

FILOSOFÍA DE LA MENTE Y PRUEBA DE LOS ESTADOS MENTALES: UNA DEFENSA DE LOS CRITERIOS DE «SENTIDO COMÚN»

Daniel González Lagier

Catedrático de Filosofía del Derecho de la Universidad de Alicante
daniel.gonzalez.lagier@gmail.com

RESUMEN: La prueba de los hechos psicológicos o estados mentales plantea dificultades de carácter ontológico (¿existen realmente hechos mentales, distintos de los hechos físicos?), epistemológico (¿cómo puedo conocer los estados mentales de los demás?) e ideológico (¿son jurídicamente relevantes?). Una parte de la doctrina y de la jurisprudencia ha dicho que, en realidad, los estados mentales no son —o no deben ser tratados como— “hechos” propiamente, por lo que no se prueban (esto es, no se demuestra que hayan ocurrido o hayan existido), sino que para el Derecho son entidades ficticias que se imputan o adscriben a partir de ciertos criterios normativos. Este escepticismo contrasta con el hecho de que todos atribuimos con frecuente éxito estados mentales a los demás y conseguimos adaptar y coordinar nuestra conducta a la de los otros gracias a esas atribuciones. Sabemos atribuir estados mentales, pero no sabemos cómo lo hacemos. En este trabajo trato de analizar qué luz puede arrojar la filosofía de la mente (en particular, las teorías acerca del llamado “problema de las Otras Mentes”) a la prueba de los hechos psicológicos en el Derecho.

PALABRAS CLAVE: Prueba; estados mentales; filosofía de la mente; otras mentes; atribución psicológica.

ABSTRACT: The proof of psychological or mental facts poses challenges of an ontological nature (do mental facts really exist as distinct —and not reducible to— physical facts?) epistemological (how can we know others’ mental states?) and ideological (are those facts legally relevant?). Part of legal doctrine and case law holds that states of mind are actually not —or ought not to be approached as— proper “facts”, therefore they are not to be proved (their occurrence or existence cannot be

demonstrated). From a legal perspective, they are fictional entities to be ascribed or imputed using certain normative criteria. This scepticism is openly in contrast with the fact that we quite successfully and frequently attribute states of mind to others, as well as we all manage to adapt and coordinate our behavior following those attributions. We know how to attribute states of mind, however we ignore how we do it. In this paper, I try to explore how Philosophy of Mind (theories on the so-called “problem of other minds”) can contribute to shed some light on the proof of psychological facts in the realm of Law.

KEYWORDS: Proof; evidence; mental states; Philosophy of Mind; Other Minds; psychological attribution.

SUMARIO: 1. UNA APROXIMACIÓN A LA FILOSOFÍA DE LA MENTE: 1.1. ¿Qué son los «estados mentales»? 1.2. ¿Existen los estados mentales?. 1.3. ¿Existen otras mentes, además de la mía?. 1.4. ¿Cómo conozco mis propios estados mentales?. 1.5. ¿Cómo conozco los estados mentales de los demás?.— 2. LA PRUEBA DE LOS ESTADOS MENTALES EN EL DERECHO: 2.1. El argumento ontológico contra la prueba de los estados mentales. 2.2. El argumento epistemológico contra la prueba de los estados mentales. 2.3. El argumento ideológico contra la prueba de los estados mentales. 2.4. Entonces... ¿prueba o imputación?. 2.5. La prueba de los estados mentales con contenido representacional. 2.6. La prueba de los estados mentales de carácter fenomenológico. 2.7. La prueba de los estados mentales mixtos (las emociones). 3. CONCLUSIONES.— BIBLIOGRAFÍA.

RECOMMENDED CITATION: DANIEL GONZÁLEZ LAGIER, 2022: «Filosofía de la mente y prueba de los estados mentales: Una defensa de los criterios de «sentido común»», in *Quaestio facti*, 3: 49-80. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: http://dx.doi.org/10.33115/udg_bib/qf.i3.22731

La prueba de los hechos psicológicos o estados mentales es un problema complejo de la teoría de la prueba jurídica, un problema que plantea profundas dificultades y perplejidades de carácter ontológico (¿existen realmente hechos mentales, distintos de los hechos físicos?), epistemológico (¿cómo puedo conocer los estados mentales de los demás?) e ideológico (¿son jurídicamente relevantes?). Hasta tal punto parecen irresolubles estas dificultades que una parte de la doctrina y de la jurisprudencia ha dicho que, en realidad, los estados mentales no son —o no deben ser tratados como— “hechos” propiamente, por lo que no se prueban (esto es, no se demuestra que hayan ocurrido o hayan existido), sino que para el Derecho son entidades ficticias que se imputan o adscriben a partir de ciertos criterios normativos¹. Este

¹ Lo cual tiene importantes consecuencias prácticas. Por ejemplo, en algunas sentencias del Tribunal Supremo y el Tribunal Constitucional español se ha argumentado que, al no tratarse propiamente de hechos, no les alcanza la presunción de inocencia: “Los juicios de valor sobre intenciones no pueden someterse a las exigencias de la presunción de inocencia por ser circunstancias inaprehensibles por los sentidos, razón por la cual solo los hechos en sí, sobre los que se base la inducción, pueden ser objeto de prueba...” (STS de 17 de febrero de 1995); “Tales juicios de valor sobre intenciones o juicios de inferencia sobre pensamientos, querer o deseos escondidos en el intelecto humano, no son hechos en sentido estricto porque no constituyen datos aprehensibles, por ello no pueden ser objeto de prueba propiamente”.

escepticismo contrasta con el hecho de que todos atribuimos con frecuente éxito estados mentales a los demás y conseguimos adaptar y coordinar nuestra conducta a la de los otros gracias a esas atribuciones. Adscribir estados mentales a los demás es la estrategia que ponemos en práctica para explicar y predecir su comportamiento, lo que a su vez es necesario para coordinarnos con ellos y participar de la vida social. Que en buena medida logremos comunicarnos, entendernos, colaborar, emprender tareas comunes, construir el mundo de reglas, convenciones e instituciones en el que vivimos requiere poder conocer los estados mentales de los otros. Por supuesto, abundan los desencuentros y malentendidos; pero eso no debe ocultar la importancia de los aciertos. Esta es una circunstancia que me parece que no debe perderse de vista cuando discutimos sobre la prueba de este tipo de hechos, porque, frente a las posturas escépticas, nos recuerda que el sentido común parece, en muchos casos, una estrategia adecuada para adscribir hechos mentales.

