* muestra de sangre, al desarrollo de métodos más específicos que consideran *una secuencia completa* de resultados de análisis del individuo. En ese sentido, se ha dicho que antes la regla de decisión se basaba en “el umbral arbitrario de la especificidad del parámetro sanguíneo y no en una verdadera prueba de manipulación de la sangre” (Sottas, et.al., 2008: 193). Es más, se ha considerado que se trataba de un método que servía básicamente para “identificar correctamente los casos en los que no había dopaje [*the negative cases*]”, pero no los casos positivos.

Hay que señalar que la distinción entre pruebas directas e indirectas no puede ser vista ni como una diferencia categorial ni como un criterio de fiabilidad de las pruebas (González Lagier, 2022: 386-388). No hay una diferencia de categoría porque en todas las pruebas hay pasos inferenciales que se deben hacer, inclusive cuando se trata de un testigo que presencié los hechos. Tampoco puede asumirse sin más que la prueba directa es más fiable que la indirecta: un testigo que supuestamente ha presenciado los hechos, por ejemplo, puede ser mucho menos fiable que una prueba indirecta sobre el ADN de un sospechoso.

Entonces, lo único que tenemos son pruebas más o menos indirectas sobre determinada hipótesis y, en ese contexto, hay que prestar atención al número de pasos inferenciales que separan la hipótesis de las pruebas. Consideremos que se pretende probar si Pedro se dopó usando un método prohibido, la manipulación de su sangre; y, para ello, tenemos que hacer una tarea de comparación. Si la comparación se hace entre una única muestra del individuo y un patrón estadístico de *normalidad* de cierto parámetro sanguíneo, se enfrentan muchos problemas para poder hacer inferencias fundadas sobre el caso concreto, para empezar porque la población de referencia empleada para calcular la frecuencia podría tener variables que no tiene la población de referencia analizada: no ser deportista, tener un origen étnico distinto, una edad que no fue medida, etc. Son tantos los problemas relevantes sobre la clase de referencia que se llega a decir que los datos estadísticos nada dicen sobre el caso concreto. Y, claro, a todas las cuestiones sobre la población de referencia analizada, hay que sumar si puede haber diversos factores que un día determinado pueden influir de manera relevante en los resultados de una única muestra. Así pues, como puede fácilmente advertirse, hay muchos pasos inferenciales que deben hacerse para conectar la hipótesis con la prueba, varios de ellos pueden ser muy débiles y, además, puede que no se hayan excluido otras hipótesis alternativas.

La situación es muy distinta, en cambio, cuando se dispone de información más rica y individualizada sobre el caso que permite fundar inferencias probatorias. En el ámbito de la prueba del dopaje mediante el pasaporte biológico, esto se da cuando:

- Se han identificado marcadores específicos del dopaje sanguíneo, que son mucho más específicos que los parámetros sanguíneos individuales resultantes en un hemograma.<sup>84</sup>

---

<sup>84</sup> La literatura especializada (y, después, también la normativa) habla sobre marcadores primarios (que son más específicos) y marcadores secundarios (que proporcionan pruebas de apoyo del dopaje de forma aislada o en combinación con otros marcadores). Un marcador primario, por ejemplo, es la concentración de hemoglobina (HGB); y, en cambio, un marcador secundario es la puntuación del perfil sanguíneo anormal.

Los marcadores sanguíneos hacen uso de varios parámetros sanguíneos. Por ejemplo, el Abnormal Blood Profile Score (ABPS) es un marcador universal que se basa en hasta 12 parámetros sanguíneos combinados con técnicas de clasificación de patrones para detectar tanto el abuso de eritropoyetina recombinante (rHuEPO) como la transfusión de sangre (Sottas, et. al., 2008: 192).

El efecto del dopaje, según Krumm y Faiss (2021: 2), no se examina directamente a través de la modificación de los parámetros biológicos, sino a través de la variación de la media y la desviación estándar de estos biomarcadores.

- La comparación se hace con los parámetros previos del propio sujeto, teniendo una secuencia completa de sus valores individuales de un parámetro sanguíneo específico (mediante lo que se conoce como el pasaporte biológico del deportista).

Hay que tomar en cuenta que se hace un procesamiento automático de los datos mediante lo que se identifica como un modelo adaptable (llamado ADAMS) que es “un modelo matemático ... diseñado para identificar resultados longitudinales inusuales de los Deportistas. El modelo calcula la probabilidad de un perfil longitudinal de valores de los Marcadores asumiendo que el Deportista tiene un estado fisiológico normal.”<sup>85</sup>

Es adaptativo porque, como lo explican Krumm y Faiss (2021: 2), “cuando se recoge una primera muestra, se determinan los umbrales superior e inferior con puntos de referencia medios basados en la población. A continuación, estos límites individuales se adaptan progresivamente en función de los valores de cada atleta a medida que se toman muestras adicionales.”

---

<sup>85</sup> Es la definición dada en las Directrices operativas y compilación de elementos necesarios del Pasaporte Biológico del Deportista.

Según el Estándar internacional de gestión de resultados, C.2.1.1, “en ADAMS, el modelo adaptativo procesa automáticamente datos sobre los marcadores biológicos del Pasaporte Biológico del Deportista. Estos marcadores incluyen marcadores primarios que se definen como los más específicos para el dopaje y marcadores secundarios que proporcionan evidencia de respaldo de dopaje en forma aislada o en combinación con otros marcadores. El modelo adaptativo predice para un individuo un rango esperado dentro del cual entra una serie de valores de marcadores que implican una condición fisiológica normal.”

- Se cuenta con criterios de interpretación de los parámetros resultantes. Como dicen Sottas, et.al. (2008: 195), “la disponibilidad de todos los factores simplemente ayuda a disminuir la variación global del parámetro sanguíneo: cuanto mayor sea la cantidad de información, mayor será la sensibilidad.”<sup>86</sup>  
Como el hecho de conocer los factores que pueden influenciar los parámetros de la sangre (el tipo de deporte que se práctica, el género, el origen étnico, la edad, la altitud, etc.)<sup>87</sup> o, mejor aún, cómo y cuánto pudieran influir.

Estamos pues ante un tipo de análisis individual, longitudinal y adaptativo que permite medir los posibles *efectos* de un dopaje e inferir, por tanto, que el deportista se dopó. Para esa medición, por ejemplo, en el pasaporte biológico se registran y analizan catorce parámetros como parte de su módulo hematológico. No obstante, como se ha dicho antes, hay muchas variables que pueden influir en los resultados y que condicionan entonces su correcta interpretación:

- a. Múltiples factores intrínsecos al deportista (por ejemplo, genéticos o ciertas enfermedades).
- b. Múltiples factores extrínsecos al deportista (el tipo de entrenamiento o las condiciones ambientales<sup>88</sup>).
- c. Condiciones de los estudios antes y durante (el funcionamiento de los instrumentos que se utilizan, cuándo se extrae la sangre o la conservación de la muestra<sup>89</sup>).

---

<sup>86</sup> La sensibilidad de una prueba o test es su capacidad de identificar correctamente los casos positivos (es decir, es más sensible cuantos menos falsos negativos da). La especificidad es su capacidad de identificar correctamente los casos negativos (es decir, es más específica cuantos menos falsos positivos da).

<sup>87</sup> El número y el tipo de estos factores puede variar en función de los desarrollos teóricos o avances tecnológicos en las áreas de conocimiento relevante.

<sup>88</sup> Según Krumm y Faiss (2021: 1), los efectos sobre las variaciones hematológicas de las condiciones ambientales (por ejemplo, la exposición al calor o a la hipoxia) siguen siendo discutibles.

<sup>89</sup> Así, por ejemplo, el ruido en los dispositivos electrónicos, el tiempo transcurrido entre la transfusión y la extracción de sangre, el procedimiento de conservación de la muestra, etc. (Sottas, et. al, 2006: 2).



Estos son, pues, factores *confundentes* (*confounding factors*) en la interpretación de los resultados. Por ejemplo, según Krumm y Faiss (2021), el volumen plasmático está en el centro de muchos cambios hematológicos y sus variaciones tienen ciertas dificultades de interpretación en el contexto deportivo, pues los hematíes diluidos en plasma se pueden ver afectados de forma diferente por factores intrínsecos del deportista como un entrenamiento prolongado y agudo. Pero, según la revisión sistemática de estudios que hicieron Krumm y Faiss (2021: 12), el origen más habitual de factores confundentes para la interpretación de los resultados son los métodos o formas en las que se practica el dopaje sanguíneo.

Y en este punto, debemos preguntarnos, ¿cuáles son los riesgos de error en la medición de los diversos parámetros sanguíneos que pueden conducirnos a un falso positivo o a un falso negativo? En efecto, claramente hay tantas variables a considerar en la determinación de un dopaje sanguíneo que no podemos esperar razonablemente, como pretende la Audiencia Nacional, alcanzar certezas en un contexto así.<sup>90</sup> Los criterios Daubert, recién traídos al ámbito español por el Tribunal Supremo, marcan también un giro importante al poner sobre la mesa la importancia del conocimiento sobre los errores,<sup>91</sup> como no puede ser de otra manera cuando se asume la falibilidad del conocimiento científico.

