

APRENDIENDO DE LOS INFORMES DE LAS ACADEMIAS DE CIENCIAS NORTEAMERICANAS SOBRE LA CIENCIA FORENSE

JOSE JUAN LUCENA MOLINA
CORONEL (R) DE LA GUARDIA CIVIL

Fecha de recepción: 26/01/2022. Fecha de aceptación: 17/11/2022

RESUMEN

Un modo privilegiado de conocer el estado del arte en la ciencia forense en un país económicamente desarrollado consiste en consultar los informes emitidos por sus organismos especializados, ya sean públicos o privados. Así ocurre en los Estados Unidos, país en el que sus Academias Nacionales proporcionan valiosa información a los investigadores. En este siglo han sido publicados dos grandes informes que analizan la situación real de la ciencia forense en los laboratorios norteamericanos que se conocen por sus siglas: NRC 2009 y PCAST 2016. Esos informes han tenido una difusión internacional que ha permitido que sus contenidos hayan sido objeto de debate en foros científicos y profesionales de todo el mundo. En este artículo recogemos, primordialmente, el eco del informe PCAST 2016 entre científicos forenses principalmente europeos con gran influencia en ENFSI (*European Network of Forensic Science Institutes*). Sirve de guía un artículo publicado en la revista *Forensic Science International* en 2017 en el que sus autores realizan un detallado estudio crítico del mencionado informe.

Palabras clave: evaluación de la evidencia; identificación forense; relación de verosimilitudes.

ABSTRACT

A privileged way to know the state of the art in forensic science in an economically developed country is to consult the reports issued by its specialized agencies, whether public or private. This is the case in the United States, where its National Academies provide valuable information to researchers. In this century, two major reports have been published that analyse the real situation of forensic science in North American laboratories that are known by their acronyms: NRC 2009 and PCAST 2016. These reports have been disseminated internationally, which has allowed their contents to be discussed in scientific and professional forums around the world. In this article we collect, primarily, the echo of the PCAST 2016 report among mainly European forensic scientists with great influence in ENFSI (*European Network of Forensic Science Institutes*). It serves as a guide an article published in the journal *Forensic Science International* in 2017 in which its authors carry out a detailed critical study of the aforementioned report.

Keywords: evaluation of evidence; forensic identification; likelihood ratio.

1. LOS INFORMES DE LAS ACADEMIAS NACIONALES DE CIENCIAS, INGENIERÍA Y MEDICINA DE LOS ESTADOS UNIDOS SOBRE LA CIENCIA FORENSE EN EL SIGLO XXI

El *National Research Council* (NRC) es el ente operativo de las Academias estadounidenses de Ciencias, Ingeniería y Medicina. Las mencionadas Academias son instituciones privadas, sin ánimo de lucro, que proporcionan asesoramiento experto ante retos científicos relevantes para nuestra época, especialmente a quienes son responsables de configurar políticas acertadas para afrontarlos, informar a la opinión pública y promocionar el avance de las ciencias, la ingeniería y la medicina¹.

En nuestro siglo, los informes NRC 2009² y PCAST 2016³ han sido los más sobresalientes dentro del ámbito de las ciencias forenses. Ambos informes han ido acompañados de medidas especiales, avaladas por el Congreso de los Estados Unidos y dotadas de fondos federales específicos, encaminadas a mejorar las instituciones relacionadas con la práctica de la ciencia forense.

El informe PCAST 2016 ha sido objeto de crítica por algunos de los expertos en ciencias forenses de mayor prestigio científico en Europa⁴, aunque también lo han hecho algunos expertos no europeos de similar categoría científica, todos ellos muy cercanos a ENFSI (*European Network of Forensic Science Institutes*) al haber colaborado ordinariamente en sus iniciativas de carácter académico y con quienes pertenecen a los laboratorios de la red.

El orden lógico que el lector debería seguir para la comprensión óptima de este artículo es el orden temporal de publicación de los informes.

2. LA IMPORTANCIA DE LA LÓGICA EN LA CIENCIA FORENSE

Durante ya más de un siglo, algunos eminentes científicos han tratado de explicar en qué consiste evaluar los datos como evidencia. Esa pregunta ha merecido una especial atención de filósofos de la ciencia, estadísticos o científicos de distintas disciplinas.

Dentro de un contexto estrictamente forense, es decir, si consideramos la prueba científica en el marco lógico que construye un procedimiento probatorio en la jurisdicción penal, la evaluación de los datos como evidencia se ha resuelto —aplicando la inferencia estadística de distintas escuelas de esa disciplina— como una relación de verosimilitudes. Cabe citar la escuela fisheriana, que desarrolla la noción de verosimilitud (e.g. Royall 1997); o la bayesiana, que desarrolla la noción de probabilidad subjetiva (e.g. Taroni 2001).

1 <https://www.nationalacademies.org/>

2 National Research Council (2009). *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington DC. <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>

3 President's Council of Advisors on Science and Technology (2016). *Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, Washington DC. <https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports>

4 Evett, I.W. et al. (2017). Finding the way forward for forensic science in the US—A commentary on the PCAST report. *Forensic Science International*, 278, 16-23. <http://dx.doi.org/10.1016/j.forsci-int.2017.06.018>

Una de las obras que mejor ha descrito el marco lógico de referencia en el que se sitúa un cotejo criminalístico es la de B. Robertson y G.A. Vignaux en 1995 y titulada *Evaluating Forensic Science in the courtroom*. Recientemente se ha publicado la segunda edición (septiembre de 2016), en la que aparece un tercer coautor, C.E.H. Berger.

La estructura lógica que describen esos autores permite entender el papel del científico evaluando la evidencia como lo explica Evett (2017, pp. 17-18)⁵: (1) *se necesita considerar la evidencia dentro de un marco de circunstancias*; (2) *el valor probatorio de las observaciones no puede evaluarse a menos que se consideren, al menos, dos proposiciones*; y (3) *es necesario que el científico considere la probabilidad de las observaciones dado que cada una de las dos proposiciones sea cierta*.

Este aspecto lógico fundante de la noción de evaluación de un dato como evidencia es el que separa la ciencia forense actual de la ejercida durante el siglo pasado, en términos generales. Este principio lógico se considera por Evett como el avance científico más relevante acaecido en la ciencia forense desde su nacimiento (Evett, 2009)⁶. Por eso, entenderlo bien es el primer paso que hay que dar para que se comprenda, por ejemplo, por qué se propone la prueba de ADN como la regla de oro que hay que seguir para la ciencia forense en su conjunto en el siglo XXI.

No obstante, nos parece que lo primero que hay que recomendar a quien no esté familiarizado con estas nociones es que asimile las leyes y teoremas de la teoría de la probabilidad tal y como, por ejemplo, el matemático ruso Andréi Kolmogórov las enunció a principios del siglo XX dentro de la teoría de conjuntos.

Además, una vez comprendido el teorema de Bayes —que se deriva matemáticamente de esas leyes— y, más concretamente, su formulación en forma de apuestas, se está en condiciones de entender las denominadas falacias de raciocinio en las conclusiones de los cotejos criminalísticos. Esas falacias fueron brillantemente expuestas en una obra escrita por los estadísticos C.G.G. Aitken y F. Taroni en el año 2004, en su segunda edición, traducida al español y coeditada por el Ministerio del Interior y la editorial Dykinson, titulada *Estadística y evaluación de la evidencia para expertos forenses* en el año 2009⁷.