De hecho, determinar qué intención tiene un agente o cuál es su estado de ánimo es algo tan natural en nuestra vida cotidiana que puede considerarse una habilidad en la que somos expertos desde niños, aunque la ejerzamos de una manera intuitiva. Sabemos atribuir estados mentales, pero no sabemos cómo lo hacemos. Explicar cómo lo hacemos es un problema propio de disciplinas como la filosofía y las ciencias cognitivas (psicología, neurociencia, inteligencia artificial, etc.), cuyas respuestas quizá puedan contribuir a resolver algunos de los problemas que se les plantean a los juristas. En este trabajo voy a tratar de analizar qué luz puede aportar la filosofía de la mente (en particular, las discusiones acerca del llamado “problema de las otras mentes”) a la prueba de los hechos psicológicos en el Derecho.

1. UNA APROXIMACIÓN A LA FILOSOFÍA DE LA MENTE

En la filosofía de la mente podemos distinguir varios tipos de cuestiones:

a) La cuestión conceptual: ¿Qué son los estados mentales?

b) La cuestión ontológica:

(b.1) ¿Existen los estados mentales?

(b.2) ¿Existen otras mentes, además de la mía?

c) La cuestión epistemológica:

(c.1) ¿Cómo conozco mis propios estados mentales?

(c.2) ¿Cómo puedo conocer los estados mentales de los demás?

Las preguntas b.2 y c.2 constituyen el aspecto ontológico y epistemológico, respectivamente, del llamado problema de las “otras mentes”. Es la vertiente epistemo-

te dicha y quedan fuera de la garantía constitucional” (STC 195/1993). Sobre la cuestión de si la prueba del dolo es genuina prueba o imputación puede verse RAGUÉS, 1999 y 2002 y GONZÁLEZ LAGIER, 2004.

lógica de este problema lo que nos interesa aquí, pero digamos algo también sobre el resto de cuestiones, dado que todas están entrelazadas.

1.1. ¿Qué son los «estados mentales»?

Comencemos con la cuestión conceptual. Son estados mentales —usando la elegante enumeración de Richard Wollheim— “las percepciones, tales como oír un coro de maitines o ver una constelación de estrellas; las sensaciones, como el dolor, los picores, las punzadas de hambre o sed; los sueños y los ensueños; los momentos de desesperación, aburrimiento o lujuria; los momentos de inspiración; las evocaciones; las imágenes fantaseadas y las melodías oídas en la cabeza; y los pensamientos, tanto los pensamientos que pensamos, como los que se cuelan en nuestra cabeza sin invitación” (WOLLHEIM 2006: 24). Con una clasificación menos lírica, pero algo más sistemática, podríamos distinguir entre (a) voliciones (como los deseos y las intenciones), (b) cogniciones (creencias, conocimientos), (c) percepciones y sensaciones, (d) estados afectivos (emociones, estados de ánimo) y (d) actos mentales (planear, deliberar, decidir, etc.).

Pese a su diversidad, parece haber una serie de propiedades que permiten diferenciar este tipo de hechos de los hechos físicos (MOYA, 2006: Cap. I; SEARLE, 1990).

a) Consciencia: Normalmente, somos conscientes de los estados mentales que tenemos. Podemos entender la *consciencia* como una especie de conocimiento “de lo que sucede dentro de nosotros” (GARCÍA CARPINTERO, 1995: 46) no fundado en la evidencia empírica acerca de rasgos del mundo externo: Si me pellizco el brazo, experimento dolor y por ello sé que el brazo me duele, y lo sé con independencia de que me vea a mí mismo sacudirlo o gritar; si deseo un vaso de cerveza, sé que lo deseo, sin necesidad de esperar a ver si me levantaré e iré a la nevera; y si estoy preocupado, sé —porque lo siento— que lo estoy.

b) Qualia: Ahora bien, hay cierta diferencia cualitativa en cómo experimento mi dolor, mi deseo o mi preocupación. Decimos entonces que estos estados mentales se corresponden con diferentes sensaciones internas o *qualia*. Los *qualia* son la manera peculiar como cada estado mental emerge a mi consciencia, el componente fenomenológico de los estados mentales. La consciencia (y los *qualia*) establecen una importante asimetría entre la manera como conocemos nuestros estados mentales y la manera como conocemos los de los demás; este último conocimiento sí requiere una inferencia a partir de la observación de la conducta del otro. Un ejemplo que ayuda a entender la noción de *qualia* puede ser el siguiente: supongamos que María tiene una extraña enfermedad que no le permite ver los colores; todo lo ve en blanco y negro, como en las antiguas televisiones². María se ha interesado mucho por el fenómeno de los colores y ha estudiado e investigado todo lo que hay que saber sobre

² Adapto un ejemplo de Frank Jackson, comentado por MOYA, 1996 .

ellos, hasta convertirse en una eminencia en el tema. Desde el punto de vista físico y neurofisiológico sabe todo lo que hay que saber de los colores. Pero aún así podemos decir que le falta algo: le falta la experiencia de haber visto el rojo, el verde, el azul... Si de repente se curara de su enfermedad, aprendería algo más sobre los colores (si realmente lo sabe todo sobre los colores desde un punto de vista físico, incluyendo las reacciones neurofisiológicas correspondientes, pero le falta algo, ese algo no puede ser físico; por ello este ejemplo se ha usado también para mostrar que las propiedades mentales no son reducibles a propiedades físicas).

c) Contenido mental: Otra característica de muchos estados mentales consiste en poseer un *contenido representacional*, esto es, en tener la capacidad de versar o ser sobre objetos y estados de cosas del mundo (que, además, no tienen por qué estar presentes o existir en realidad, sino que pueden pertenecer al pasado o al futuro o incluso ser ficticios) distintos de ellos mismos (SEARLE, 1992: 18). La creencia de que mañana lloverá se refiere a un contenido (“mañana lloverá”). Dado un estado mental con contenido representacional, podemos distinguir entre su contenido y su modo psicológico: “Creo que mañana lloverá” y “Deseo que llueva mañana” son dos estados mentales que se diferencian por su modo psicológico, pero no por su contenido (a los estados mentales con contenido mental podemos llamarlos también actitudes proposicionales).