En lugar de seguir insistiendo en objetivos inalcanzables, como las certezas racionales, o en estados netamente subjetivos, como las certezas psicológicas, deberíamos *aprender a aprender* sobre los errores y, por supuesto, exigir la información relevante al respecto. En esa línea, por ejemplo, la National

---

<sup>90</sup> Este tipo de incertidumbre también está presente en la prueba de ADN. Como dicen Prieto y Carracedo (2021: 219), sobre el estudio de la transferencia de perfiles genéticos en los que “no solo influye si la persona descama células epiteliales en mayor o menor cantidad, sino el tiempo de contacto y el tipo de superficie sobre el que se depositan las células.”

<sup>91</sup> Concretamente el criterio sobre el “rango de error conocido o posible, si se trata de una técnica científica, así como la existencia de estándares de calidad y su cumplimiento durante su práctica.” Aunque una interpretación literal del criterio nos conduciría a estimaciones cuantitativas de fiabilidad en términos de frecuencias de errores, podríamos también analizar las fuentes de errores, con el objetivo de buscar disminuir mediciones o decisiones equivocadas, o debatir sobre cómo se distribuyen los errores.

Academy of Science (2009: 188) estadounidense en su *Report on Forensic Science* ya recalca que:

una agenda de investigación completa debe incluir estudios para establecer las fortalezas y limitaciones de cada procedimiento, fuentes de sesgos y variación, cuantificación de las incertezas creadas por estas fuentes, mediciones del rendimiento, los pasos procedimentales en el proceso de analizar pruebas forenses y los métodos para continuar monitoreando y mejorando los pasos de tal proceso.

Y, Mayo (1996: xii), en el contexto del conocimiento experimental, afirma:

conviene aprender cuáles son los diferentes tipos de error, qué es lo que específicamente se aprende cuando se reconoce un error, cómo se identifica cuándo es culposos, cómo crece nuestra habilidad para detectar y corregir errores, y cómo está relacionado este conocimiento con el crecimiento del conocimiento científico”.

No es objetivo de este trabajo analizar cada uno de los puntos mencionados sobre los errores<sup>92</sup>, pero sí, como se verá, dedicar las últimas páginas a plantear algunas cuestiones sobre la distribución de los riesgos del error.

Vale la pena iniciar distinguiendo dos tipos de error, los falsos positivos (resultados que señalan que un deportista se dopó cuando en realidad no se ha dopado) y los falsos negativos (por el contrario, resultados que señalan que un deportista no se dopó cuando en realidad sí se ha dopado)<sup>93</sup>. Se puede dar un falso negativo cuando, por ejemplo, el método de dopaje es a través de microdosis de sustancias estimuladoras de la sangre, pues todo apunta a que complican el análisis de los perfiles sanguíneos debido a las limitadas variaciones hematológicas que producen (Krumm y Faiss, 2021: 12). De hecho, según los mismos autores “la eficacia del pasaporte parece significativamente inferior con protocolos de transfusión autóloga de bajo volumen debido a varios factores que pueden influir en su sensibilidad (por ejemplo, el momento en que se recogen las muestras).”

---

<sup>92</sup> Por ejemplo, sobre los sesgos cognitivos que pueden sufrir los expertos y que potencialmente los llevan a decisiones erróneas, puede verse Vázquez (2022), capítulo II, y Duce (2022).

<sup>93</sup> Lo mismo puede verse en cualquier otro razonamiento fáctico, tanto ordinario como forense. Así, por ejemplo, en el caso de la identificación de las huellas dactilares, un falso positivo ocurriría si mediante una técnica concreta se concluye que dos huellas tienen una misma fuente cuando, de hecho, no la tienen; y un falso negativo ocurre si mediante la técnica concreta en cuestión se concluye que las dos huellas no tienen una misma fuente cuando, de hecho, sí la tienen. Edmond (2020) pone sobre la mesa cómo los jueces y tribunales estadounidenses malinterpretan las tasas de error de la identificación por huellas dactilares.

Y, por el contrario, se podría dar un falso positivo en la interpretación de la variación del volumen de plasma del deportista si no se considera su exposición aguda a calor o frío extremo o la hipoxia (Krumm y Faiss, 2021: 12). Vale la pena señalar, no obstante, que los protocolos establecidos exigen que el deportista informe adecuadamente cualquier exposición a hipoxia o condiciones ambientales extremas y otras condiciones patológicas, a efectos de tener en cuenta las posibles variaciones preanalíticas.

Uno de los aspectos importantes de distinguir los tipos de error es que uno de ellos puede considerarse más costoso que el otro<sup>94</sup> y, por eso se establecen reglas de decisión que establecen criterios que dan preferencia a un tipo de error sobre otro, disminuyendo el riesgo de que se cometa uno de ellos.<sup>95</sup> En el ámbito del dopaje podemos encontrar varios ejemplos de decisiones tomadas por los expertos que tienen como objetivo evitar los falsos positivos, es decir, los

---

<sup>94</sup> Piénsese, por ejemplo, en las falsas condenas y los altísimos costos que suponen para la legitimidad de los sistemas de justicia y, por supuesto, para las personas que los sufren. Las falsas absoluciones son también un tipo de error. En el ámbito jurídico, se ha analizado el tema del error en las decisiones judiciales, tanto en lo que hace a las falsas absoluciones como a las falsas condenas, especialmente en el debate sobre los estándares de prueba. Puede verse, por todos, Laudan (2006).

<sup>95</sup> Por supuesto, también se puede pensar en los mecanismos necesarios para evitar que suceda sobre todo un tipo de error o incentivar la disminución de los errores en general.

Además de las diversas variables comentadas en párrafos arriba, una fuente de error sería el no-seguimiento de la cadena de custodia de las muestras de los deportistas. La WADA tiene previsiones amplias y claras sobre la recolección, seguridad, transporte, etc. de las muestras en el Estándar Internacional for Testing and Investigations. Ahora bien, el Código mundial antidopaje prevé, artículo 3.2.2, la presunción iuris tantum de su cumplimiento por los laboratorios oficiales: se presume que los laboratorios acreditados y otros aprobados por la WADA realizan análisis de muestras y aplican procedimientos de custodia que son conformes al Estándar internacional para laboratorios: "El deportista u otra persona podrá rebatir esta presunción demostrando que se ha producido un incumplimiento de dicho Estándar Internacional que podría razonablemente haber causado el Resultado Analítico Adverso."

Vale la pena notar que para la derrotar la presunción se prevé un estándar más bien bajo. La versión en inglés del Código mundial alude al "balance de probabilidades" en el comentario al artículo 3.2.2: "the burden is on the Athlete or other Person to establish, by a balance of probability a departure from the International Standard for Laboratories that could reasonably have caused the Adverse Analytical Finding." La versión (no oficial) en castellano comete un error importante al respecto toda vez que alude a "atender las probabilidades": "la carga de la prueba recae sobre el Deportista u otra Persona, quien debe demostrar, atendiendo a las probabilidades, que se ha producido un incumplimiento del Estándar Internacional para Laboratorios que podría haber causado razonablemente el Resultado Analítico Adverso."

Por supuesto, se deben probar dos hechos: primero, que no se cumplieron los Estándares y, segundo, que eso pudo haber causado el resultado adverso. Esto segundo, supuestamente, exigiría un "el criterio probatorio algo menos estricto – 'podría razonablemente haber causado'".

resultados que señalan que un deportista se dopó cuando en realidad no se ha dopado. Así, por ejemplo:

- “Una interpretación objetiva es difícil en el caso de dos picos superpuestos y cuando la población de glóbulos rojos menores es pequeña en comparación con la población de glóbulos rojos mayores. Estos aspectos pueden conducir a una disminución de la sensibilidad final de la prueba, ya que cualquier resultado dudoso se devuelve negativo para evitar cualquier falso positivo” (Sottas, et. al, 2006: 2 y 3).
- Como dicen Sottas, et.al (2008: 192) los hemogramas que se usan para la identificación de dopaje son los mismos que se emplean en la clínica. Ambos “se basan en estructuras de datos similares: los recuentos/valores anormalmente altos o bajos de los parámetros sanguíneos pueden indicar la presencia de una enfermedad o de una manipulación sanguínea, pero difieren en las reglas de decisión asociadas. En las pruebas de diagnóstico médico, el 95% de la población normal se encuentra dentro de los límites que definen el intervalo de referencia para un parámetro determinado. En el caso de la regla de “no inicio”, los intervalos de referencia suelen corresponder a un percentil superior de la población de referencia, con el fin de reducir la probabilidad de excluir de la competición a un deportista inocente. Dado que las poblaciones de referencia suelen estar compuestas únicamente por deportistas no dopados, el riesgo de un falso positivo es elevado: un aumento del número de controles antidopaje eleva la probabilidad de encontrar una coincidencia por pura casualidad”.
- Otra referencia de Sottas, et. al (2006: 7): “en la práctica, es posible incluir un margen para disminuir aún más el riesgo de dar un falso positivo. Los falsos negativos se caracterizan por una pequeña área de este segundo pico, una situación que se produce cuando la cantidad de sangre transfundida sigue siendo pequeña en comparación con la sangre original. La sensibilidad disminuye al 66% (6 sobre 9 muestras) cuando la relación entre el área del segundo pico (población menor de GR) y el área del pico principal (población mayor de GR) es inferior al 2%. Esto significa que el 66% de estos histogramas se devuelven positivos cercanos al 100%.
- Según el Estándar internacional de laboratorios, los factores por considerar para que la WADA evalúe el rendimiento de todos los Laboratorios que certifica son, entre otros, sus resultados falsos negativos y sus Resultados Analíticos Adversos falsos, es decir, falsos positivos.