Aunque este apartado se titula *¿por qué es tan importante la lógica?*, pudiera parecer, tras lo expuesto, que es más relevante la teoría de la probabilidad. Sería un craso error llegar a esa conclusión porque la teoría de la probabilidad es una teoría matemática fundamentada en la ciencia que llamamos lógica. Su interpretación en un contexto forense —la probabilidad de que, por ejemplo, dos perfiles de ADN coincidan, uno procedente de un vestigio recogido en la escena del crimen tras una inspección ocular

5 <https://enfsi.eu/hall-fame/distinguished-forensic-scientist-award/2000-ian-evett/>

6 Como una forma asequible, en tiempo y grado de dificultad, de aproximarse hacia la comprensión de este nuevo paradigma para quien tenga conocimientos jurídicos y en español, recomendamos la lectura de este artículo: Gascón, M. et al. (2010). Razones científico-jurídicas para valorar la prueba científica: una argumentación multidisciplinar. Diario La Ley, 7481, Sección Doctrina, Año XXXI. https://www.academia.edu/27079041/Razones_cient%C3%ADfico_jur%C3%ADdicas_para_valorar_la_prueba_cient%C3%ADfica_una_argumentaci%C3%B3n_multidisciplinar

7 Al menos un ejemplar de esa obra en su primera y única edición en español, hasta el presente, se distribuyó oficialmente a todos los laboratorios pertenecientes a ENFSI (Red Europea de Institutos de Ciencias Forenses) de España.

y otro precedente de un sospechoso conocido— exige un análisis lógico cuidadoso, cuya estructura ha sido sintetizada por Evett en su artículo.

La interpretación de la probabilidad como evidencia estadística es la que ayuda a comprender por qué los paradigmas de la coincidencia y de la identificación —que se exponen a continuación— son erróneos desde la lógica y, consecuentemente, son errores científicos y matemáticos, aunque se hayan empleado con profusión en el pasado y se sigan utilizando en algunos laboratorios oficiales en el presente.

No es infrecuente encontrar en la literatura científica, o incluso jurídica, explicaciones sobre las falacias de lógica de raciocinio en cotejos criminalísticos con ejemplos que facilitan su intelección y que podríamos llamar triviales. Pero no sería buena táctica pedagógica minusvalorar la dificultad que la ciencia de la lógica, en general, tiene siempre para cualquier intelecto. Los buenos manuales de lógica aplicada, como podría calificarse la obra de Aitken y Taroni anteriormente mencionada, recuerdan que *la intuición no es buena consejera para calcular probabilidades*.

3. LOS PARADIGMAS DE LA COINCIDENCIA Y DE LA IDENTIFICACIÓN

Sin duda, las explicaciones que el artículo de Evett presenta bajo estas denominaciones constituyen un conjunto de razones de máxima importancia para ayudar a desenredar la madeja que se esconde bajo términos tan aparentemente inofensivos como “coincidencia”, “cotejo positivo” o “identificación”, comúnmente empleados en los informes forenses.

La relación de verosimilitudes es la contrapropuesta, ofrecida desde el avance reconocido por la comunidad científica especializada en inferencia lógica durante el transcurso del siglo pasado, a formas de entender las comparaciones criminalísticas y sus resultados desde lógicas obsoletas y erróneas.

A este respecto, resulta interesante la rotunda crítica que el informe PCAST recoge con respecto a formas de practicar la ciencia forense desvinculadas de la estadística. Aunque el artículo de Evett critica —fundada y contundentemente— el empleo de la estadística frecuentista para evaluar la evidencia de un cotejo criminalístico en el informe PCAST, el informe tiene especial interés para quien tenga una buena base jurídica pero esté poco familiarizado con la estadística —que suele ser lo corriente—. A nuestro juicio, las consideraciones de carácter estadístico del informe PCAST pueden ayudar a ese tipo de personas a valorar más la importancia del uso de la estadística en la ciencia forense. Concretamente, la insistencia del informe PCAST en que los métodos de comparación de características sean evaluados para conocer su precisión y fiabilidad, exigiendo que se cuantifiquen las tasas de falsos positivos y negativos en experimentos controlados, supone un avance enorme con respecto a los métodos en los que esa evaluación empírica sencillamente nunca ha existido o ha sido realizada de forma muy deficiente.

Entender por qué se dice en el artículo de Evett que la estadística contemplada bajo la noción de precisión en el informe PCAST es científicamente deficiente es, sin duda, un paso adicional absolutamente necesario que hay que dar, pero nos parece que, a efectos didácticos —también porque la estadística que se contempla en el informe PCAST sigue siendo muy utilizada hoy día por peritos oficiales ante los

tribunales— la lectura del informe PCAST puede ser mucho más útil que perjudicial. Una buena parte de su contenido continúa siendo válido y aprovechable para quien consiga entender finalmente la necesidad del uso del paradigma de la relación de verosimilitudes, sobre todo sus estudios empíricos.

Basta haber trabajado unos años en un laboratorio forense policial para conocer que muy pocos laboratorios de ese tipo informan a los tribunales sobre tasas empíricas de posibles errores en las conclusiones de comparaciones de características gracias a experimentos controlados y todavía menos que la mayoría de sus peritos sepan interpretar correctamente los resultados. La principal razón de la falta de esos estudios y de la escasa preparación de los peritos para evaluar certeramente la evidencia se debe a una extendida creencia —no sustentada en estudios empíricos debidamente diseñados, pero muy reforzada por la carencia ordinaria de cuestionamiento de sus resultados por ninguna de las partes en los procesos judiciales— en la eficacia y solvencia de sus métodos.

3.1. PROFUNDIZANDO EN EL PARADIGMA DE LA COINCIDENCIA

Los cotejos de huellas dactilares son pedagógicamente útiles para explicar este paradigma. En un cotejo de huellas dactilares relacionado con la investigación de un delito, la huella dubitada revelada tras la inspección ocular se compara con la huella indubitada del sospechoso o se realiza una búsqueda automática en el sistema automático de identificación dactilar del cuerpo policial investigador. Nos quedamos en el primer caso.

En el artículo de Evett (p. 18) se dice lo siguiente:

1. “Resulta conveniente llamar a estos vestigios como muestras de *referencia* y muestras *cuestionadas*, respectivamente. La materia de interés para el tribunal es la determinación del origen de la muestra cuestionada. Esta cuestión se resuelve científicamente realizando observaciones sobre ambas muestras”.
2. Más adelante, dice: “El paradigma de la coincidencia pide un juicio, por parte del científico, sobre si dos conjuntos de observaciones son coincidentes dentro de un rango esperado si la muestra cuestionada tuviera el mismo origen que la muestra de referencia (...) Si dos conjuntos de observaciones se consideran fuera del rango esperado si las dos muestras hubieran procedido de la misma fuente, entonces el dictamen del resultado sería el de una ‘no coincidencia’. Teniendo en cuenta la dependencia de la naturaleza de las observaciones, esto proporciona el fundamento para una fuerte implicación de que las muestras, la cuestionada y la de referencia, procedieron de fuentes diferentes. En muchas ocasiones, esta conclusión no será controvertida en el sentido de que tanto la acusación como la defensa aceptarán su contenido”.
3. Y finalmente: “Sin embargo, cuando el resultado de la comparación sea una ‘coincidencia’, no se sigue lógicamente que las muestras compartan el mismo origen o incluso que probablemente procedan del mismo origen. Es posible que las dos muestras procedan de fuentes distintas y que por casualidad tengan propiedades similares. A lo largo de la historia de la ciencia forense, ha existido la noción, a menudo imperfectamente expresada, de que cuanto más pequeña

sea la probabilidad de tal coincidencia, más grande es el valor probatorio asociado a la coincidencia observada”.