d) Subjetividad: Otra característica propia de los estados mentales es su *modo subjetivo de existencia*. Mis dolores, temores, odios, deseos y creencias son exclusivamente míos. En opinión de John Searle se trata no solo de una subjetividad epistemológica (la consciencia), sino también ontológica: “Consideremos, por ejemplo, el enunciado ‘tengo dolor de espalda’. Tal enunciado es completamente objetivo en el sentido de que lo convierte en verdadero la existencia de un hecho real y no depende de las actitudes de los observadores. Sin embargo, el fenómeno mismo, el dolor real mismo, tiene un modo subjetivo de existencia (SEARLE 1992:21)

e) Causalidad mental: A todas estas propiedades hay que añadir el hecho de que los estados mentales no son causalmente inertes, sino que interactúan entre sí y con estados físicos; esto es lo que los filósofos de la mente llaman *causalidad mental*. Los estados mentales parecen causar hechos físicos: nuestras emociones, creencias y deseos causan nuestro comportamiento. Mi miedo a que un perro me muerda causa que salga corriendo y mi deseo de saciar la sed, junto con las creencias adecuadas, causa que vaya a la nevera a por una cerveza. Pero para que se hable de causalidad mental el efecto debe producirse precisamente en virtud del contenido mental. Una manera de explicar que los estados mentales causan hechos físicos consiste en afirmar que tienen un componente o sustrato físico (neuronal) y atribuir su poder causal a ese sustrato. Por ejemplo, el miedo a que el perro me muerda causa mi conducta porque se corresponde con ciertos cambios neurofisiológicos que disparan señales eléctricas que estimulan los músculos. Sin embargo, la relación entre esos estados neuronales y la conducta no sería un caso de causalidad mental, porque en esta explicación el contenido de la emoción (el miedo al perro, la creencia de que ese perro pone en riesgo

mi integridad física) no juega ningún papel. Supongamos una soprano que hace estallar un vaso cantando el aria *Il dolce suono* de Donizetti. El aria posee un significado (“*Il dolce suono mi colpi di sua voce! Ah, quella voce m’e qui nel cor discesa!...*”), pero no es el significado, sino las vibraciones provocadas por la voz de la soprano, las que rompen la copa (con total independencia de su significado), por lo que este no sería un caso de causalidad mental. El desafío que plantea la causalidad mental es explicar que nuestras razones para actuar tienen influencia o producen una diferencia en la probabilidad de que actuemos de una u otra manera que varía según su contenido.

Tomemos como ejemplo los estados mentales que llamamos intenciones: La consciencia de la intención consiste en la experiencia, no basada en evidencia empírica externa, de nuestra pretensión de hacer algo: si se trata de la intención de hacer algo en un futuro próximo o lejano, la certeza de que –a menos de que algo me lo impida o cambie de opinión– efectivamente lo haré; si se trata de la intención en el momento de actuar, el conocimiento del tipo de acción que estoy realizando (HAMP-SHIRE y HART, 1958: 1 y ss.). El contenido representacional de la intención consiste en las acciones u objetivos que procuro realizar o alcanzar. La subjetividad ontológica consiste en el carácter privado y particular de las intenciones (hay cierto sentido en que otro sujeto y yo no podemos compartir la misma intención, porque mi intención es solo mía: aunque tengamos intenciones de hacer lo mismo). Por último, la intención tiene alguna relación (en virtud de su contenido) con mi conducta física, ejerce algún control sobre ella.

Es importante observar que estas propiedades no son condiciones necesarias y suficientes, ni individual ni conjuntamente, que sirvan para completar una definición de la noción de “hecho psicológico”. En realidad, no todos los estados mentales tienen estas propiedades. Por ejemplo, quizá tenga sentido hablar de estados mentales no conscientes. También es dudoso que todos los estados mentales conscientes posean un contenido representacional, o que todos posean un aspecto fenomenológico o *qualia*. Es discutible si las sensaciones, como los dolores, la picazón, etc., tienen en algún sentido un contenido mental (hay quien dice que un dolor en el brazo tiene como contenido la parte del brazo que te duele, pero resulta ciertamente forzado), o acerca de si las actitudes proposicionales (como las intenciones, creencias, etc.) tienen *qualia* o carecen por completo de él. Una definición sugerente de estado mental es la que establece una condición necesaria disyuntiva: x no puede ser considerado un estado mental si (a) no tiene un contenido representacional o (b) no posee carácter fenomenológico (*qualia*). En todo caso, estas dos propiedades de lo mental son sumamente importantes y se puede trazar una distinción entre aquellos estados mentales cuya principal característica es que poseen un contenido representacional articulado o conceptual y aquellos cuya característica más llamativa es poseer un componente fenomenológico. Esto permitiría clasificar los deseos, las intenciones y las creencias en el primer grupo; las sensaciones en el segundo, y las emociones (que, muy a grandes rasgos, se pueden ver como asociaciones entre creencias y sensaciones) se clasificarían como fenómenos mixtos.

1.2. ¿Existen los estados mentales?

Las propiedades de lo mental son difíciles de encajar en nuestra imagen científica y materialista del mundo. ¿Cómo puede un cerebro, algo completamente material, tener consciencia? ¿Cómo puede versar sobre algo, esto es, tener un contenido semántico? ¿Cómo puede ser real algo esencialmente subjetivo? ¿Y cómo puede lo mental interactuar con lo físico? Cuando nos planteamos estas cuestiones, surge el problema ontológico de lo mental: Si los estados mentales están correctamente caracterizados por las anteriores propiedades, ¿existen realmente? Las distintas concepciones de la mente pueden interpretarse como intentos de responder a esta pregunta³.

Las concepciones dualistas consideran que existen dos tipos de sustancias o entidades: las entidades materiales o físicas y las entidades mentales. Para estas teorías, los estados mentales existen, con las propiedades señaladas anteriormente, y tienen un tipo de existencia distinto e independiente de las entidades físicas. Tampoco plantea un problema su existencia para el monismo idealista (que, sin embargo, niega la existencia de los hechos físicos). El monismo materialista, por el contrario, al sostener que solo existen entidades físicas, se enfrenta al problema de dar cuenta de las propiedades de lo mental. Por ello, dentro del materialismo podemos distinguir tres tipos de actitudes ante los estados mentales: las posiciones eliminacionistas, las posiciones reductivistas y las no reductivistas.

El *materialismo eliminacionista* cree que no hay estados mentales, tal como lo considera la psicología y el sentido común. Hay estados cerebrales, actividad neuronal, pero presuponer que la neurociencia permitirá establecer una correlación entre estados cerebrales y mentales es dar erróneamente carta de existencia a estos últimos: la psicología está equivocada, trata de dar explicaciones de la conducta que la investigación científica mostrará que no logran decir nada verdadero sobre la realidad. Como la teoría del flogisto (una supuesta sustancia, en realidad inexistente, que se creía que se liberaba en los procesos de combustión u oxidación), las teorías y los conceptos mentales están condenados a desaparecer, simplemente porque hacen referencia a cosas que no existen. A medida que avance la neurociencia, podremos sustituir el lenguaje de los estados mentales por el lenguaje de los estados cerebrales, de igual manera que el descubrimiento de los virus y las bacterias permitió rechazar las explicaciones de las enfermedades en términos de demonios, maldiciones y espíritus. Este sería, simplemente, un ámbito más donde la extensión de la ciencia permitiría la expulsión de la superstición.