En el ámbito de la prueba de los efectos del dopaje a través del pasaporte biológico del deportista conviene distinguir tres escenarios: los resultados adversos a partir del pasaporte; los resultados atípicos o anómalos en el pasaporte; y cuando no hay un resultado atípico en el pasaporte, pero se solicita la revisión de la secuencia de datos del deportista. El Estándar Internacional sobre gestión de resultados define un resultado adverso como un informe que da cuenta de la presencia en una muestra “de una sustancia prohibida o de sus metabolitos o marcadores, o pruebas del uso de un método prohibido.” En cambio, según la misma normativa, artículo C.2.1.2, un resultado atípico es “un

resultado generado por el modelo adaptativo en *ADAMS* que identifica los valores de los marcadores primarios como fuera del rango intraindividual del deportista o un perfil longitudinal de los valores de un marcador primario (desviaciones de secuencia) como estar fuera de los rangos esperados, asumiendo una condición fisiológica normal.” Un resultado atípico en el pasaporte puede, como veremos, convertirse en un resultado adverso.

Sobre esta normativa, la Agencia Mundial antidopaje, en el caso resuelto por la sentencia que origina este informe, preguntó a la Audiencia Nacional española si inferir el dopaje del resultado adverso en el Pasaporte Biológico, de acuerdo con lo previsto por la legislación vigente (por ejemplo, en el art. 39.3 LO 11/2021), resultaría compatible con el derecho a la presunción de inocencia. Pues bien, hay tres posibilidades para sostener una respuesta negativa a esta pregunta, esto es, para sostener que la prueba científica del resultado adverso no es suficiente para derrotar la presunción de inocencia en el ámbito jurídico:

- 1) Que la prueba del resultado adverso es probabilística y, en cambio, la prueba necesaria para derrotar la presunción de inocencia exige que se llegue a la certeza de la comisión del ilícito. Esta es la respuesta de la Audiencia Nacional, como vimos: “(...) el derecho sancionador no puede operar sobre el terreno de las probabilidades sino sobre el plano de las certezas” (F.J. 3º). El problema es que, como hemos visto ya también, esta respuesta es insostenible.
- 2) Que la inferencia que va de la existencia de un resultado adverso en el Pasaporte biológico a la comisión de un ilícito sancionable administrativamente tenga carácter presuntivo, siendo entonces una presunción de culpabilidad que sería, claro, contraria a la presunción de inocencia.
- 3) Que el estándar de prueba científico (para el resultado adverso) y el jurídico (para el ilícito administrativo sancionable) sean distintos y más exigente este último. En efecto, sin necesidad de sostener que el estándar jurídico sea la certeza, podría sostenerse que este es más exigente que el científico y que, por tanto, la prueba del resultado adverso no permita inferir, por sí sola, la prueba de la infracción jurídica.

Analicemos estas dos últimas posibilidades. La primera cuestión que vale la pena plantear es si efectivamente es una presunción o no. Esto es relevante en este punto porque en el ámbito jurídico podemos identificar tres reglas probatorias que hacen la función de distribuir el riesgo del error: la carga de la prueba, las presunciones y los estándares de prueba.

El término presunción en el ámbito probatorio es ambiguo.<sup>96</sup> Puede hacer alusión a las llamadas “presunciones judiciales” (también conocidas como presunciones hominis)<sup>97</sup> o a las “presunciones legislativas”; también se puede distinguir entre las presunciones que admiten prueba en contrario y aquellas que no la admiten (que algunos autores proponen considerar como ficciones jurídicas o incluso como reglas sustantivas)<sup>98</sup>. En el contexto de la regulación sobre dopaje citada, si la interpretáramos como presunción, claramente estaríamos hablando de una presunción establecida legislativamente y que admitiría prueba en contrario.<sup>99</sup>

Las presunciones legislativas que admiten prueba en contrario, siguiendo a Ullmann–Margalit (2019: 17ss), tendrían las siguientes características: (i) se trata de normas; (ii) que establecen una relación entre un hecho base y un hecho

---

<sup>96</sup> Dada la gran diversidad de reglas, inferencias y razonamientos que suelen etiquetarse como “presunciones”, Allen (1981) sostiene que “[l]as presunciones han sido utilizadas para alcanzar distintos fines. Han sido utilizadas para construir reglas de decisión destinadas a evitar *impasses* fácticos en el juicio oral; para adjudicar la carga de la prueba; para instruir al jurado sobre la relación entre determinados hechos; y para adjudicar cargas de producción de pruebas. Sin embargo, ... ninguno de los resultados pretendidamente alcanzados a través del uso de las presunciones son, de hecho, alcanzados *a causa* de las presunciones. En cambio, lo que sucede, o al menos lo que debería suceder, es que los variados problemas probatorios son resueltos sobre la base de las consideraciones políticas particulares involucradas en cada caso, más que a causa de lo que una presunción es, siendo la etiqueta “presunción” luego añadida al resultado”. Y, por ello, concluye que “[l]a palabra “presunción” es simplemente una etiqueta que ha sido aplicada a un conjunto ampliamente dispar de decisiones relacionado con el modo apropiado de juzgar y la manera en la cual pueden ser establecidos los hechos a fin de resolver una disputa jurídica.” Una tesis opuesta a Allen puede verse en Dei Vecchi (2020: 77ss), discutiendo la tesis reduccionista sobre las presunciones.

<sup>97</sup> Las llamadas presunciones hominis, en realidad, son las inferencias probatorias realizadas por los jueces (no se trata pues de reglas) a partir de los elementos de juicio y se dice que permiten pasar de un hecho conocido a otro hecho no conocido. Este uso del término “presunción”, por supuesto, resulta muy confundente pues el supuesto hecho presumido no es más que los hechos probados como resultado de la inferencia.

<sup>98</sup> Al respecto, puede verse Ferrer Beltrán (2022, capítulo 9).

<sup>99</sup> Así lo establece la dicción literal del mismo artículo en el inciso g): “El presunto infractor podrá refutar todos los hechos y presunciones que le perjudiquen y probar los hechos y circunstancias necesarios para su defensa.”

presumido; (iii) y obligan a sus destinatarios a aceptar el hecho presumido como si fuera verdadero siempre que se pruebe el hecho base; y, esto último, (iv) excepto si se prueba que el hecho presumido no sucedió.

Si aplicamos lo anterior a la norma prevista en el artículo 39.3 b) de la Ley Orgánica 11/2021, que establece que “[u]n resultado adverso en el pasaporte biológico del deportista constituirá prueba de cargo suficiente a los efectos de considerar existente la infracción tipificada en el artículo 20.b) de esta ley”, podemos concluir muy rápidamente que no se trata de una presunción, puesto que no hay un hecho base a probar que se conecte con la obligación de aceptar la ocurrencia de otro hecho.<sup>100</sup> En efecto, no hay dos hechos que sean relacionados por la regla, como sí los hay, por ejemplo, en la denominada presunción de conmorienza: si dos sujetos “murieron en el mismo accidente” entonces “murieron a la vez”<sup>101</sup>. En el caso del resultado adverso en el ámbito del dopaje, en cambio, lo que la regla establece es que la prueba de un resultado adverso equivale o es lo mismo que la prueba de que el deportista se dopó; así, según la previsión jurídica, “el deportista ha tenido un resultado adverso” es sinónimo de “el deportista se dopó”. En otros términos, no hay aquí presunción alguna (y mucho menos una presunción contraria a la presunción de inocencia), sino una regla jurídica que tiene el efecto práctico de igualar el estándar de prueba jurídico para el dopaje con el estándar de prueba científico para el resultado adverso. Lo que quedará por determinar, y abordaré más adelante, es cuál es el umbral de exigencia probatoria, el estándar de prueba, que permita concluir que está probado el resultado adverso, esto es, el dopaje.

Toda presunción *iuris tantum* de la que se concluya que un sujeto cometió un hecho ilícito atenta contra la presunción de inocencia porque supone que, si el hecho base llega a probarse, la carga de probar que *no* se ha producido el hecho

---

<sup>100</sup> Lo mismo puede decirse de la norma contenida en el artículo 39.3 a) de la misma Ley: “[u]n resultado analítico adverso en un control de dopaje constituirá prueba de cargo suficiente a los efectos de considerar existentes las infracciones tipificadas en el artículo 20, letras a) y b), de esta ley”.