En este punto se hace una específica mención a la diferencia entre un análisis de la coincidencia entre vestigios (dubitado e indubitado) en una sola etapa o en dos etapas. Como antecedente histórico se cita un artículo de Evett en 1977 en el que distingue una primera etapa comparativa de otra, secuencialmente posterior, que llama etapa de significancia. La primera centrada en medir la semejanza entre las características comparadas y la segunda en medir la diferenciación entre esas características (es decir, la desemejanza entre las características comparadas y las de la población de referencia).

Evett eligió unos criterios previos al cotejo para determinar cuándo existía una coincidencia. En ese caso, eligió el criterio de que las mediciones sobre las muestras comparadas estuvieran dentro de tres desviaciones estándar en la distribución de probabilidad normal que las representaba. Y, después, valoró la probabilidad de encontrar el grado de similitud observado en la población de referencia.

Aitken et al. (2021, p. 26) explican que el establecimiento de un estadístico que se interpreta como un umbral de decisión tras la etapa comparativa (el criterio de las tres desviaciones estándar mencionado), para pasar o no a la etapa de significancia, produce un fenómeno que se ha denominado “efecto acantilado”, muy conocido en la teoría moderna de la decisión, y que es calificado por estos expertos como arbitrario.

La segunda crítica de Aitken et al. (2021, pp. 27-28) se refiere a que, con la elección del umbral, ciertas medidas procedentes de la misma fuente y otras procedentes de distintas fuentes pueden quedar descartadas. Por tanto, la segunda etapa no calcula la probabilidad de toda la evidencia, sino de una parte de la evidencia, para la que el estadístico utilizado para la toma de la decisión (pasar o no a la etapa de significancia) no supera el valor umbral, asumiendo que los dos conjuntos medidos proceden de distintas fuentes.

Aitken et al. (íbid.) argumentan que es necesario comparar esta probabilidad con la de obtener el mismo resultado si los dos conjuntos medidos proceden de la misma fuente. Así se consigue medir la probabilidad de la evidencia en su totalidad, tanto si los dos conjuntos medidos proceden de distintas fuentes como si proceden de la misma fuente.

Finalmente, resaltan (íbid.) que asumir que una probabilidad de la evidencia pequeña cuando los dos conjuntos medidos provienen de distintas fuentes implica que haya una alta probabilidad de que los dos conjuntos procedan de la misma fuente es un error de lógica de raciocinio, es decir, no respeta las reglas de la lógica matemática como los axiomas y teoremas de la teoría de la probabilidad.

Además del trabajo de Evett, otros autores han propuesto análisis en dos etapas en los que, además de distinguirlas secuencialmente, es decir, ordenarlas en serie, en cuanto a su ejecución, se miden sus resultados utilizando diferentes escalas y sin que esas escalas estén ordenadas entre sí: Morrison (2012), por ejemplo, hace esta crítica respecto a una propuesta de valoración de la evidencia acústica realizada por fonetistas forenses en el Reino Unido en 2008 con respecto a la prueba de voz.

Frente a esta propuesta de doble etapa se alza la relación de verosimilitudes que, en solo una única etapa, proporciona formalmente una valoración del peso de la evidencia como resultado de un cotejo.

No basta, no obstante, formalizar la relación de verosimilitudes correctamente desde la lógica matemática para una aplicación determinada. Hace falta asegurarse de la calibración de las relaciones de verosimilitudes mediante experimentos controlados para que esas valoraciones se correspondan realmente con lo que teóricamente representan. Este aspecto lo subraya Evett en los dos últimos apartados de su artículo y será objeto de un artículo adicional.

3.2. PROFUNDIZANDO EN EL PARADIGMA DE LA IDENTIFICACIÓN

El artículo de Evett describe perfectamente en qué consiste este paradigma (p. 18):

Históricamente, la comparación mediante las huellas dactilares fue vista como el estándar de oro a través del cual la potencia de cualquier otra técnica forense podría ser juzgada. El paradigma aquí fue la noción de “identificación” o “individualización” (los términos se utilizan aquí de forma sinónima). Con tal de que se observase suficiente nivel de detalle en la correspondencia, el resultado de una comparación entre una marca de huella dactilar de origen cuestionado y una impresión tomada de una persona conocida sería dictaminado como una opinión categórica: las dos fueron realizadas, sin género de duda, por la misma persona.

Y resalta que los paradigmas de la coincidencia y de la identificación llegan así a ser el mismo.

Los avances en inferencia lógica a los que hemos hecho referencia en la segunda mitad del pasado siglo lograron explicar en qué consiste evaluar la evidencia y demarcar esa explicación de los mencionados paradigmas.

El paradigma de la identificación sigue utilizándose en nuestros días disfrazado de moderación. En lugar de utilizar expresiones categóricas de identificación como, por ejemplo, “el cotejo positivo hallado entre la muestra dubitada e indubitada implica que ambas tengan un mismo origen”, se utilizan expresiones probabilísticas como “el cotejo positivo hallado entre la muestra dubitada e indubitada implica que es altamente probable que tengan un mismo origen”. La incertidumbre se admite porque puede que las muestras comparadas estén incompletas o contaminadas con ruido que dificulte su descripción, o que los expertos puedan fallar por inexperiencia o falta de atención, o incluso que la propia praxis forense, bien ejecutada, admita una remota probabilidad de error más teórica que real⁸.

8 De acuerdo con el modo de pensar de muchos de los expertos en dactiloscopia en los laboratorios forenses, la elección de un límite mínimo de puntos característicos coincidentes en los cotejos de huellas dactilares para que se dictamine una identificación no implica que no se reconozca, al mismo tiempo, que cuantos más puntos característicos coincidan en un cotejo, más segura es la identificación. Lo que sucede es que ese trabajo se considera inútil a los efectos de lo que se le pide al perito que dictamine. Por otro lado, se acepta la existencia de una remota posibilidad de que la identificación declarada pudiera ser errónea por la existencia de otra u otras huellas indubitadas con los mismos puntos característicos que los hallados en el cotejo. Sin embargo, se argumenta que la vasta praxis pericial realizada en miles de laboratorios forenses en el mundo durante más de cien años de existencia de la dactiloscopia y sus efectos posteriores, corrobora la seguridad de los resultados emitidos en los informes periciales.

Desde que los expertos en estadística forense de nuestra época, ligados a instituciones preponderadamente académicas, elevan su voz reclamando un cambio de paradigma en la evaluación forense de la evidencia, el paradigma de la identificación es un vestigio histórico de la creencia, infundada en la ciencia de la lógica, en que el perito es quien debe pronunciarse sobre la identificación de un vestigio que sea una pieza de convicción.

La permanencia de la mentalidad identificativa clásica en la jurisdicción penal en muchos países no tiene otra explicación que el permanente uso del paradigma de la identificación por parte de los peritos oficiales. Por eso los laboratorios oficiales son los principales responsables de esa situación y mientras quienes los dirigen no quieran aceptar los argumentos de la comunidad científica forense internacional, seguiremos viendo sentencias fundamentadas en una convicción que hace décadas los científicos calificaron como falacia.

4. ACERCA DE LOS CONCEPTOS ERRÓNEOS, FALACIAS Y CONFUSIONES EN EL INFORME PCAST

El apartado 5 del artículo de Evett es el que describe mejor las debilidades del informe PCAST. Lo que pretendemos hacer, a continuación, es complementar lo que en ese apartado se dice, intentando adaptar sus contenidos a la realidad española.