El *materialismo reduccionista* no impugna el sentido de los conceptos mentales (mantiene la psicología como una disciplina epistemológicamente adecuada o conveniente, al menos en cierto grado), pero cree que las teorías mentales ganarán en precisión y poder explicativo si son traducidas o reducidas a un lenguaje en última

³ Una presentación de estas concepciones y sus problemas puede verse en MOYA, 2006.

instancia no mental. Una teoría es reducida a otra, de acuerdo con Paul Churchland, cuando “una teoría nueva y eficaz logra abarcar un conjunto de proposiciones y principios que reflejan perfectamente (o casi perfectamente) las proposiciones y principios de una teoría o marco conceptual anteriores” (CHURCHLAND 1999: 43). Si el ejemplo preferido de los eliminacionistas es el flogisto, el de los reduccionistas es la explicación del calor en términos de promedio de energía cinética de las moléculas (a diferencia de lo que ocurría con el flogisto, las teorías o leyes que usan el concepto de calor, como la que afirma que el calor dilata los metales, son útiles, pero su reducción a la nueva teoría proporciona una descripción más precisa del fenómeno). El conductismo (reducción a conducta), el materialismo de la identidad de tipos (reducción a actividad cerebral) y el funcionalismo (reducción a funciones y roles causales) son concepciones reduccionistas.

El *materialismo no reduccionista* sostiene que la traducción del lenguaje mental al lenguaje físico es imposible sin pérdidas significativas: aunque haya una sola sustancia, lo físico, podemos hablar de estados o propiedades cerebrales y de estados o propiedades mentales. El materialismo no reduccionista puede verse como un monismo respecto de las sustancias y, simultáneamente, un dualismo respecto de las propiedades. Su desafío, por tanto, es explicar cómo de lo físico pueden surgir las propiedades de lo mental. Una estrategia interesante (dejo de lado el monismo anómalo de Davidson) es la del emergentismo. El *emergentismo* es una postura materialista que identifica los estados mentales con propiedades emergentes o sistémicas, es decir, propiedades que surgen a partir de cierto grado de complejidad de un sistema. Son propiedades del conjunto del sistema, pero no de cada una de sus partes. Para explicar qué es una propiedad emergente John Searle —cuya concepción, que él mismo llama *naturalismo biológico*, es una teoría emergentista— recurre al siguiente ejemplo: “Supongamos que tenemos un sistema S, compuesto de los elementos *a, b, c...* Por ejemplo, S podría ser una piedra y los elementos podrían ser moléculas. En general, habrá rasgos de S que no son, o no necesariamente, rasgos de *a, b, c...* Por ejemplo, S podría pesar 20 kilogramos, sin que las moléculas individualmente pesen 20 kilogramos. Denominemos a estos rasgos ‘rasgos del sistema’. Algunos rasgos del sistema pueden ser deducidos, o determinados, o calculados a partir de los rasgos *a, b, c...*, simplemente por la forma en que se componen y ordenan (y, a veces, por sus relaciones con el entorno). Ejemplos de ellos serían la forma, el peso y la velocidad. Pero algunos otros rasgos del sistema no pueden ser determinados solo a partir de los elementos que los componen y de las relaciones con el entorno: han de ser explicados a partir de las relaciones entre los elementos. Llamémosles ‘rasgos del sistema causalmente emergentes’. La solidez, la liquidez y la transparencia son ejemplos de rasgos del sistema causalmente emergentes” (SEARLE, 1996: 121). De manera que, para Searle, una propiedad emergente es una propiedad que surge a partir de los elementos que componen el sistema, de las relaciones del sistema con el entorno y —esto parece lo decisivo— de las relaciones de esos elementos entre sí. Los estados mentales se identifican con propiedades emergentes de los cerebros. Los cerebros son sistemas físicos

que han ido adquiriendo complejidad a lo largo de su historia evolutiva. El naturalismo biológico es descrito por Searle a partir de cuatro tesis (SEARLE, 2006: 147 y ss.):

1) Los estados mentales, con su ontología subjetiva de primera persona, son fenómenos reales del mundo real.

2) Los estados mentales son causados en su totalidad por procesos neurobiológicos de nivel inferior localizados en el cerebro.

3) Los estados mentales son rasgos del sistema cerebral en su conjunto y existen, por tanto, en un nivel superior al de las neuronas y sus sinapsis (“Por sí misma —dice Searle—, una neurona no es consciente, pero las partes del sistema cerebral compuestas por ella sí lo son”).

4) Los estados mentales, en la medida en que son reales, tienen eficacia causal (“Mi sed consciente, por ejemplo, me lleva a tomar agua”).

El materialismo emergentista no es reductivista y, por tanto, no elimina la perspectiva de primera persona, a diferencia del conductismo o la tesis de la identidad de instancias. De acuerdo con Searle, aunque los estados mentales pueden ser reducidos causalmente a —en el sentido de explicados por— los componentes del cerebro y sus relaciones, esta explicación no es suficiente, porque lo característico de lo mental es que es un fenómeno en primera persona, requiere un lenguaje subjetivo que no puede ser descrito con el lenguaje objetivo de la neurociencia.

El siguiente esquema trata de dar cuenta de todas estas concepciones de la mente:

1. Dualismo (dos sustancias: materia y mente)
2. Monismo (una sustancia).
 - 2.1. Monismo idealista (todo es espíritu/mente).
 - 2.2. Monismo materialista (todo es materia).
 - 2.2.1. Materialismo eliminacionista (no hay estados mentales, es una descripción falsa).
 - 2.2.2. Materialismo reductivista (los estados mentales son estados de un tipo distinto).
 - 2.2.2.1. Conductismo (reducción a conducta).
 - 2.2.2.2. Tesis de la identidad de tipos (reducción a eventos neurofisiológicos)
 - 2.2.2.3. Funcionalismo (reducción a roles causales).
 - 2.2.3. Materialismo no reductivista.
 - 2.2.3.1. Tesis de la identidad de instancias (los estados mentales son estados físicos, pero no hay correlación entre tipos).
 - 2.2.3.2. Emergentismo (los estados mentales son propiedades emergentes. Monismo de sustancias, dualismo de propiedades).