<sup>101</sup> Me refiero a la presunción prevista en el artículo 33 del Código Civil: “Si se duda, entre dos o más personas llamadas a sucederse, quién de ellas ha muerto primero, el que sostenga la muerte anterior de una o de otra, debe probarla; a falta de prueba, se presumen muertas al mismo tiempo y no tiene lugar la transmisión de derechos de uno a otro.”

ilícito presumido correspondería al imputado. Por ello, precisamente, se ha defendido que en materia penal las presunciones están prohibidas, pues exigiría que el acusado pruebe que lo presumido en realidad es falso. Como dice Ullman-Margalit (2019: 33), las presunciones legales son incompatibles con el derecho penal pues supondrían desfavorecer al acusado, dado que:

[s]on requeridas en situaciones que, si se las describe genéricamente, presentan un patrón recurrente, caracterizado por un problema de decisión en el cual es previsible que los compelidos a actuar se vean atascados en un proceso de deliberación inconcluyente o abortado. Son requeridas, por tanto, ahí donde su ausencia probablemente conllevaría la toma de decisiones arbitrarias o aleatorias. ... Lo que estas reglas ofrecen, en efecto, es una política de escape basada en un sesgo sistemático a favor de uno de los tipos de alternativas disponibles por encima del otro u otros.

En cambio, como se ha dicho ya, los estándares de prueba son reglas de decisión que indican cuándo un conjunto de elementos de juicio es *suficiente* para considerar como probado un hecho, derrotando así la presunción de inocencia. Si el razonamiento probatorio es probabilístico, entonces deviene imprescindible dotarse de reglas que determinen el grado de probabilidad a partir del cual estamos dispuestos a dar por probada una hipótesis. En el caso concreto del resultado adverso en el pasaporte del deportista, vale advertir la peculiaridad de que se hace coincidir el estándar de prueba jurídico con el científico.<sup>102</sup> Es decir, jurídicamente estará probado el dopaje cuando científicamente esté probado el resultado adverso (por efecto de la sinonimia analizada unos párrafos más arriba), pero uno todavía se puede preguntar cuánta prueba se exige para que los expertos puedan llegar a considerar probado un resultado adverso.

Para el análisis de esta cuestión es necesario distinguir dos supuestos, que se rigen en los instrumentos legales internacionales por estándares de prueba diversos, en función de si la hipótesis del dopaje está basada en el hallazgo de un resultado atípico o anómalo en el pasaporte biológico del deportista o si esa hipótesis se sostiene en ausencia de un resultado atípico en el pasaporte. Veamos primero qué sucede con los llamados resultados atípicos o anómalos en el pasaporte biológico.

---

<sup>102</sup> Como lo advertí en Vázquez (2015: 268), “las comunidades expertas tienen estándares propios para dar por probada una hipótesis; sin embargo, estos estándares no necesariamente coinciden con los estándares de prueba jurídicos, por lo que podría plantearse una discrepancia entre ambos”. También advierte sobre estas cuestiones Ferrer Beltrán (2007: 48).



Siguiendo el Estándar internacional de gestión de resultados, en primer lugar, un resultado atípico en el pasaporte se da cuando el valor de un marcador primario cae fuera del rango intraindividual del atleta o cuando un perfil longitudinal de valores de marcadores primarios está fuera de los rangos esperados (desviaciones de secuencia), asumiendo una condición fisiológica normal (C.2.1.2)<sup>103</sup>. Ahora bien, ese resultado será revisado por un experto “en función de lo expresado en el pasaporte y otra información básica (por ejemplo, los horarios de la competición), que pueden estar disponibles, de modo que el experto no conozca la identidad del deportista” (C.2.2.1). Para ello, la normativa parece exigir el planteamiento de algunas hipótesis alternativas “uso de una sustancia o método prohibido” o “resultado de una condición fisiológica” o “resultado de una condición patológica normal” (C.2.2.5.1); y, en función de ello, decidir:

- a. normal
- b. sospechoso
- c. probable dopaje
- d. probable afección médica

Según el Estándar internacional de gestión de resultados (C.2.2.5.1), la conclusión de "probable dopaje", exige que la probabilidad de que el resultado atípico en el pasaporte sea la consecuencia del uso de una sustancia o método prohibido *supere la probabilidad* de que el resultado atípico sea la consecuencia de una condición fisiológica o patológica normal. Una conclusión así supondrá, entonces, que el pasaporte se envíe a dos expertos adicionales para su revisión. Tal revisión se hará sin conocimiento de la revisión inicial; y, entonces, se

---

<sup>103</sup> La misma normativa en lo que respecta al módulo hematológico (C.2.1.4.1) establece que “el modelo adaptativo procesa automáticamente en ADAMS dos *marcadores* primarios, la concentración de hemoglobina (HGB) y el índice de estimulación OFF-score (OFFS), y dos *marcadores* secundarios, el porcentaje de reticulocitos (% RET) y el puntaje anormal del perfil sanguíneo (ABPS). Un *Resultado Atípico en el Pasaporte* se genera cuando un valor HGB y/o OFFS del último control queda fuera de los rangos intraindividuales esperados. Además, el perfil longitudinal compuesto por (hasta) los últimos cinco valores válidos de HGB y/u OFFS también se considera un *Resultado Atípico en el Pasaporte* cuando se desvía de los rangos esperados, según lo determinado por el modelo adaptativo (secuencia de *Resultado Atípico en el Pasaporte*). Un *Resultado Atípico en el Pasaporte* solo es generado por el modelo adaptativo basado en los valores de los *marcadores* primarios HGB y OFFS o la secuencia de los mismos.

constituye un panel de tres expertos independientes que han revisado el pasaporte (C.3.3). En ese contexto, se prevé que tales expertos puedan solicitar más información para llevar a cabo su revisión, vgr., aquella relacionada con afecciones médicas del deportista, el calendario de la competición y/o resultados de análisis de muestra(s).

El Estándar internacional de gestión de resultados establece que solo se puede llegar a un resultado adverso en el pasaporte si los tres expertos emiten una opinión de “probable dopaje” (C.3.4), evaluando el pasaporte del deportista con los mismos datos. Luego, la llamada Unidad de Gestión del Pasaporte del *Deportista* informa al custodio del pasaporte si mantiene la opinión de los expertos después de revisar toda la información disponible en esa etapa (C.1.3.e). También se notifica al deportista el Resultado Adverso en el Pasaporte y se le da la oportunidad de dar explicaciones (C.1.3.f). Finalmente, si luego de del análisis de las explicaciones del deportista, los expertos mantienen su conclusión unánime de que es muy probable que el deportista haya usado una sustancia o método prohibido, entonces se declara la infracción del deportista de las normas antidopaje (C.1.3.g).

En resumen, en el caso de un resultado atípico en el pasaporte biológico de un deportista, un experto debe evaluar si ese resultado es consecuencia del dopaje o de una condición fisiológica o patológica normal, decidiendo el caso conforme al estándar de la probabilidad prevaleciente. Si su conclusión es favorable al dopaje, toda la información se envía a dos expertos adicionales (que no se conocen ni comunican entre ellos ni con el primero) para que realicen la misma evaluación y con el mismo estándar de probabilidad prevaleciente. Si su conclusión es coincidente con el primero se comunica a la Unidad de Gestión del Pasaporte y si esta coincide también en la evaluación se comunica al deportista y se le permite aportar información de descargo. En caso de que no la aporte, se confirma la decisión de dopaje, y, si aporta información, esta es evaluada de nuevo por los expertos junto con los resultados del pasaporte biológico. Si de nuevo emiten por unanimidad una conclusión de “probable dopaje” se entiende que hay un resultado adverso, lo que significa que está probado el dopaje. Con ello, podemos identificar un estándar de prueba con dos requisitos: 1) que la

hipótesis de que el resultado atípico sea consecuencia del dopaje sea más probable que la hipótesis de que sea consecuencia de una condición fisiológica o patológica normal, y 2) que por unanimidad los expertos consideren cumplida esa primera condición.