4.1. SOBRELACONFUSIÓNENTRELOSPARADIGMASDELACOINCIDENCIA Y DE LA IDENTIFICACIÓN

Puede leerse en el mencionado apartado lo siguiente (p. 18):

Hemos visto que declarar una coincidencia y declarar una identificación no son la misma cosa. Declarar una coincidencia no implica nada respecto a su peso como prueba, mientras que declarar una identificación implica un peso como prueba que asciende a la completa certeza.

La noción de “propuesta de identificación” del informe PCAST se corresponde con el modo de pensar que hemos descrito en el pie de página nº 8.

Evett et al. afirman (p. 19):

El paradigma de PCAST requiere que el científico deba emitir una afirmación categórica (una identificación) que no puede justificarse con fundamento en la lógica como ya hemos explicado. La mayoría de los científicos se sentirían cómodos con la noción de observar que las dos muestras coinciden, pero, con razón, se negarían a dar el paso lógicamente insostenible de inferir que esta observación equivale a una identificación.

Nos parece muy clarificador subrayar, como se hace incluso en el informe PCAST, que una coincidencia entre características no tiene valor probatorio en sí mismo. Lo que dota de ese valor a la coincidencia es su rareza en la población. Por eso lo que se ha explicado en el apartado 3.1 sobre la interpretación falsa que supone implicar una baja probabilidad de coincidencia (rareza en la población) con una alta probabilidad de que las muestras comparadas procedan de una misma fuente constituye el mayor peligro de confusión.

Ese tipo de razonamientos se formalizan como probabilidades condicionales. Esas probabilidades son difíciles de entender correctamente pese a su aparente sencillez. Familiarizarse con los teoremas de la teoría de la probabilidad tiene la enorme ventaja de poder entender la crítica que los coautores del artículo de Evett realizan con respecto a alguna de las afirmaciones del informe PCAST.

La confusión entre los paradigmas de la coincidencia y de la identificación es efecto, en última instancia, del desconocimiento de la lógica inherente a la teoría de la probabilidad.

4.2. SOBRELARELEVANCIADELJUICIO, ADEMÁSDELOSDATOSEMPÍRICOS

Toda probabilidad sobre un suceso real es una probabilidad condicional. Las probabilidades pueden calcularse en condiciones idealizadas y, en ese caso, aquello que nos imaginemos puede llegar a estar perfectamente definido por nuestro pensamiento. Sin embargo, sobre la realidad no tenemos ese control.

Por eso cuando pretendemos conocer una parte de la realidad y utilizamos nuestra capacidad de abstracción, forzosamente tenemos necesidad de *idealizarla*.

Cuando pensamos en las probabilidades de que salga un as, cuando una persona cualquiera lanza un dado, no es costoso reconocer que necesitamos reducir la complejidad de la realidad para poder dar una respuesta útil y satisfactoria. Por ejemplo, suponer que el dado sea equilibrado, que los lanzamientos sean análogos, que el estado anímico y físico del lanzador sea estable, etc.

Resulta conveniente fijarse en la notación matemática utilizada frecuentemente por los estadísticos forenses cuando escriben sobre cálculo de probabilidades en contextos propios de su ciencia —por ejemplo, la probabilidad de una coincidencia entre las características que se analizan de dos muestras que se comparan entre sí—.

Llamemos E a la evidencia consistente en la coincidencia entre las características comparadas de dos muestras distintas entre sí, una procedente de la escena del crimen y otra procedente de un sospechoso. La probabilidad condicional de que observemos esas coincidencias si el sospechoso fue el origen de ambas, puede expresarse así: $P(E | I)$, donde P significa probabilidad, E evidencia o cotejo positivo entre las características de las muestras comparadas, I información de contexto (entre la que se encuentra saber que el origen de las muestras es el sospechoso) y la barra vertical la notación que indica que E está condicionada a I , es decir, la evidencia está condicionada a la información de contexto.

Esa notación puede completarse (y complicarse) algo más si, en lugar de considerar incluido en la I (información de contexto) el hecho de saber que el sospechoso es el origen común de ambas muestras, introducimos un nuevo suceso que llamamos H_p como la hipótesis que mantiene la acusación de que el sospechoso sea el origen común de las muestras y H_d la hipótesis contraria.

Así: $P(E | H_p, I)$ es la notación que significa la probabilidad de que se observe la coincidencia entre las características de las dos muestras comparadas condicionada a que la hipótesis que mantiene la acusación sea cierta y a la información de contexto (la unión mediante la letra y griega en el texto y una coma en la fórmula entre H_p e I

suele llamarse “intersección” y significa que ambos sucesos suceden conjuntamente). En este caso, en que hemos desligado a la hipótesis condicionante H_p de la información de contexto, subrayamos que cuando decimos que toda probabilidad real es condicional, podemos agrupar lo que condiciona o desagruparlo. En nuestro ejemplo, la información de contexto desagrupada podría ser la circunstancia de que la muestra dubitada sufriera degradación por las condiciones de humedad y temperatura en el lugar de los hechos hasta que pudiera ser recogida para su análisis.

Lo que dice Evett en su artículo es que la información de contexto condiciona el resultado del cotejo. Traduciéndolo a nuestro ejemplo, las condiciones reales en las que hagamos la prueba del lanzamiento del dado condicionan el resultado o las condiciones en las que las muestras se encuentren al ser recogidas para su análisis condicionan el resultado del cotejo. Por tanto, no es suficiente un control experimental idealizado para explicar un suceso real concreto en el que estemos interesados en calcular su probabilidad.

4.3. SOBRE LO SUBJETIVO VERSUS OBJETIVO

La estadística frecuentista clásica afirmaba que el fin principal de las técnicas estadísticas aplicadas a la ciencia era el de objetivar los resultados. Así, la noción de lo objetivo como opuesto a lo subjetivo se entiende como lo que evita eficazmente —en la medición del valor probatorio de una prueba— toda posible influencia procedente del juicio de un evaluador.

Sin embargo, esta concepción de la estadística es consecuencia de una concepción de la ciencia y del conocimiento filosóficamente reduccionista⁹. Los datos y la interpretación de los datos a partir de teorías requieren una fundamentación racional lógico-técnica y lógico-filosófica que los trasciende. La lógica está intrínsecamente relacionada con el ejercicio de la razón en su más amplio sentido: en la abstracción, en el juicio y en el raciocinio. Todo juicio humano, incluido el que un empirista llamaría objetivo, es esencialmente subjetivo.

El problema epistemológico que debería preocupar a un científico no es tanto si su juicio es objetivo o subjetivo, sino si su juicio es verdadero o no. Y si está convencido de que su juicio es verdadero, tendría que dar razón de su convicción. La noción de verdad como adecuación del juicio subjetivo a la realidad —esa es la noción de la psicología cognitiva realista— no puede dejar de ser esencialmente subjetiva por definición, porque hay una parte de la noción ligada esencialmente a la subjetividad. Los clásicos han distinguido entre verdad ontológica y verdad lógica. La verdad de las proposiciones es una verdad lógica y es a la que normalmente nos referimos en los juicios que se vierten en los informes periciales.

Para ilustrar cómo lo subjetivo impregna el quehacer científico más elemental, puede servir como ejemplo lo que hacen los policías cuando desempeñan su rol de policía judicial investigando la posible comisión de un delito en la escena del crimen. La denominada inspección ocular descansa, muy principalmente, sobre sus capacidades sensitivas y el juicio humano sobre lo que esas capacidades le suministran.