En lo que sigue se dará por sentado que es posible conciliar, a la manera del emergentismo, la existencia de estados mentales (tal como han sido caracterizados) con una concepción básicamente materialista del mundo (de acuerdo con la cual el sustrato físico es una condición de posibilidad de la existencia de los estados mentales, que son generados por él).

1.3. ¿Existen otras mentes, además de la mía?

El problema de las “otras mentes” comienza a plantearse cuando asumimos que tiene sentido hablar de la existencia de estados mentales y nos preguntamos si es posible que, además de mi mente, existan otras mentes, esto es, si el resto de sujetos poseen también una mente. Como hemos visto, parece que debe aceptarse una asimetría entre la forma como conocemos nuestros propios estados mentales (conocimiento sin observación ni inferencia) y la forma como conocemos los estados mentales de los demás (conocimiento indirecto, que requiere una inferencia a partir de su comportamiento) (DAVIDSON, 2003: 25 y ss.; MOYA, 1996: 157). Además, en cierto sentido, el conocimiento que tenemos de nosotros mismos, aunque puede discutirse si es infalible, está claramente dotado de una autoridad de la que carece el conocimiento de los estados mentales de terceros. No puedo dudar de que yo pienso, siento, percibo..., pero no puedo tener esa misma convicción respecto de los demás. ¿Cómo puedo saber si existen “otras mentes”, además de la mía?

La respuesta clásica a esta pregunta, defendida por autores como Agustín de Hipona o John Stuart Mill, es que lo sabemos (lo suponemos) por analogía (CASTRO, 2017: 48). Del mismo modo que mi mente y mi conducta están relacionadas, de manera que sé que son mis estados mentales los que producen mi comportamiento, puedo asumir que la conducta que observo en los demás, que es básicamente análoga a la mía, está causada igualmente por sus estados mentales. John Stuart Mill lo expuso de la siguiente manera:

Soy consciente en mí mismo de una serie de hechos conectados por una secuencia uniforme, el principio de la cual es la modificación de mi cuerpo, el punto medio son las sensaciones, y el final, la conducta exterior. En el caso de los otros seres humanos, tengo la evidencia de mis sentidos para el primer y último eslabón de esa serie, pero no para el eslabón intermedio. Sin embargo, descubro que la secuencia entre el primero y el último es tan regular y constante en esos casos como en el mío. En mi propio caso, sé que el primer eslabón produce el último a través del eslabón intermedio y que no lo podría producir sin él. Por tanto, la experiencia me obliga a concluir que debe existir un eslabón intermedio; que debe ser en otros, o bien el mismo que en mí, o bien otro diferente; [...] al suponer que el eslabón es de la misma naturaleza [...] actúo de acuerdo con las reglas legítimas de la investigación experimental (MILL, 1867; tomado de DANCY, 1993: 87).

El argumento proyecta a los demás una relación entre estados mentales y comportamiento a partir de mi conocimiento fiable de que mis propios estados mentales están detrás de mi propio comportamiento; sin embargo, se trata de un argumento demasiado débil para satisfacer al escéptico, que puede señalar que, en realidad, se trata más bien de una inducción a partir de un solo caso⁴. Se podría aducir en respuesta a esta objeción que no se trata en realidad de un único caso: lo que proyecto son *múltiples* experiencias en las que observo esta conexión entre mis estados mentales y mis acciones; no obstante, todos los casos están referidos a mí, lo que limita la fuerza de la conclusión.

La anterior crítica puede sortearse completando el argumento con una inferencia a la mejor explicación. Habría varias formas de hacerlo. Podríamos decir, por ejemplo, que la mejor explicación de que la conducta de los demás sea semejante a la mía, frente a alternativas del tipo de que sean autómatas o marionetas de algún dios, es que posean mentes como la mía; o podría decir que la mejor explicación del éxito de la estrategia de atribuir estados mentales como los míos a los demás para explicar y predecir su conducta es que efectivamente tienen estados mentales como los míos. En todo caso, la existencia de otras mentes, aunque es un importante desafío filosófico, suele aceptarse como un presupuesto de nuestra comprensión del mundo: no parece existir un solipsista convencido que lleve su convicción a su conducta cotidiana.

1.4. ¿Cómo conozco mis propios estados mentales?

Pasemos al problema epistemológico de los estados mentales. Como hemos visto, se pueden plantear dos preguntas acerca del conocimiento de los estados mentales: Cómo conocemos nuestros estados mentales y cómo conocemos los estados mentales de los demás.

La primera cuestión suele responderse, como hemos adelantado, recurriendo a un conocimiento directo de nuestros estados mentales, un conocimiento que no requiere la observación de nuestro comportamiento por parte de nosotros mismos, ni ninguna inferencia a partir del mismo; y además un conocimiento privilegiado, en el sentido de dotado de una especial autoridad o una “presunción de veracidad” (MOYA, 1996). A veces se dice que este modo de conocimiento consiste en “volver los ojos hacia dentro”, “hacia uno mismo”, en un ejercicio de introspección. Pero esta forma de describir las cosas hace pensar que este autoconocimiento es el resultado de un esfuerzo intencional de examinarse a sí mismo; esto puede ser así en ocasiones, pero la consciencia de nuestros propios estados mentales suele ser más “directa”, sin requerir un esfuerzo consciente. Incluso puede sostenerse, como sugiere Carlos Moya, que en el momento en que me pregunto por mis estados mentales y hago el

⁴ Sobre los problemas planteados por este argumento puede verse DANCY, 1993: 87 y ss.

esfuerzo de sacarlos a la luz por medio de una inferencia, estos se convierten en algo distinto de lo que eran (MOYA, 1996: 182). Podemos hacer auto—atribuciones de estados mentales respecto del pasado, del presente, del futuro o respecto de situaciones hipotéticas, pero cuando se postula un conocimiento sin inferencia este parece darse solo a propósito de los estados mentales presentes, los que ocurren en ese momento (las autoatribuciones hacia el pasado son recuerdos o reconstrucciones, y el resto de autoatribuciones parece resultado de inferencias; esto no excluye, sin embargo, que el conocimiento de nuestros estados mentales pasados o las predicciones hacia el futuro de los mismos carezcan necesariamente de una autoridad privilegiada, en la medida en que ya tenemos experiencias acerca de cómo normalmente pensamos y sentimos) (SKIDELSKY, 2011: 35).