El Estándar internacional de gestión de resultados prevé también la posibilidad de que la Unidad de Gestión del Pasaporte Biológico solicite la revisión de un experto a pesar de la ausencia de un resultado atípico en el pasaporte, por ejemplo, porque se hizo una tarea de inteligencia con relación al deportista en cuestión o por cualquier nivel anormal y/o variación de marcador(es)<sup>104</sup>. En ese caso, dado que el punto de partida es menos claro que en el anterior, porque no hay resultado atípico, los estándares de prueba para el experto individual y para el panel de expertos son más elevados. Así, el artículo C.2.2.5.1 del Estándar Internacional de Gestión de Resultados exige al experto en cuestión que dictamine que:

- a. es *muy probable* que los datos del pasaporte sean el resultado del uso de una sustancia o método prohibido y
- b. que es *muy poco probable* que el pasaporte sea el resultado de una condición fisiológica o patológica normal

Si el experto encuentra pruebas suficientes sobre ambas cuestiones, entonces el pasaporte es revisado también por otros dos expertos para conformar un panel de tres expertos revisores. Dicho panel de expertos, según el artículo C.3.5 de la misma normativa, debe concluir unánimemente que:

- a. es *muy probable* que el pasaporte sea el resultado del uso de una sustancia o método prohibido
- b. que *no existe una hipótesis razonablemente concebible* según la cual el pasaporte sea el resultado de una condición fisiológica normal
- c. *altamente improbable* de que sea el resultado de una condición patológica

Por tanto, fijándonos, por resumir, en esta última previsión, el estándar de prueba para un resultado adverso en el pasaporte biológico en ausencia de resultado atípico, requiere también dos condiciones: 1) que la hipótesis de la culpabilidad (que sostiene que el pasaporte es el resultado del uso de una sustancia o método

---

<sup>104</sup> Estos supuestos y otros los prevé el Estándar Internacional de Gestión de Resultados, artículo C.2.2.4.1.

prohibido) sea muy probable y, a la vez, que no exista una hipótesis razonablemente concebible de que el pasaporte sea resultado de una condición fisiológica normal que sea altamente improbable que sea el resultado de una condición patológica; y 2) que por unanimidad los expertos consideren cumplidas estas primeras condiciones. A pesar de la evidente vaguedad de las expresiones “muy probable” y “altamente improbable”, parece claro que en este caso se requiere que la diferencia de probabilidad entre las hipótesis acusatoria y defensivas sea mayor que la requerida en el caso de la presencia de resultados atípicos.

Y ahora, la pregunta del millón, todo el análisis delineado anteriormente para la decisión de un conjunto de expertos sobre el probable dopaje de un deportista, ¿es suficiente para satisfacer el estándar de prueba *jurídico* para la sanción administrativa? Suponiendo una respuesta afirmativa a esta pregunta, la autoridad administrativa incoó un expediente sancionador al Sr. Salas en el caso resuelto en apelación por la sentencia de apelación 39/2021 de la Audiencia Nacional que es objeto de este informe:

En el inicio del procedimiento sancionador se especificaba que, en el marco de las diligencias previas, recibida la evaluación del pasaporte biológico del interesado emitida el 28 de febrero de 2018 por el Laboratorio Suizo de Control de Dopaje, el Jefe del Departamento de Control de Dopaje dirigió el 10 de abril de 2018 una comunicación al sr. Salas diciéndole que se había recibido *la evaluación de su pasaporte biológico* por parte de un *panel formado por tres expertos científicos independientes*, que declaraban, de forma unánime, que su perfil hematológico indicaba que era «[a]ltamente probable que usted haya usado una Sustancia y/o Método Prohibido y que no es probable que sea resultado de cualquier otra causa [...]». El acuerdo de incoación del expediente parte de esta premisa y así se refleja en su antecedente primero. (F.J. 3ª de la sentencia 39/2021 de la Audiencia Nacional).

Normalmente la pregunta sobre la satisfacción o no del estándar de prueba jurídico es muy difícil de responder porque no es extraño en los sistemas de tradición romano-germánica que las reglas procesales no especifiquen un concreto estándar de prueba y se limiten a hacer referencia a la sana crítica como sistema de valoración del conjunto de pruebas (sin indicar un umbral de suficiencia probatoria para la decisión). Así, por ejemplo, el artículo 39.4 de la Ley Orgánica 3/2013, ya derogada pero aplicable a los hechos resueltos por la sentencia de la Audiencia Nacional 39/2021, establecía que:

En el procedimiento sancionador en materia de dopaje la Administración y la persona afectada por aquél podrán servirse de todos los medios de prueba admisibles en derecho, incluido el pasaporte biológico si existiesen datos sobre el mismo. *Dichas pruebas deberán valorarse de modo conjunto de acuerdo con las reglas de la sana crítica* (cursiva añadida).

Y un texto parecido, con un añadido importante que veremos más tarde, se encuentra en el artículo 39.2 de la vigente Ley Orgánica 11/2021, de 28 de diciembre, de lucha contra el dopaje en el deporte.

La mejor interpretación que se puede hacer de la sana crítica es que se trata de una apelación a los criterios de la racionalidad general aplicados, en este contexto, al razonamiento inductivo. Esto son, en suma, el tipo de criterios que analizamos en el epígrafe segundo sobre las hipótesis a comparar, las diversas pruebas que han sido admitidas y practicadas y las conexiones o enlaces para relacionar las proposiciones en juego. Sucede, sin embargo, que la apelación a esos criterios de la sana crítica no responde a la pregunta por la suficiencia probatoria: toda decisión sobre los hechos durante un procedimiento administrativo o judicial debe adoptarse respetando las reglas de la sana crítica (i.e., los criterios de racionalidad), las leyes de la lógica y de la ciencia, etc.; pero, a pesar de ello, por ejemplo, no se exige lo mismo para una decisión inicial del procedimiento, para una medida cautelar o para la decisión final (Ferrer Beltrán, 2022: 434).

Necesitamos, pues, determinar cuál es el estándar de prueba aplicable a la decisión jurídica para considerar probado el dopaje. La Audiencia Nacional se limita a apelar a la falta de certeza, exigiendo lo imposible. Además, ni siquiera realiza un análisis de qué pruebas fueron consideradas, de qué manera las pruebas existentes daban o no sustento a la hipótesis del dopaje del deportista, y, por supuesto, tampoco consideró analizar la información sobre el descarte de otras hipótesis. Es más, en la sentencia no hay alusión alguna a las pruebas que sustentarían la hipótesis contraria (no-dopaje) con la que debía compararse mediante las “reglas de la sana crítica” la hipótesis del dopaje.

Ahora bien, como he dicho, no basta con comparar las hipótesis, es necesario también decidir entre ellas en función de aquella que alcance el nivel de

suficiencia exigible (Vázquez, 2015: 267)<sup>105</sup>. Y, desde luego, las “reglas de la sana crítica” no establecen en sí mismas un umbral sobre la suficiencia de las pruebas y, por lo tanto, habría que buscarlo en otras fuentes.

En tal contexto se ha concebido que el estándar de prueba penal es aplicable al ámbito administrativo sancionador dado que en ambos casos se ejercería el ius puniendi estatal para sancionar a los ciudadanos<sup>106</sup>. La aplicación al ámbito administrativo sancionador del derecho a la presunción de inocencia y la asunción de que este derecho exige, para su derrota, el cumplimiento del estándar del más allá de toda duda razonable, completarían el argumento a favor de la exigencia de ese preciso estándar, también en el ámbito de las sanciones administrativas.

Es claro que esto no fue lo que resolvió la Audiencia Nacional (que requirió, en cambio, la certeza), pero supongamos por un momento que efectivamente la Audiencia Nacional debió considerar si los hechos del caso estaban probados “más allá de toda duda razonable”. Sobre este punto, vale la pena traer a colación la STS 680/222, que determina qué condiciones deben darse para que una hipótesis alternativa a la de la culpabilidad pueda fundar una duda razonable respecto de esta:

La cuestión, por tanto, obliga despejar si en el caso la decisión absolutoria del Tribunal Superior se basó en una duda razonable. *No en cualquier duda extraída de una hipótesis abstractamente posible construida* mediante una inverosímil combinación de circunstancias. Como apuntábamos, la duda razonable debe fundarse en razones intersubjetivamente compartibles y justificarse a la luz de las circunstancias del caso.

Para ello, la hipótesis alternativa sobre la que se sustente deberá: primero, ofrecer una explicación que abarque todos los datos constatados que sean relevantes; segundo, las consecuencias que de tal hipótesis se deriven no podrán ser incompatibles con los datos existentes; y, tercero, deberá resistir, al menos, intentos de falsación proveniente de las pruebas que en el proceso se han tenido por acreditadas.

---

<sup>105</sup> Esto evita tener que decidir entre dos hipótesis que están respaldadas por pruebas débiles, un escenario en el que nos veríamos abocados a decidir entre dos malas reconstrucciones de los hechos. Véase al respecto las críticas de Laudan (2007) a la llamada “inferencia a la mejor explicación”, tanto para la aceptación y rechazo de teorías científicas como sustituto del estándar de prueba en materia jurídica.

<sup>106</sup> Como afirma Letelier (2018: 218): “La afirmación de que ambas sanciones son manifestaciones del mismo ius puniendi –a pesar del simplismo lacónico de su formulación– ha campado con pocas matizaciones en la doctrina y la jurisprudencia comparada. Dicha identidad provocaría una suerte de asimilación regulatoria de ambos tipos de sanciones en materia de garantías o, en subsidio, una aplicación atenuada o matizada de las garantías penales al ámbito de las sanciones administrativas.”