9 Un buen manual para entender la objetividad y la verdad en las ciencias, a la par que la influencia en ellas de la subjetividad y del convencionalismo, es el del profesor Mariano Artigas (1992).

La utilización de medios técnicos en esa inspección es también fundamental y su relevancia está fuera de toda duda, pero nadie puede negar que el proceso de búsqueda de indicios relevantes en una investigación real está supeditado a un escrutinio sobre la información disponible al investigador que persigue averiguar cómo pudieron suceder los hechos que investiga, cuya naturaleza trasciende los datos —su mera enumeración y descripción— y forma parte de un proceso de inferencia lógica que intenta explicar lo observado.

En definitiva, los datos se interpretan a la luz de hipótesis que los explican plausiblemente y ese proceso requiere el permanente ejercicio de juicios subjetivos. Otra cosa es cómo se sustancia el proceso judicial mediante el ejercicio de las facultades jurídicamente reconocidas a las partes.

4.4. UNAFALACIA OMNIPRESENTE: LA TRANSPOSICIÓN DEL CONDICIONAL

Esta falacia de lógica se ha hecho especialmente famosa por el número de ocasiones en que se comete, por toda clase de actores procesales: fiscales, abogados, jueces, periodistas —estos son los colectivos literalmente citados en el artículo de Evett— a los que añadimos los peritos e investigadores policiales en todas partes.

La enumeración de ejemplos de falacias en el informe PCAST que se recoge en el artículo de Evett es suficientemente expresiva como para que nos haga reflexionar sobre cómo es posible que un colectivo selecto de científicos de nuestros días —el informe PCAST nombra a cada uno de los expertos intervinientes en su redacción y resalta sus currículos— pueda respaldar un informe con ese número de casos flagrantes de errores de lógica de raciocinio.

Este error es una auténtica pesadilla para muchos científicos forenses, los cuales tienen una especial sensibilidad para detectarlo. El artículo de Evett es un magnífico ejemplo porque cada uno de esos errores supone un atentado contra la racionalidad cuando lo sostienen los actores procesales. En este aspecto, es especialmente lamentable la profusa comisión de esta falacia por peritos en sus testimonios o en sus informes, máxime cuando se supone que disponen de la mejor formación e información sobre este problema entre los colectivos mencionados en el artículo de Evett.

Singularmente, en el caso de España, esta falacia se comete diariamente ante los tribunales de la jurisdicción penal por numerosos peritos oficiales que siguen defendiendo el paradigma de la identificación en sus informes y testimonios. Solamente la falta de formación de los abogados en inferencia lógica puede explicar la falta de respuesta de este colectivo ante la comisión de esta falacia en España por parte de jueces, fiscales, policías y peritos oficiales.

4.5. UN EJEMPLO DE CONFUSIÓN: “COINCIDENCIA PROBABLE”

Verdaderamente es notable hasta qué punto la terminología que utilizan algunos expertos forenses puede tergiversarse hasta perder su sentido más elemental.

Este ejemplo no requiere que hagamos mucho esfuerzo para explicar que se trata de un sinsentido. Cuando se comparan dos huellas dactilares en un cotejo, lo que coincide o no son el tipo de núcleo, el tipo de delta, el tipo y situación de los puntos

característicos, la forma de las crestas, etc. Las coincidencias pueden describirse, al igual que las no coincidencias. Cuando se habla de “coincidencia probable”, el único sentido posible de esa expresión es que manifieste que hay duda sobre esos aspectos descriptivos por diversos motivos, como puede ser la falta de calidad en la impresión disponible de las huellas dactilares que se comparan.

Una vez que coinciden todos esos aspectos descriptivos de las huellas dactilares que se comparan, hay que realizar una inferencia lógica para pronunciarse sobre la posible identidad de la huella dactilar, pero ese proceso inferencial no lo puede resolver el perito sino el tribunal. Para comprender esto solo hay que tener en cuenta lo que se ha dicho sobre la relevancia de la lógica: en esto es clave el teorema de Bayes.

4.6. SOBRE LA VALIDEZ DE LOS FUNDAMENTOS Y PRECISIÓN

En este apartado se definen las nociones de repetibilidad, reproducibilidad y precisión. Las dos primeras no ofrecen dudas sobre su significado, pero sí la tercera. De entrada, la repetibilidad y la reproducibilidad se consideran expresiones de precisión, que consiste en “la cercanía entre las medidas o los resultados” (p. 21).

Cuando un tirador dispara contra una diana y consigue que los disparos aparezcan agrupados, con independencia de si esa agrupación esté cerca o no del centro de la diana, podemos decir que sus disparos han sido precisos. Cuando, además de estar agrupados, los disparos se sitúan en torno al centro de la diana, hablamos de que han sido exactos. En este ejemplo, lo que se supone que el tirador tenía que hacer es intentar que sus disparos estuvieran lo más cerca posible del centro de la diana.

Dice el artículo de Evett al respecto (p. 21):

La exactitud es una medida de como una medida o un conjunto de medidas están cercanas a las respuestas verdaderas. Esto tiene un significado obvio cuando conocemos o podemos conocer las respuestas verdaderas.

En el ejemplo, estaba claro dónde estaba el centro de la diana.

Y sigue Evett (p. 21):

Sin embargo, si queremos responder a una cuestión como “¿cuál es la probabilidad de que se hubiera producido una coincidencia con el calzado del sospechoso si no hubiera dejado la marca en la escena del crimen?, entonces no tiene sentido que haya una “respuesta verdadera”. Los valores que el experto asigne a tales probabilidades variarán dependiendo del específico conocimiento de los expertos y de la naturaleza de cualquiera de las bases de datos que el experto pueda usar para informar sus probabilidades.

Con este ejemplo, Evett intenta conseguir que abandonemos la idea de que existan respuestas verdaderas cuando se calculan probabilidades en un caso forense real.

Efectivamente, las probabilidades son entes de razón con fundamento en la realidad, no son entes reales que puedan existir independientemente de nuestro pensamiento. Por eso no tiene sentido que tratemos de entender las probabilidades como respuestas verdaderas a un problema que intentamos resolver. Lo que sí tiene sentido es que nos preguntemos por la coherencia de nuestros valores probabilísticos y cálculos con las leyes y teoremas de la probabilidad: es un problema lógico, no un problema ontológico.

Acto seguido, el artículo de Evett alude a “la difícil cuestión sobre el significado de la precisión”. Califica de “simplista” la alusión a la precisión que se hace en el informe PCAST en la página 46:

Sin las apropiadas estimaciones de acierto en las comparaciones, una proposición del experto que diga que dos muestras comparadas son similares -o, incluso, indistinguibles- carece de significado científico: carece de valor probatorio y tiene potencial considerable de causar un impacto perjudicial. Nada -ni personal formado y con experiencia personal, ni prácticas profesionales- pueden sustituir a una demostración empírica adecuada sobre el nivel de precisión.

Y Evett remarca: “Hemos visto que el informe se equivoca aquí —no es cuestión de precisión sino de peso de la prueba—” (p. 21).

Si las probabilidades no son realidades singulares autónomas, con existencia propia independiente, si se trata de entes de razón con fundamento en la realidad que dependen absolutamente del pensamiento para existir, la noción de precisión a la que se alude en el informe PCAST, noción vinculada a la metrología, no tiene sentido aplicarla en una gran mayoría de las cuestiones que se plantean en los informes periciales forenses que son de naturaleza probabilística, no de naturaleza métrica.