Una forma sugerente (pero no exenta de problemas) de explicar este tipo de conocimiento consistiría en verlo como una relación entre estados mentales de diverso orden (SKIDELSKY, 2011: 36). Así, por ejemplo, si yo sé que creo que la tierra es redonda tengo dos estados mentales: la creencia de que la Tierra es redonda, que sería un estado mental de primer nivel, y el estado mental de segundo nivel consistente en ser consciente de que creo que la tierra es redonda (el conocimiento de que creo que la tierra es redonda). La consciencia, entonces, es un estado mental de nivel superior, un “pensamiento” sobre los estados de primer nivel. Pero ¿cómo se relacionan ambos niveles?

De acuerdo con el modelo de la consciencia como “percepción” entre ambos niveles hay una relación de carácter epistemológico y empírico, que consistiría en alguna relación causal entre los estados de nivel inferior y los de nivel superior. La consciencia sería una especie de “percepción” de estados mentales de primer nivel (MOYA, 1996: 159 y ss.). La presunción de veracidad de las autoatribuciones de estados mentales descansaría entonces en esta “percepción” especial que solo yo poseo respecto de mis estados mentales, en el hecho de que somos espectadores privilegiados de nuestra mente. Pero también se ha sugerido que la prioridad epistemológica de la primera persona sobre sus propios estados mentales tiene una dimensión constitutiva o conceptual (es parte de lo que significa tener un estado mental que el sujeto que lo tiene tenga una autoridad especial sobre el mismo) o normativa (las autoatribuciones implican aceptar el compromiso de actuar conforme al estado mental autoatribuido), o simplemente descansaría en que somos expertos en nosotros mismos (ya sabemos qué estados mentales hemos tenido en circunstancias semejantes en el pasado y cómo hemos actuado, lo que da más peso e inmediatez a las inferencias que hacemos sobre nuestros propios pensamientos presentes) (SKIDELSKY, 2011: 35 y ss.).

Considerar que nuestro conocimiento de los estados mentales es una especie de “percepción” (cuyo objeto no sería un hecho externo sino otro estado mental o ciertas propiedades del mismo) plantea varios problemas: por ejemplo, parece iniciar un regreso al infinito (¿cómo somos conscientes de que somos conscientes de tener una determinada creencia?) y, respecto de algunos estados mentales, parece redundante (puede decirse que no hay distinción entre tener un dolor y ser consciente de que

tengo un dolor: son exactamente el mismo estado mental). Además, esta concepción del autoconocimiento parece casar bien con el tipo de estados mentales en los que predomina su aspecto fenomenológico, como los dolores, la sensación de hambre, las punzadas, etc., pero no con las actitudes proposicionales o estados representacionales. Puede pensarse que el “qualia” asociado a los primeros nos “informa” del estado mental, pero no parece haber ninguna sensación asociada a los estados mentales representacionales. Creer que la tierra es redonda puede generar, en ciertas situaciones, estados mentales (emociones, por ejemplo) que sí tienen asociada alguna sensación, pero la creencia misma no la tiene. Luego el conocimiento de los estados mentales proposicionales no parece encajar bien con la idea de percepción causada por el estado mental subyacente, salvo que se establezca una correlación necesaria entre estos estados mentales y estados fenomenológicos. Por ejemplo, si se concluyera que creencias, deseos y emociones suelen estar asociadas, es decir, se presentan generalmente de manera simultánea, podría decirse que estas últimas cumplen un papel de “informantes” de nuestros estados mentales no fenomenológicos.

Sea como sea, vale la pena dejar establecido que si se rechaza el conductismo como explicación de la mente (y creo que hay buenas razones para hacerlo. Véase GONZÁLEZ LAGIER, 2003: 651), las autoatribuciones de conocimiento de nuestros estados mentales presentes no están basadas en la observación de nuestro comportamiento, a diferencia de lo que parece ocurrir con la atribución de estados mentales a terceros. Vale la pena también recordar que la autoridad de las primera persona respecto de sus propios estados mentales no implica necesariamente infalibilidad: respecto de los estados psicológicos presentes, solo tendría sentido predicar infalibilidad si son conscientes; cuando nos atribuimos estados mentales futuros o hipotéticos hacemos una inferencia que podría estar equivocada; y cuando nos atribuimos estados mentales pasados, la intermediación de los recuerdos y la posibilidad de reconstrucción de tales estados mentales (por ejemplo, interpretando nuestro comportamiento bajo su mejor luz) también pueden distorsionar la corrección de la atribución.

Sin embargo, a pesar de la asimetría, la capacidad de conocer nuestros propios estados mentales parece cumplir, como veremos, un papel importante en el conocimiento de estados mentales de terceros. Esto nos lleva al problema epistemológico de las otras mentes.

1.5. ¿Cómo conozco los estados mentales de los demás?

Aceptemos que existen los estados mentales, con las propiedades que hemos mencionado anteriormente, sean o no reducibles unas a otras; y aceptemos que, como yo, los demás agentes también tienen una mente: si es así, ¿cómo conocemos el contenido de sus mentes? ¿cómo conocemos los estados mentales de los demás? Esta es una cuestión que recientemente se ha revitalizado gracias a ciertos descubrimientos neurocientíficos, como veremos enseguida. Es, además, la cuestión que más nos interesa para la prueba de los estados mentales.

La filosofía y las ciencias cognitivas se han planteado esa pregunta desde la perspectiva de quien asume que los seres humanos hacemos atribuciones psicológicas frecuentemente, con la finalidad de explicar y predecir la conducta de los demás, coordinarnos con ellos, entenderlos, cooperar y, en definitiva, construir una vida social. Desde esta perspectiva, lo que les interesa no es si podemos hacer atribuciones psicológicas exitosas, sino cómo lo hacemos. Las respuestas que se han dado se pueden clasificar en función de que se adopte una perspectiva de *tercera persona* o de *primera persona*. Para las primeras, el intérprete adopta una perspectiva externa respecto del sujeto interpretado, la perspectiva de un observador; para las segundas, el intérprete trata de ponerse en el lugar del interpretado (una exposición y discusión en detalle de las distintas teorías puede verse en GOLDMAN, 1989 y 2006; SKIDELSKY, 2001; BRUNSTEINS, 2008 y 2011; QUINTANILLA, 2019).