Desgraciadamente, en cambio, la Audiencia Nacional simplemente consideró que si no le ofrecían certezas acerca del dopaje entonces había dudas, lo que ni siquiera es eso que el Tribunal Supremo señala como una *hipótesis abstractamente posible*, como ya se argumentó a lo largo de estas páginas. La Audiencia Nacional mucho menos analizó si la hipótesis alternativa del deportista integraba explicativamente los datos relevantes, etc. No puede concluirse otra cosa que estamos ante una decisión judicial que no está debidamente justificada ni en cuanto a la valoración de las pruebas ni respecto al cumplimiento del umbral de suficiencia probatoria.

Es más, la propia aplicabilidad automática del estándar del más allá de toda duda razonable al ámbito administrativo sancionador es cuestionada con muy buenas razones por no pocos penalistas, administrativistas y especialistas en razonamiento probatorio. Como hemos visto en un epígrafe anterior, las razones para situar el estándar de prueba en un punto u otro de exigencia probatoria tienen que ver con la gravedad de la afectación a bienes jurídicos en caso de error en la decisión (tanto de falsos positivos como de falsos negativos). Y puede sostenerse muy razonablemente que la gravedad de la afectación al ciudadano en el caso de una decisión penal errónea es mayor que la gravedad de la afectación en el caso de una decisión administrativa sancionadora en el ámbito del deporte. En esa línea, por ejemplo, Duce (2018: 92):

En el caso de Chile se pueden observar diferencias importantes entre un proceso penal y un proceso administrativo sancionatorio que justifican que el estándar de prueba exigido sea diferente. Así, es posible constatar que el riesgo y la seriedad de las consecuencias que surgen producto de una decisión de órganos administrativos son mucho más bajos que los de una sanción penal, en varios sentidos. En primer lugar, a diferencia del sistema penal, las sanciones administrativas no establecen como forma de reacción la privación de libertad de sus infractores, pena que hoy en día constituye el mayor gravamen dispuesto en nuestro sistema. En segundo lugar, las medidas que establecen los órganos administrativos, por su propia naturaleza, no producen el mismo nivel de afectación a la dignidad y honor de las personas como tampoco el nivel de reproche social que lleva aparejado una sanción penal.

Y, por otro lado, debe considerarse también la afectación a los bienes sociales protegidos por el derecho administrativo en el caso de absoluciones erróneas (o falsos negativos). Por ello, Letelier (2018: 222 - 223) se pregunta:

¿es, entonces, justificable defraudar o sacrificar las reglas de conducta impuesta por las reglas jurídicas en asuntos de derecho administrativo por una especial aversión a la sanción? ¿Es justificable una preferencia por un statu quo de no sanción (libertad o inocencia) en materia administrativa, como el que rige en materia penal?

La respuesta a estas preguntas, creo, debe ser, en principio, negativa. En efecto, es posible apreciar en la realidad social un número importantísimo de sectores donde un especial sacrificio a la aplicación coactiva de la regla de conducta administrativa no se encontraría para nada justificado, y menos incluso si de ello se obtiene una decisión de calidad epistémica deficiente. Me refiero, por ejemplo, a las reglas de conducta que pretenden corregir las desviaciones o las fallas de mercado que se dan en la prestación de servicios públicos ...

En casos como estos, el statu quo existente antes de la regla de conducta –cuyo *enforcement* es la sanción administrativa y cuya defraudación es la tolerancia de falsos negativos– es el peor escenario posible en términos de bienestar social. Asimetría informativa, monopolios naturales, beneficios inesperados (*windfall profits*), presencia de externalidades, continuidad o disponibilidad del servicio, comportamientos anticompetitivos, bienes públicos o riesgo moral, desigualdades en los poderes de negociación, eventos de escasez y racionamiento, son solo algunas de las circunstancias que muestran no solo la imperiosa necesidad de regulación jurídica en estos sectores, sino también la inexistencia de una aversión a la aplicación coactiva de estas regulaciones, aun asumiendo el costo de los errores –ya minimizados– que arroja una forma eficiente de reconstrucción epistémica.

En efecto, en el ámbito administrativo se busca también corregir desviaciones, como los comportamientos anticompetitivos, en el ámbito del deporte esto redundaría en la promoción del desarrollo de competiciones deportivas en términos de igualdad, lealtad y de acuerdo con las capacidades naturales y preparación de los deportistas. Objetivos nada desdeñables e indispensables para considerar en la identificación de las exigencias probatorias con las que se distribuye el riesgo de error.

Pero es que, más allá de estas consideraciones teóricas para impugnar la aplicabilidad automática del estándar de prueba del más allá de toda duda razonable al procedimiento administrativo sancionador en casos de dopaje, hay que recordar que el artículo 39.2 de la Ley 11/2021 se remite al Código mundial antidopaje, en los siguientes términos:

En el procedimiento sancionador en materia de dopaje, la Administración y la persona afectada por aquél podrán servirse de todos los medios de prueba admisibles en derecho, incluido el pasaporte biológico, si existiesen datos sobre el mismo. Dichas pruebas deberán valorarse de modo conjunto de acuerdo con las reglas de la sana crítica, *de acuerdo con los principios y criterios de interpretación establecidos en el Código Mundial Antidopaje* (cursiva añadida).

Y el Código mundial antidopaje, en su artículo 3.1, sí se pronuncia explícitamente sobre el estándar de prueba aplicable:



The *Anti-Doping Organization* shall have the burden of establishing that an anti-doping rule violation has occurred. The standard of proof shall be whether the *Anti-Doping Organization* has established an anti-doping rule violation to the comfortable satisfaction of the hearing panel, bearing in mind the seriousness of the allegation which is made. *This standard of proof in all cases is greater than a mere balance of probability but less than proof beyond a reasonable doubt* (cursiva añadida).

Así pues, el estándar de prueba en todos los casos es más exigente que la mera probabilidad prevaleciente y menos exigente que el más allá de toda duda razonable<sup>107</sup>. No se trata de una formulación muy precisa, limitándose a fijar un espacio de exigencia probatoria dentro del que deberían fijarse los estándares de prueba para cada tipo de casos. Por eso, es necesario acudir a normativas más específicas. En esa línea, el artículo 40 de la LO 11/2021 establece, en su apartado 2º que:

Sin perjuicio de lo dispuesto en el artículo anterior, en la tramitación de los procedimientos sancionadores como consecuencia de resultados adversos en el pasaporte biológico, se respetará, en todo caso, el contenido esencial de las Normas y los Estándares Internacionales sobre gestión de resultados, sobre controles e investigaciones y sobre laboratorios de la Agencia Mundial Antidopaje.

En conclusión, resulta aquí de aplicación de acuerdo con el derecho español lo previsto sobre los estándares de prueba por el Estándar internacional de gestión de resultados<sup>108</sup>, que ha sido expuesto algunas páginas atrás para los casos en

---

<sup>107</sup> Desafortunadamente, la versión castellana del Código tiene una mala traducción del artículo 3.1. y, en lugar de hacer referencia al “estándar de prueba”, lo hace al “criterio de valoración de la prueba”: “Recaerá sobre la Organización Antidopaje la carga de probar que se ha producido una infracción de las normas antidopaje. Con respecto al *criterio de valoración*, la Organización Antidopaje deberá acreditar la infracción de la norma a plena satisfacción del tribunal de expertos, teniendo en cuenta la gravedad de la acusación que se formula. *Dicho criterio, en todo caso, no consistirá en una mera ponderación de probabilidades, pero tampoco será necesaria una demostración que excluya toda duda razonable* (cursivas añadidas).”

<sup>108</sup> La situación legal no es tan clara en la Ley Orgánica 3/2013, vigente para los hechos resueltos por la sentencia 39/2021 de la Audiencia Nacional. Esa Ley no contiene una expresa remisión al Código mundial Antidopaje en materia de prueba, como hemos visto que sí contiene la Ley Orgánica 11/2021. Y tampoco hace expresa mención del estándar de prueba aplicable a las decisiones sobre los hechos en materia de dopaje. Pero las razones expuestas en este informe muestran la inadecuación de la exigencia del estándar de prueba penal para esos casos, el Código mundial ya lo excluía, y el preámbulo de la propia LO 3/2013 ya preveía también lo siguiente:

Asimismo, las sucesivas modificaciones operadas en este ámbito en el plano internacional, con el que España ha venido asumiendo un alto compromiso en la lucha contra el dopaje y que se tradujo, entre otros aspectos, en la ratificación, pocos meses después de la promulgación de la Ley, de la Convención Internacional contra el dopaje en el deporte de la Unesco, hacían ineludible abordar estas reformas. Así, este texto impone a los Estados que lo hayan ratificado una serie de obligaciones en materia de lucha contra el dopaje, no siendo la menos importante aquella que obliga a los Estados miembros a garantizar la eficacia del Código Mundial Antidopaje.