Y respecto al peso de la prueba, la lógica a la que se ha aludido desde el principio en el informe de Evett es la que explica su posición en este aspecto: el perito tiene la función de evaluar el peso de la evidencia que consiste en conocer en qué medida es más probable observar lo hallado tras el proceso analítico correspondiente sobre las muestras si la hipótesis de la acusación se considera cierta frente a lo mismo si se considera cierta la hipótesis alternativa de la defensa.

La relación de verosimilitudes mide la fuerza de la evidencia y las probabilidades (o las densidades de probabilidad) miden incertidumbres. Los peritos han de proporcionar a los tribunales lo que es propio de su función: cuantificar, a ser posible, la fuerza de la evidencia mediante una relación de verosimilitudes y los miembros del tribunal han de intentar cuantificar, en lo posible, las probabilidades de las tesis de las partes en el proceso, una vez conocidos los resultados periciales y el resto de información relevante para dirimir la causa.

4.7. SOBRE EL PARADIGMA DEL PCAST

Dice Evett que el informe PCAST propone una fusión de los paradigmas de coincidencia y de identificación, y pone el foco en este párrafo (p. 21):

Como consecuencia de que el término “match” o coincidencia es probable que implique un valor probatorio inapropiadamente alto, debiera utilizarse un término más neutral para expresar la creencia del experto en que las dos muestras proceden de la misma fuente. Sugerimos el término “identificación propuesta” para transmitir adecuadamente la conclusión del experto, junto con la posibilidad de que quizá sea errónea. Usaremos este término a lo largo de este informe.

Argumenta Evett: (1) el término “match” (coincidencia), utilizado propiamente, no tiene implicación alguna de valor probatorio; (2) no es misión del científico asignar una probabilidad a la proposición de si las dos muestras procedieron de la misma fuente; (3) se pregunta por el significado de la expresión “identificación propuesta” y llega a la conclusión de que se trata de una afirmación categórica —si fuera probabilística, no

podría ser “errónea”—¹⁰. Y si no es probabilística, el científico proporciona una opinión categórica diciéndole al tribunal ¡que quizá se equivoque!: lo cual es un sinsentido.

Una vez más, la lógica que permite expresar propiamente el valor de prueba del resultado de un cotejo exige que se tenga en cuenta tanto la probabilidad de lo observado, si la hipótesis de la acusación se considera cierta, como lo mismo si se considera cierta la hipótesis de la defensa. Pronunciarse sobre las probabilidades de las proposiciones hipotéticas, una vez conocidos los resultados de las pruebas practicadas, es una misión que le corresponde al tribunal. A los peritos les pertenece pronunciarse sobre las probabilidades de lo que observan con sus métodos analíticos considerando ciertas cada una de las hipótesis planteadas por las partes.

Es un error tremendo entender que el perito tenga como misión específica de su función, fundamentada en sus conocimientos y experiencia, la de calcular probabilidades de las hipótesis planteadas por las partes, una vez conocidos los resultados de los peritajes.

Por eso toda afirmación de un perito, relacionada con asumir el paradigma de la identificación, es una falacia de lógica de raciocinio que ya hemos dicho que se conoce como falacia de transposición del condicional y ya va siendo hora de que esa forma de pensar desaparezca de los informes periciales emitidos por los laboratorios oficiales españoles.

4.8. SOBRE EL CIENTÍFICO COMO UNA “CAJA NEGRA”

La subjetividad en el juicio y en el raciocinio es inherente al modo en que el ser humano conoce la realidad. En el proceso de conocimiento de la realidad el ser humano emplea sus facultades de conocimiento: las sensitivas y la intelectual.

El modo en que la inteligencia trata de conocer la realidad forma parte esencial de la psicología cognitiva. Es un error ver en la subjetividad una fuente de problemas en el proceso de conocimiento que hay que tratar de evitar. Sencillamente, sin la subjetividad, el ser humano no podría conocer la realidad.

Cualquier procedimiento científico de conocimiento tiene una carga subjetiva necesaria. El proceso de objetivación que el científico realiza para poder llegar a resultados fiables —aplicando el método científico— tiene siempre una necesaria vinculación con la subjetividad. Por ejemplo, la inspección ocular que realiza un perito en el lugar de los hechos nunca es neutral bajo cualquier punto de vista. No es posible analizar todo lo que pueda materialmente recogerse en el lugar de los hechos como si solo así pudiera garantizarse una idealizada neutralidad en la inspección. De exigirse semejante cosa, posiblemente no podría resolverse caso alguno, dado el elevado número de vestigios que podrían analizarse y el nivel de sofisticación de las técnicas hoy día disponibles. El riesgo es inherente al pensamiento científico porque su metodología no puede evitarlo, aunque sí minimizarlo.

La investigación de un crimen está inserta en un proceso judicial dirigido por un juez o un fiscal y está sometido a reglas a las que las partes se someten para garantizar

10 Aunque, desde luego, sería lógicamente incorrecta a causa de que implicaría una transposición del condicional.

un juicio justo. Al tratarse de un proceso reglado, el procedimiento científico que se utilice como auxilio para el esclarecimiento de los hechos y la identificación de los delincuentes está dirigido por las partes y controlado por el juez o fiscal.

Todo eso hace que el científico no pueda imaginarse como una “caja negra” en la que resulta cognoscible lo que entra y sale de ella, pero no lo que hay en su interior. Es más, el proceso judicial está ideado con la finalidad de que las pruebas se diriman en el juicio entre las partes. Por tanto, es inherente al proceso que las pruebas se practiquen bajo los principios de inmediatez, oralidad, contradicción, libre valoración de la prueba según las reglas del criterio racional, presunción de inocencia y el de *in dubio pro reo*, si bien la única opacidad admisible sería la personal falta de capacidad de comprensión de aspectos técnicos de cualquier naturaleza para los que fuera necesaria una formación especializada y de la que se carece.

Precisamente, el científico concurre como perito por poseer conocimientos especiales necesarios para dirimir aspectos técnicos determinados del proceso y, por consiguiente, su misión es arrojar luz —no tinieblas (por aquello de la “caja negra”)— a los actores procesales para que puedan defender sus tesis apropiadamente.

No obstante, los estudios de “caja negra” son útiles y necesarios como parte de la experimentación científica imprescindible para evaluar la idoneidad del uso de un determinado método en la administración de justicia. Precisamente su falta generalizada ha ocasionado lo que podríamos denominar “alarma social”, toda vez que el desconocimiento de las capacidades de los sistemas empleados por los peritos para resolver cotejos criminalísticos, concretado en las tasas de error mencionadas en el informe PCAST, deja a los tribunales a merced del crédito que le merezcan quienes les presentan los informes periciales y no bajo el soporte que cabe esperar del estado del arte de las ciencias y las técnicas en el siglo XXI.

4.9. ESTUDIOS DE “CAJA NEGRA”

Dice Evett: “PCAST propone que las tasas de error de tales experimentos serían utilizadas para asignar valor probatorio en los tribunales” (p. 21).

Proponer que los estudios de “caja negra” sirvan para que los peritos asuman sus resultados para las asignaciones de valor probatorio en cada pericia es un error de lógica (por consiguiente, también metodológico). Esos estudios proporcionan información muy útil y necesaria para valorar el estado del arte de la ciencia y técnica en la que estemos interesados, pero no sirven para evaluar la evidencia en un caso determinado.