Las teorías que adoptan la perspectiva de *tercera persona* suelen recurrir a las siguientes estrategias:

a) La primera, que suele denominarse “teoría de la teoría”, hace una analogía entre la manera como conocemos el mundo físico y la manera como conocemos los estados mentales de los demás, en el sentido de que en ambos casos hacemos inferencias basadas en observaciones que subsumimos en teorías o generalizaciones empíricas. De la misma manera que inferimos que este trozo de metal se ha calentado del hecho de que observamos que se ha dilatado y de una ley de acuerdo con la cual “el calor dilata los metales”, o predecimos que el metal se va a dilatar a partir de la observación de que ha aumentado su temperatura y de la misma ley, explicamos que x echó a correr porque vio un perro, tiene miedo a los perros y aquellos que tienen miedo a algo normalmente lo rehúyen. La idea es que todos tenemos (al menos, a partir de cierta edad) una “teoría de la mente” (que a veces se denomina “psicología popular”). Esta teoría puede ser aprendida o innata (así, algunos autores postulan que, al igual que poseemos un módulo para el lenguaje, nuestro cerebro posee también un módulo o sistema encargado de la atribución psicológica), y se puede entender en un sentido más o menos articulado. Su contenido sería un conjunto de conceptos y “leyes” o regularidades nomológicas que correlacionarían estados mentales internos y conducta externa. A partir de estas “leyes” y de la observación de la conducta de los demás, podemos inferir sus estados psicológicos. Estas generalizaciones, que serían “de sentido común”, dirían cosas como “Las personas con dolor tienden a aliviar ese dolor”, “Las personas que desean determinado fin F y creen que M es un medio adecuado para dar lugar a F, intentarán hacer M” o, en general, “Cuando alguien está en una tal y cual combinación de estados mentales y recibe estímulos sensoriales del tipo tal y cual, tiende con tal y cual probabilidad a pasar a tales y cuales estados mentales y producir tales y cuales respuestas” (SKIDELSKY, 2001: 39).

La correlación entre estados mentales y conductas que expresan estas “leyes” no puede ser más que una aproximación probabilística, porque es posible que un mismo estado mental dé lugar a acciones distintas o que una misma acción pueda explicarse

por estados mentales distintos. Por ello, la teoría de la mente puede completarse con una concepción de la racionalidad. Esto nos lleva a la segunda estrategia.

b) La segunda es la estrategia de la “racionalización”. Varios autores han sugerido que la atribución psicológica se basa en la idea de que los seres humanos son racionales y actúan siempre bajo parámetros explicables a la luz de la racionalidad.

Existen muchos modelos de interpretación de la conducta (y del significado) basados en la idea de que hay que tratar a los agentes como sujetos “racionalizadores”. Uno de los más conocidos y discutido es el modelo de Donald Davidson. De acuerdo con esta idea, la atribución de estados mentales es una interpretación o comprensión de la conducta de los otros basada en el “principio de caridad” (Davidson ofreció varias formulaciones de este principio, pero se considera clásica la que puede encontrarse en DAVIDSON, 1995: 281; véase también GOLDMAN, 1989: 163 y ss.). Este principio estaría compuesto por un conjunto de asunciones (relacionadas con la racionalidad del agente) que serán condición de posibilidad de la interpretación de la conducta. De acuerdo con Pablo Quintanilla (QUINTANILLA, 2004: 86), el principio de caridad requiere que quien atribuye los estados mentales al otro asuma los siguientes postulados:

- (i) El comportamiento del agente es consistente.
- (ii) El agente mantiene los mismos supuestos lógicos que el intérprete.
- (iii) Las creencias del agente son verdaderas.
- (iv) El agente se comporta siguiendo su mejor juicio.

Los postulados (i) y (iv) nos permitirían, por ejemplo, inferir las intenciones de un agente a partir de sus creencias y de su conducta, o las creencias a partir de su intención y su comportamiento; el postulado (iii) nos permitiría atribuirle las mismas creencias que tengo yo (porque son las que yo considero verdaderas); etc.

Es importante darse cuenta de que, desde una perspectiva racionalizadora, un rasgo relevante de los estados mentales es la coherencia que mantienen entre sí: intenciones, deseos, creencias y acción mantienen relaciones de consistencia, de manera que la posesión de uno de estos estados mentales lleva a, o presupone, la de otro. A veces se ha llamado a esta característica “el holismo de lo mental”. La atribución consiste en buena medida en encontrar un ajuste coherente (racional) entre los estados mentales implicados y la acción.

El principio de caridad, en definitiva, nos pide que aceptemos que el agente que tratamos de entender se comporta racionalmente. Su justificación puede buscarse en un argumento trascendental, como hacen muchos autores (es un presupuesto que hay que aceptar si queremos atribuir estados mentales a los demás y explicar su conducta) o puede tomarse como una generalización empírica (de hecho, los agentes suelen comportarse de forma racional). Más adelante volveremos sobre el papel de la racionalidad en la atribución de los estados mentales y en la justificación de esta atribución.

Frente a las teorías que adoptan una perspectiva de tercera persona, otros autores han propuesto una perspectiva de *primera persona*. Si desde la anterior perspectiva las nociones centrales en la explicación de nuestro modo de atribuir estados mentales eran la noción de teoría o la de racionalidad, aquí son las de empatía e imaginación. El punto de partida de la perspectiva de primera persona es el autoconocimiento que las personas tienen de sus estados mentales y su capacidad de ponerse en lugar del otro. La atribución psicológica consistiría en un proceso de imaginarse en la situación del otro para atribuirle estados mentales. Esta atribución podría hacerse de dos formas: imaginando que yo, con mis características, estoy en la posición del otro y atribuyéndole por tanto al otro los estados mentales que yo creo que tendría en esa situación o tratando de ponerme completamente en el lugar del otro (es decir, asumiendo las características del otro) y atribuyéndole los estados mentales que creo que tendría si fuera él (SKILDESKY, 2011). Es decir: (a) imaginando que yo estoy en su lugar o (b) imaginando que yo soy él. En todo caso, el proceso constaría de tres fases: (a) Imaginarse en el lugar del otro, o siendo el otro; (b) determinación de los estados mentales que tendríamos en esa situación y (c) atribución de esos estados mentales al otro.