que se den resultados atípicos en el pasaporte biológico y en ausencia de esos resultados atípicos<sup>109</sup>. Ni el Código mundial ni los Estándares internacionales hacen alusión a los fundamentos de tales niveles exigencia probatoria, que a final de cuentas son fruto de decisiones político-morales<sup>110</sup>. Ahora bien, ¿qué razones podrían contar para tomar la decisión sobre cuál sería el nivel de exigencia adecuado? Si tomamos en cuenta las señaladas por Ferrer Beltrán (2021: 441-446) y consideramos el ámbito del dopaje, se puede señalar:

- Ciertas dificultades probatorias, sobre todo ante métodos sofisticados de dopaje, aconsejaría bajar el estándar de prueba. Lo que quizá no se suele dar con métodos más “tradicionales” como la ingesta de ciertas sustancias<sup>111</sup>.
- La gravedad de no sancionar una conducta antideportiva como el dopaje (los falsos negativos) aconsejaría bajar el estándar probatorio, cuya incidencia en la salud de los deportistas valdría la pena también considerar.
- Las reglas de decisión científicas que ya buscan evitar los falsos positivos, como las que se mencionaban párrafos arriba, aconsejan reducir el estándar. Es decir, si ya se tienen reglas que distribuyen el riesgo del error,

---

Dado, pues, que España había ya ratificado la Convención Internacional contra el dopaje en el deporte de la Unesco y que esta obliga a los estados miembros a garantizar la eficacia del Código mundial, creo que puede concluirse que, en ausencia de previsión legal expresa sobre el estándar de prueba aplicable, debía también acudir a lo previsto en el Código mundial y los documentos internacionales a los que este se remite. Así se establece también muy claramente, de nuevo, en el preámbulo de la LO 3/2013, diciendo lo siguiente:

[E]l Código Mundial Antidopaje debe configurarse como un elemento central de interpretación de las normas de la ley que se ocupan de esta cuestión, de manera que las dudas que su aplicación pueda plantear deberán resolverse a la luz de los preceptos, comentarios y principios del Código.

<sup>109</sup> Es muy interesante también notar que el propio Código mundial antidopaje prevé un estándar de prueba más bajo (la probabilidad prevaleciente) para el deportista que quiera probar circunstancias o hechos específicos que sustentarían una hipótesis de no-dopaje.

<sup>110</sup> En los comentarios al artículo 3.1 lo único que se dice, y que no resulta informativo, es que el estándar de prueba es similar al que se exige en la mayoría de los países en casos de conducta profesional indebida.

<sup>111</sup> La Ley Orgánica 3/2013, de 20 de junio, de protección de la salud del deportista y lucha contra el dopaje en la actividad deportiva, en su artículo 39, preveía que “*un resultado analítico adverso en un control de dopaje constituirá prueba de cargo suficiente a los efectos de considerar existentes las infracciones tipificadas en el artículo 22.1.a) y b) de esta Ley.*

La propia normativa prevé que, por ejemplo, para las sustancias sujetas a un umbral, el “resultado Analítico Adverso o Resultado Anómalo del hallazgo en la *Muestra "A"* se basará en el promedio de los valores medidos de análisis (p. ej., concentraciones) o cociente calculado a partir de los promedios de los valores analíticos medidos (p. ej., concentraciones, alturas del pico o área del cromatograma) de tres Alícuotas.”

un estándar de prueba también exigente podría dificultar fuertemente la prueba de los hechos.

- La gravedad del error en caso de un falso positivo para el deportista y su carrera profesional sería, en cambio, una razón para subir el estándar de prueba.

Esta última cuestión, desde luego, estará muy relacionada con el tipo de sanciones que están previstas, pues cuanto mayor sea la sanción asociada, mayor será también el costo del riesgo del error para el deportista. No es lo mismo, la anulación de los resultados de una competición o prueba que la anulación de todos los resultados de todas las pruebas durante un evento; o la imposición de la pérdida de una medalla y la imposición de un periodo de inhabilitación; o una sanción económica y la divulgación a la ciudadanía del dopaje del deportista, etc. En todo caso, el tipo de sanciones previstas me parece que justifica ampliamente la opinión de Duce y Lettelier sobre lo lejos que queda su gravedad de las sanciones penales, como la pérdida de libertad. Y, en esa línea se sitúa la regulación del Código mundial antidopaje, al que la Ley Orgánica 11/2021 se remite. Es por ello que puede decirse que la sentencia 39/2021 de la Audiencia Nacional no sólo no está debidamente motivada en lo que hace a la valoración de la prueba, sino que es también contraria a las reglas procesales aplicables al caso en lo que al estándar de prueba concierne.

### **Bibliografía**

- Aitken, Colin; Taroni, Franco; y Bozza, Silvia (2022): "Evidence, probability and relative plausibility", en *Evidence and Proof International Journal*, vol. 26, no. 4.
- Aliste Santos, Tomás (2021): "Probabilidad y certeza: Análisis crítico sobre la aplicación del teorema de Bayes en los estados unidos para la motivación de la prueba de los hechos susceptibles de enjuiciamiento criminal", en *Revista ítalo-española de derecho procesal*, vol. 1.
- Allen, Ronald (1981): "Presumptions in Civil Actions Reconsidered", en *Iowa Law Review*, 66.

- American Academy of Pediatrics (1993): "Shaken Baby Syndrome: Inflicted Cerebral Trauma", en *Pediatrics*, 92.
- Bauer, H.H., 1992: *Scientific Literacy and the Myth of the Scientific Method*, Chicago: University of Illinois Press.
- Black, Max (1984): *Induction and Probability*: Macmillan Publishing Company Inc. Citado por la traducción al castellano de Casan, P. y Beneyto, R.: *Inducción y probabilidad*, Madrid: Cátedra.
- Carlizzi, Gaetano (2019): *La valutazione della prova scientifica*, Milano: Giuffrè Francis Lefebvre.
- Champod, Ch. y Vuille, J., 2010: "Scientific Evidence in Europe. Admissibility, Appraisal and Equality of Arms, Comparative Study on Scientific Evidence Drawn Up to the Bureau of the Council of Europe's: European Committee on Crime Problems", en *International Commentary on Evidence*, vol. 9, no 1.
- Cohen, Jonathan, 1989: *Introduction to the Philosophy of Induction and Probability*, Oxford: Oxford University Press.
- Cordeiro, Alessandro, 2013: "Neurociencias y derecho penal desde el prisma de la dimensión procesal". Citado por la traducción al castellano de Nieva Fenoll, J., en Taruffo, M. y Nieva Fenoll, J.: *Neurociencia y proceso judicial*, Barcelona: Marcial Pons.
- De La Oliva, Andrés (2002): "La sentencia", en De La Oliva, Andrés, Aragonese Martínez, Sara, Hinojosa Segovia, Rafael, Muerza Esparza, Julio y Tomás García, José Antonio: *Derecho procesal penal*, 5ª ed., Madrid. Centro de Estudios Ramón Areces.
- De Finetti, b. (1970): *Teoria delle probabilità*, Torino: Einaudi.
- Dei Vecchi, Diego (2020): "La diversidad de lo probable: apuntes sobre prueba y probabilidad", en *Revista de Brasileira de Direito Procesal*, año 28, no, 109.
- Dei Vecchi, Diego (2020b): *Los confines pragmáticos del razonamiento probatorio*, Lima: ZELA y Editorial CEJI.
- Denti, V., 1972: "Cientificidad de la prueba en relación principalmente con los dictámenes periciales y la libertad de apreciación del juzgador". Citado por la traducción al castellano de Oñate L., S., en *Revista de Derecho Procesal Iberoamericana*, no. 2 – 3.

- Donohoe, Mark (1998): "Evidence Based Medicine and Shaken Baby Syndrome. Part 1. Literature review 1966-1998", en *American Journal of Forensic Medicine and Pathology*, 24.
- Duce, Mauricio (2018): "Reflexiones sobre el proceso sancionatorio administrativo chileno: debido proceso, estándar de convicción (prueba) y el alcance del sistema recursivo", en *Diritto Penale Contemporaneo. Rivista Trimestrale*, no.2.
- Duce, Mauricio (2022): "Los errores. La aplicación al caso concreto y los sesgos cognitivos de los expertos", en Vázquez, Carmen: *Manual de prueba pericial*, México: Suprema Corte de Justicia de la Nación.
- Dwyer, Deidre (2008): *Judicial Assessment of Expert Evidence*, Cambridge: Cambridge University Press.
- Edmond, Gary (2012): "Is Reliability Sufficient? The Law Commission and Expert Evidence in International and Interdisciplinary Perspective (Part 1)", en *The International Journal of Evidence & Proof*, no. 16.
- Edmond, Gary (2020): "Cuando el derecho es poco fiable: Respuestas jurídicas a la prueba de huellas dactilares latentes", en *Quaestio facti. Revista Internacional sobre Razonamiento Probatorio*, vol. 1.
- Ferrajoli, Luigi (1997): *Derecho y razón. Teoría del garantismo penal*, Barcelona: Editorial Trotta.
- Ferrer Beltrán, Jordi (2003): *Prueba y verdad en el derecho*, Madrid - Barcelona: Marcial Pons.
- Ferrer Beltrán, Jordi (2007): *La valoración racional de la prueba*, Madrid - Barcelona: Marcial Pons.
- Ferrer Beltrán, Jordi (2021): *Prueba sin convicción. Estándares de prueba y debido proceso*, Madrid - Barcelona: Marcial Pons.
- Ferrer Beltrán, Jordi (Coord.), (2022): *Manual de razonamiento probatorio*, México: Suprema Corte de Justicia de la Nación.
- Gascón Abellán, Marina (2004): *Los hechos en el derecho. Bases argumentales de la prueba*, 2ª ed., Madrid - Barcelona: Marcial Pons.
- Gascón A., M.; Lucena Molina, J.J.; González Rodríguez, J. (2010): "Razones científico-jurídicas para valorar la prueba científica. Una argumentación multidisciplinaria", en *La Ley*, no. 5619.