Como se ha señalado anteriormente, el informe PCAST aporta información de gran relevancia para hacerse una idea, lo más objetiva posible, sobre hasta qué punto los métodos de comparación de características utilizados en el ámbito forense son fiables. Por tanto, sus aportaciones son enriquecedoras y no pueden pasarse por alto, aunque, al mismo tiempo, hay que resaltar que el enfoque metrológico que inspiran sus páginas no es el adecuado para resolver las necesidades de los peritos para poder evaluar la evidencia de sus cotejos de acuerdo con las exigencias de los avances de la lógica en nuestra época.

4.10. SOBRE LA GOBERNANZA

El problema científico de la evaluación de los datos como prueba está muy relacionado con la historia de la filosofía de la ciencia y de la estadística en los últimos dos siglos. La envergadura de los problemas que intenta resolver y las aportaciones más sobresalientes que se han presentado para su solución solo se vislumbran cuando se tiene una perspectiva tan amplia como la referida.

Por eso, la metrología es una parte de la ciencia que, por sí sola, es incapaz de adentrarse en la profundidad de la parte de la realidad que intenta conocer.

Sobre este tema es relevante conocer cómo cada instituto forense tiene resuelto su régimen de personal, en concreto a lo referente a quién gobierna los laboratorios y quién puede trabajar en ellos como experto. En países del primer mundo hay instituciones oficiales que permiten que los laboratorios forenses estén dirigidos por investigadores policiales. Eso ocurre, sin ir más lejos, en España. El problema no es que sean policías, sino que no tengan la suficiente formación científica y carezcan de experiencia como peritos.

5. LAS NOCIONES DE INFERENCIALÓGICA Y CONOCIMIENTO CALIBRADO

El apartado 6 del artículo de Evett se titula “Nuestra visión del futuro” y dice esto en el subapartado 6.1 (p. 22):

Las recomendaciones del informe PCAST están fundamentadas en una conjunción de dos paradigmas forenses clásicos: coincidencia e identificación. Estos paradigmas son tan antiguos como la ciencia forense, pero sus inadecuaciones y carencias de lógica han sido comprensivamente expuestas en los últimos 50 años más o menos. Todos nosotros mantenemos, y hemos hecho eso en nuestros artículos, que el futuro de la ciencia forense debe estar fundamentado, primero, en la noción de inferencia lógica, y, segundo, en la noción de conocimiento calibrado. El primero conduce a un marco de principios (que han sido adoptados por ENFSI¹¹) y estamos perplejos de que PCAST haya aparentemente decidido ignorarlos, o, a lo sumo, mantenerse en silencio, a este cambio fundamental. El segundo es un concepto más profundo y mucho más rico que la noción profundamente limitada de las tasas de error de falsos-positivos y falsos negativos: se trata de la noción de calibración.

Estando plenamente de acuerdo con el contenido de este subapartado, sin embargo, no se hace mención alguna a los esfuerzos que se han hechos en el seno de algunos laboratorios, públicos y privados, sobre todo en la red ENFSI—aunque también merecen ser mencionados algunos laboratorios no pertenecientes a esa red—, para que la noción de inferencia lógica se implante en la práctica forense¹².

11 ENFSI GUIDELINE FOR EVALUATIVE REPORTING IN FORENSIC SCIENCE (2015). <https://enfsi.eu/documents/forensic-guidelines>

12 En España se ha logrado plena implantación de esa noción en el ámbito de las pruebas de ADN—como ha sucedido en todos los países en los que se practican esas pruebas— y es bien conocido que en el Laboratorio Central del Servicio de Criminalística de la Guardia Civil se utiliza esa inferencia en los informes de cotejo de voces desde el año 2004. En ese laboratorio, en estrecha cooperación con el grupo de investigación universitario denominado ATVS (Área de Tratamiento de Voz y Señales)—y que hoy día se denomina AUDIAS (<http://audias.ii.uam.es/>)—, perteneciente a la Universidad Politécnica de Madrid (1997-2005) y a la Universidad Autónoma de Madrid (2005-2013) secuencialmente— en los años en los que un Convenio de Colaboración entre la Secretaría de Estado para la Seguridad del Ministerio del Interior estuvo en vigor activamente (1997-2013), se logró ese objetivo.

Este hecho resalta que la implantación de la inferencia lógica de la que habla Evett no es futuro porque ya hay pasado y presente. El que un área de la criminalística en un laboratorio de nuestra época avance más o menos hacia ese objetivo no es una cuestión preferentemente de recursos económicos o capacidad de desarrollo industrial. Es una cuestión de convencimiento personal y, por influencia de muchos, de convencimiento institucional. Ese convencimiento se ha extendido ampliamente, desde hace décadas, en el ámbito académico próximo a los laboratorios forenses, pero aún es muy débil en el ámbito forense profesional.

Fuera de España, podemos destacar las contribuciones en esa línea de laboratorios forenses de ENFSI como el laboratorio forense nacional sueco (antiguo SKL), el de la Garda irlandesa, el instituto de investigación forense (IFR) de Cracovia (Polonia) y el instituto forense holandés (NFI). En el ámbito privado, es relevante la influencia de algunos laboratorios en el Reino Unido (no pocos de sus principales investigadores actuales pertenecieron al extinto y prestigioso Servicio Forense Británico).

5.1. LA CRÍTICA DE EVETT ET AL. A LA VISIÓN DE FUTURO DE PCAST

Los autores del artículo de Evett hacen estas significativas afirmaciones (p. 22):

- (1) aunque están “a favor del estudio de la opinión del experto en circunstancias controladas (...), las pruebas de competencia son mucho más que el recuento de errores”;
- (2) “la interpretación forense es mucho más rica y más informativa que un simple sí/sin respuestas”;
- (3) “la noción de una tasa de error que se presenta ante los tribunales es engañosa porque falla en reconocer que la ciencia avanza como resultado de las pruebas de competencia”;
- (4) “ha habido grandes avances en los últimos años sobre la forma en que las distribuciones de relaciones de verosimilitudes de tales experimentos pueden compararse y evaluarse (Ramos [26]¹³, Brümmer [27]¹⁴ ver también Robertson et al. [28]¹⁵ para una introducción para no expertos sobre la calibración). La elegancia y el rendimiento de tales métodos trasciende con creces la cruda noción de PCAST de tasas de error de “falsos positivos” y “falsos negativos” (p. 22).

Esto es justamente lo que creemos que motiva que este artículo deba continuarse con un segundo, específicamente técnico, que se fundamenta en los contenidos de los cursos de especialización impartidos por los investigadores de AUDIAS al Departamento de Ingeniería de la Guardia Civil.

5.2. SOBRE EL CONOCIMIENTO Y LOS DATOS

El artículo de Evett destaca la importancia de que los peritos tengan en cuenta proposiciones a nivel de actividad (véase también Cook et al., 1998) cuando tengan que

13 Coincide con nuestra referencia (Ramos D. et al., 2013b).

14 Coincide con nuestra referencia (Brümmer N. et al., 2006).

15 Coincide con nuestra referencia (Robertson G.A. et al., 2016).

resolver un caso real. Este es uno de los aspectos más relevantes en el avance en valoración de la evidencia experimentado en las últimas décadas.

Los autores resaltan (p. 22) que:

1. “La duración de la experiencia no es una medida de la fiabilidad de la opinión científica: la base es el conocimiento fiable”. En eso están de acuerdo con PCAST.
2. “La visión de que los datos pueden reemplazar al juicio es engañosa. Una colección de datos puede utilizarse para informar conocimiento fiable —no para reemplazarlo—”.
3. Se oponen a considerar al científico como “caja negra” porque reducen sus juicios a “un todo o nada”. “El juicio cualitativo estará siempre en el centro de la evaluación forense de la evidencia científica”.
4. Por tanto, rechazan la visión de PCAST del científico como aquel que emite una opinión categórica o una afirmación sobre la probabilidad de que la opinión sea errónea.