- Giles, Alejo J. (2021): Los usos de la estadística en los casos de discriminación: un análisis de tres sentencias de la Suprema Corte de Justicia de la Nación, en Vázquez, C: Ciencia y justicia. El conocimiento experto en la Suprema Corte de Justicia de la Nación, México: Centro de Estudios Constitucionales.
- González Garay, A. (2014): “La Red Regional Mexicana Cochrane”, en Revista CONAMED, vol. 19, núm 1, pp. 43-45.
- González Lagier, Daniel (2022): “Inferencia probatoria y valoración conjunta de la prueba”, en Ferrer Beltrán, Jordi (Coord.): Manual de razonamiento probatorio: Suprema Corte de Justicia de la Nación.
- González Rodríguez, Joaquín (2022): “Las ciencias forenses a la luz del ADN”, en Vázquez, Carmen: Manual de prueba pericial, México: Suprema Corte de Justicia de la Nación.
- Haack, Susan (2013): Putting Philosophy to Work. Inquiry and its Place in Culture, New York: Prometheus Books.
- Haack, Susan (2020): Filosofía del derecho y de la prueba. Perspectivas pragmatistas, Madrid - Barcelona: Marcial Pons.
- Haack, Susan (2021): “El largo brazo del sentido común. En lugar de una teoría del método científico”, en Vázquez, C: Ciencia y justicia. El conocimiento experto en la Suprema Corte de Justicia de la Nación, México: Centro de Estudios Constitucionales.
- Hastie, Reid (1993): “Algebraic Models on Decision Processes”, en Hastie, R. (Ed.): Inside the Juror: The Psychology of Juror Decision Making, Cambridge: Cambridge University Press.
- Howick, J. (2011): The Philosophy of Evidence Based Medicine, UK: Wiley-Blackwell.
- Igartua Salaverría, Juan (2021): Indicios, duda razonable, prueba científica. Perspectivas sobre la prueba en el proceso penal, Valencia: Tirant lo Blanch.
- Krumm, Bastien y Faiss, Raphael (2021): “Factors Confounding the Athlete Biological Passport: A Systematic Narrative Review”, en Sports Medicine Open.

- Laudan, Larry (2006): Verdad, error y proceso penal. Un ensayo sobre epistemología jurídica, Madrid – Barcelona: Marcial Pons. Citado por la traducción de Carmen Vázquez y Edgar Aguilera.
- Laudan, Larry (2007): “Aliados extraños: la inferencia a la mejor explicación y el estándar de prueba penal”, en Problema. Anuario de Filosofía y Teoría del Derecho, no. 1.
- Law Commission (2009): “The Admissibility of Expert Evidence in Criminal Proceedings in England and Wales. A New Approach to the Determination of Evidentiary Reliability”, Consultation Paper, no. 190.
- Letelier, Raúl (2018): “El precio del statu quo. Sobre el estándar probatorio en las sanciones administrativas, en Revista de derecho”, volumen XXXI, no. 1.
- Limardo, Alan (2021): “Repensando las máximas de la experiencia”, en Quaestio facti. Revista Internacional sobre Razonamiento Probatorio, vol. 2.
- Mayo, Deborah (1996): Error and the Growth of Experimental Knowledge, Chicago: University of Chicago Press.
- McCormick, Charles (1954): Handbook on the Law of Evidence, St. Paul Minnesota: West.
- Medina Alcoz, Luis (2016): “Los hechos en el derecho administrativo. Una aproximación”, en Revista española de derecho administrativo, no. 177.
- Mnookin, Jennifer L.; Cole, Simon A.; Dror, Itiel E.; Fisher, Barry A. J.; Houck, Max M.; Inman, Keith; Kaye, David H.; Koehler, Jonathan J.; Langenburg, Glenn; Risinger, D. Michael; Rudin, Norah; Siegel, Jay; y. Stoney, David A (2011): “The Need for a Research Culture in the Forensic Sciences”, en UCLA Law Review, no. 58.
- Montero Aroca, Juan (2006): “La valoración de la prueba como garantía en el proceso civil, III Congreso panameño de derecho procesal”, Instituto Colombo Panameño de derecho procesal.
- Murphy, P. (1980): Murphy on Evidence, 7ª ed., London: Blackstone Press Limited.
- NAS Report (2009): “Strengthening Forensic Science in the United States: A Path Forward”, National Academy of Science.
- Nieva, Jordi (2010): La valoración de la prueba, Barcelona: Marcial Pons.

- PCAST Report (2014): Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, President's Council of Advisors on Science and Technology.
- Pérez Gil, Julio (2010): El conocimiento científico en el proceso civil: Ciencia y tecnología en tela de juicio, Valencia: Tirant lo Blanch.
- Popper, Karl (1962): La lógica de la investigación científica, 2ª ed., Madrid: Editorial Tecnos. Citado por la traducción al castellano de Sánchez de Zavala, Víctor (2008).
- Posner, R.A. (2001): Frontiers of Legal Theory, Cambridge: Harvard University Press.
- Prieto, Lourdes y Carracedo, Ángel (2022): "La prueba de ADN", en Vázquez, Carmen: Manual de prueba pericial, México: Suprema Corte de Justicia de la Nación.
- Ramírez Ortiz, José Luis (2023): "¿Son fiables los testimonios policiales?", en Noelia Corral Maraver, Deborah Isabel García Magna, María Fátima Pérez Jiménez, Bertha Prado Manrique y Pablo Rando Casermeiro: Estudios político-criminales, jurídico-penales y criminológicos. Libro homenaje al profesor José Luis Díez Ripollés, Valencia: Tirant lo Blanc.
- Redmayne, Mike (2001): Expert Evidence and Criminal Justice, New York: Oxford University Press.
- Reed, Baron (2008): "Certainty", en Stanford Encyclopedia of Philosophy.
- Rivera Morales, R., 2011: La prueba. Un análisis racional y práctico, Madrid: Marcial Pons.
- Roberts, Paul (2013): "¿Fue el bebé sacudido? Prueba, pericia y epistemología jurídica en el proceso penal inglés", en Vázquez, Carmen (Ed.): Estándares de prueba y prueba científica, Madrid - Barcelona: Marcial Pons.
- Sottas, Pierre-Edouard; Robinson, Neil; Giraud, Sylvain; Taroni, Franco; Kamber, Matthias; Mangin, Patrice; y Saugy, Martial (2006): "Statistical Classification of Abnormal Blood Profiles in Athletes", en The International Journal of Biostatistics, vol. 2, no. 1.
- Sottas, Pierre-Edouard; Robinson, Neil; Saugy, Martial; y Niggli, Olivier (2008): "A forensic approach to the interpretation of blood doping markers", en Law, Probability and Risk, 7.



- Taruffo, Michele (2010): *Simplemente la verdad. El juez y la construcción de los hechos*, Madrid - Barcelona: Marcial Pons.
- Taruffo, Michele (2023): *Contribución al estudio de las máximas de la experiencia*, Madrid – Barcelona: Marcial Pons.
- Tuerkheimer, Deborah (2014): *Flawed convictions. “Shaken Baby Syndrome” and the Intertia of Injustice*, New York: Oxford University Press.
- Twinning, William (1994): *Rethinking Evidence. Exploratory Essays*, Illinois: Northwestern University Press.
- Ullmann – Margalit, Edna (2019): “Sobre la presunción”, en Mendonca, D; Ullman-Margalit, E.; y Aguiló Regla, J.: *Presunciones*, México: Ed. Fontamara.
- Vázquez, Carmen (2015): *De la prueba científica a la prueba pericial*, Madrid - Barcelona: Marcial Pons.
- Vázquez, Carmen (Coord.) (2022): *Manual de prueba pericial*, México: Suprema Corte de Justicia de la Nación.
- Vázquez, Carmen (2022b): “La conformación del conjunto de elementos de juicio: admisión de pruebas”, en Ferrer Beltrán, Jordi (Coord.): *Manual de razonamiento probatorio: Suprema Corte de Justicia de la Nación*.
- Villanueva ELA y Lezana FMÁ, 2017: «Introducción a la Medicina Basada en Evidencias», *Rev CONAMED*, 22(3), pp.152-154.



*Carmen Vázquez*