Al científico lo caracterizan como: (1) profundamente conocedor de su dominio de pericia y capaz de racionalizar la opinión en términos que el jurado entienda; (2) sus opiniones se rigen por los principios de equilibrio, lógica, solidez y transparencia. Por tanto, no tiene sentido entenderlo como una “caja negra”.

Terminan el subapartado diciendo que están de acuerdo con PCAST en que subraye que el científico proporcione al tribunal lo que le solicite con pruebas de rendimiento bajo condiciones controladas, pero la solución técnica a esta necesidad está en la calibración de las valoraciones probabilísticas.

5.3. SOBRE LA CONCLUSIÓN DEL INFORME PCAST

La 44ª solicitud del presidente de los Estados Unidos fue “sobre la posibilidad de dar pasos adicionales que pudieran ser útiles desde el punto de vista científico para fortalecer las disciplinas forenses y así asegurar la validez de la prueba científica utilizada en el sistema jurídico nacional” ([1], p.1). Pensemos que el informe aporta muy poco para dar pasos positivos y refuerza mucho un pensamiento y una terminología pobres (p. 23).

El principal problema que los autores del artículo ven en el informe PCAST es su desvinculación con un marco lógico de referencia como el de la inferencia bayesiana que permita a los peritos evaluar la evidencia científica de acuerdo a los ya mencionados principios de equilibrio, lógica, solidez y transparencia.

Puede consultarse en la bibliografía una selección de artículos de Taroni et al., sobre evaluación de la evidencia forense de acuerdo con la inferencia bayesiana.

Explícitamente dicen, con respecto a ese marco lógico de referencia necesario, lo siguiente:

1. Ese marco lo proporciona la teoría de probabilidad junto con el reconocimiento de que la probabilidad es necesariamente subjetiva y condicionada por el conocimiento y el juicio. De ello se deduce que nuestra visión del científico forense

es el de una persona con conocimiento, lógica y razonable. Mientras que las colecciones de datos son valiosas, deben verse dentro de un contexto de conocimiento fiable. El paradigma general del conocimiento fiable debe fundamentarse en la noción de gestión del conocimiento, que incluye sistemas comprensivos para la calibración de la opinión de un experto (p. 23).

2. Los juicios subjetivos de los expertos no basados en datos también pueden calibrarse. El propio Evett señala cómo hacerlo en la práctica explicándolo en uno de los artículos recomendado en la bibliografía. Sobre este tema ha habido contribuciones relevantes en publicaciones recientes de psicología cognitiva.

6. CONCLUSIONES

Los informes de las Academias de Ciencias, Ingeniería y Medicina norteamericanas son ejercicios de escrutinio del estado del arte de la ciencia forense en el país al que pertenecen que no tienen parangón.

El ejercicio de crítica científica realizado por los autores del artículo publicado en la revista *Forensic Science International* en 2017, sobre el contenido del informe PCAST 2016, muestra lo lejos que aún se encuentran las formas de entender la estructura lógica de evaluación de la evidencia como prueba científica a ambos lados del Atlántico. El debate científico primordial que divide a los científicos está centrado en la lógica que debe aplicarse. Por eso, la formación en estadística —hasta donde hemos sido capaces de desarrollarla en nuestros días— es, a nuestro juicio, la asignatura más importante que debe asimilar y superar un perito del siglo XXI.

No obstante, es evidente el cambio positivo que la comunidad científica norteamericana está experimentando en lo que va de siglo por el rigor con el que intentan practicar la ciencia forense en sus laboratorios. Es también muy de agradecer la transparencia con la que comunican sus debates, sus logros y sus debilidades, así como su compromiso con la mejora de la justicia.

BIBLIOGRAFÍA

Aitken C.G.G., Taroni F., & Bozza S. (2021). *Statistics and the Evaluation of Evidence for Forensic Scientists*. 3rd ed., Wiley, Chichester (UK).

Artigas M. (1992). *Filosofía de la ciencia experimental: la objetividad y la verdad en las ciencias*. 2^a edición ampliada, EUNSA.

Brümmer N., du Preez J. (2006). Application-independent evaluation of speaker detection. *Computer Speech and Language*, 20, 230-275.

Cook R., Evett I.W., Jackson G., Jones P.J., & Lambert J.A. (1998). A hierarchy of propositions: deciding which level to address in casework. *Science and Justice* 38(4), 231-240.

ENFSI guideline for evaluative reporting in forensic science (2015). <https://enfsi.eu/documents/forensic-guidelines>

Evett I.W. (2009). Evaluation and professionalism. *Science & Justice*, 49(3), 159-160.

- González Rodríguez J., Drygajlo A., Ramos Castro D., García Gomar M., & Ortega García J. (2006). Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. *Computer Speech and Language*, 20(2-3), 331-355.
- González Rodríguez J., Rose P., Ramos Castro D., Toledano D.T., & Ortega García J. (2007). Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 15(7), 2104-2115.
- González Rodríguez J. (2014). Evaluating Speaker Recognition systems: An overview of the NIST Speaker Recognition Evaluations (1996-2014). *Loquens*, 1(1), e007. <http://loquens.revistas.csic.es/index.php/loquens/article/download/9/21>
- Morrison G.S. (2012). Response to DRAS 5388.3 Forensic Analysis – Part 3 – Interpretation. [https://geoff-morrison.net/documents/Morrison,%20et%20al%20\(2012\)%20Response%20to%20Australian%20Draft%20Standards%20DR%20AS%205388.3%20Forensic%20analysis%20-%20Part%203-%20Interpretation.pdf](https://geoff-morrison.net/documents/Morrison,%20et%20al%20(2012)%20Response%20to%20Australian%20Draft%20Standards%20DR%20AS%205388.3%20Forensic%20analysis%20-%20Part%203-%20Interpretation.pdf)
- Ramos Castro D. (2007). Forensic evaluation of the evidence using automatic speaker recognition systems. Tesis doctoral. Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid.
- Ramos Castro D., González Rodríguez J. (2013a). Reliable Support: Measuring Calibration of Likelihood Ratios. *Forensic Science International*, 230, 156-159.
- Ramos Castro D., González Rodríguez J., Zadora G., & Aitken C. (2013b). Information-theoretical assessment of the performance of likelihood ratio computation methods. *Forensic Science International*, 58(6), 1503-1518.
- Robertson B., Vignaux G.A. (2016). *Interpreting Evidence – Evaluating Forensic Science in the Courtroom*. 2nd edition, John Wiley & Sons, Chichester (UK).
- Royall R. (1997). *Statistical Evidence: A Likelihood Paradigm*. Chapman&Hall, London (UK).
- Taroni F., Aitken C.G.G., & Garbolino P. (2001). De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. *Science and Justice*, 41(3), 145-150.
- Taroni F., Bozza S., Biedermann A., Garbolino P., & Aitken C. (2010). *Data Analysis in Forensic Science*. J. Wiley & Sons, Statistics in Practice, Chichester (UK).
- Taroni F., Biedermann A., Bozza S., Garbolino P., & Aitken C. (2014). *Bayesian Network for Probabilistic Inference and Decision Analysis in Forensic Science*. 2nd edition, John Wiley & Sons, Statistics in Practice, Chichester (UK).
- Taroni F., Garbolino P., Biedermann A., Aitken C., & Bozza S. (2018). Reconciliation of subjective probabilities and frequencies in forensic science. *Law, Probability and Risk*, 17(3), 243-262.