En todo caso, lo que esta teoría sostiene es que la atribución psicológica se realiza de una manera espontánea, sin inferencia o reflexión; de una manera semejante, por tanto, al proceso de autoconocimiento, pero simulando que soy el otro o que estoy en su lugar (por ello esta teoría recibe el nombre, no muy apropiado en español, de “*teoría de la simulación*”), lo que requiere capacidades empáticas del intérprete. Se postula que este conocimiento es más intuitivo que racional, más “sentido” que “inferido”. Esta concepción de la atribución de estados mentales se ha visto fortalecida con el descubrimiento de las neuronas espejo: neuronas que se activan tanto cuando se realiza un movimiento o se tiene una experiencia como cuando se observa que otro realiza ese movimiento o tiene esa experiencia (RIZZOLATI y CRAIGHERO, 2004, IACOBONI, 2013). Estas neuronas no se limitan a la percepción de las acciones, sino que se postula que activan las mismas intenciones y emociones que motivan la acción observada, aunque existe también un mecanismo que inhibe la realización de la acción (es decir, algún mecanismo que informa de que se trata de una situación ficticia). Además, se postula que las neuronas espejo no solo se activan cuando “vemos” a la otra persona, sino también en situaciones ficticias (por ejemplo, leyendo una novela).

Aunque la perspectiva de tercera y de primera persona se han presentado como excluyentes, es evidente que resulta fácil conciliarlas. Por ejemplo, se puede sostener que son recursos distintos que los sujetos podemos poner en marcha sucesivamente: los procesos directos y automáticos que postula la teoría de la simulación podrían proporcionar la primera información sobre los estados psicológicos de los demás, que luego se verían respaldados o corregidos por las inferencias basadas en generalizaciones empíricas y/o perspectivas racionalizadoras. Esto casaría bien con las concepciones de los neurocientíficos y psicólogos que postulan dos sistemas generales

de conocimiento y toma de decisiones, uno más rápido e intuitivo y otro lento y reflexivo (KAHNEMAN, 2012; HAIDT, 2012). Otra manera de compatibilizar ambas teorías consistiría en recordar la distinción entre estados mentales con contenido representacional y estados mentales en los que predomina el aspecto fenomenológico. Las teorías basadas en los principios de racionalidad o de caridad parecen más plausibles para el primer tipo de estados mentales que para los del segundo tipo, dado que la racionalidad parece requerir estados proposicionales. Una creencia o una intención pueden ser racionales o irracionales, pero un dolor o una punzada no. La teoría de la simulación y la teoría de la teoría sí serían útiles para la adscripción de estados fenomenológicos. La tendencia, en general, ha sido la de construir teorías mixtas que integran elementos de una y otra perspectiva.

Tampoco son teorías exhaustivas sobre los modos de atribución de estados mentales, esto es, cabe hablar de otros modos de conocimiento de la mente de los demás: por ejemplo, se puede plantear una “percepción directa” de sus pensamientos a partir de gestos y expresiones, que serían “significantes naturales” de los estados mentales (KALPOKAS, 2019). Y también se han propuesto teorías desde la perspectiva de la *segunda persona*, que ponen el acento en la relevancia de la interacción y comunicación entre los sujetos a la hora de interpretar mutuamente sus estados psicológicos.

Por último, las tres teorías acerca de cómo atribuimos estados mentales a los demás parecen presuponer la capacidad de conocer los propios estados mentales: las tres estrategias parecen presuponer una similitud entre los demás y nosotros, es decir, asumen el argumento de la analogía que hemos mencionado acerca de la existencia de las otras mentes. La construcción de las teorías basadas en regularidades empíricas o “leyes” que usamos para explicar la conducta de los demás, según la “teoría de la teoría”, probablemente combinan la observación de la conducta de los demás con la reflexión sobre nuestra propia conducta; la estrategia de la racionalización, como hemos visto, presupone que consideramos que los deseos de los demás son razonables y sus creencias verdaderas, esto es, que son aproximadamente los mismos deseos y las mismas creencias que tenemos nosotros; y la teoría de la simulación proyecta a terceros los estados mentales que, visto desde una perspectiva de primera persona, tendríamos contrafácticamente en determinadas circunstancias.

Hasta aquí un rápido repaso a algunas cuestiones centrales de la filosofía de la mente. ¿Puede ser útil algo de todo esto para la prueba de los estados mentales en el Derecho?

2. LA PRUEBA DE LOS ESTADOS MENTALES EN EL DERECHO

2.1. El argumento ontológico contra la prueba de los estados mentales

Como hemos visto al principio de este trabajo, es una opinión bastante extendida entre los juristas que las dificultades para probar los estados mentales son de tal calibre que se debe renunciar a esa pretensión y sustituirla por una imputación a partir de criterios normativos. La prueba de la verdad no sería realmente un caso de prueba genuina, en el sentido de que no consistiría en una actividad cognoscitiva, orientada a establecer juicios que son verdaderos o falsos, sino una actividad dirigida a calificar la acción del agente de acuerdo con ciertos criterios que pueden satisfacer intereses muy variados (como el carácter ejemplificativo de la pena, la resolución de conflictos, etc.), pero no orientados al descubrimiento de una realidad psicológica. Se suele sostener esta posición con tres argumentos: El primero, de carácter ontológico, toma las peculiaridades de lo mental para negar su compatibilidad con una perspectiva materialista del mundo y sostener su inexistencia; el segundo, de carácter epistemológico, gira en torno a la imposibilidad de conocer los estados mentales de los demás, especialmente dadas las limitaciones de la prueba judicial; el tercero, de carácter ideológico, sostendría que los estados mentales no son realmente relevantes para el Derecho (aunque pudieran ser conocidos) (GONZÁLEZ LAGIER, 2006: 178 y ss.).

La filosofía de la mente esbozada en el anterior apartado proporciona objeciones contra los dos primeros argumentos. Así, frente al argumento eliminacionista, cabe aducir que negar la existencia de estados mentales es sostener algo muy implausible, lejos del sentido común, pues se estaría negando todas las experiencias que tenemos en primera persona: se estaría negando que tenemos consciencia de nuestros pensamientos, de nuestras emociones, de nuestros deseos; no se podría explicar, por ejemplo, la sensación que sentimos cuando nos pellizcamos un brazo. No podríamos decir tampoco que a María, la científica del anterior ejemplo con visión en blanco y negro que sabe todo sobre los colores desde un punto de vista físico, todavía le falta experimentar o aprender algo sobre ellos.

También muestra la filosofía de la mente que la alternativa entre eliminacionismo o reduccionismo, por un lado, y dualismo cartesiano, por otro (que es uno de los argumentos clásicos para negar la existencia de los estados mentales) es un falso dilema: como hemos visto a propósito de las distintas concepciones sobre la mente, cabe la posibilidad teórica de un materialismo no eliminacionista ni reductivista, una tercera vía que podría presentar los estados mentales como hechos que tienen propiedades peculiares no reductibles que supervienen a, o emergen de, hechos físicos.

Por último, puede decirse que el eliminacionismo deja sin explicar el aparente éxito de la estrategia de atribuir estados mentales a los demás para interpretar su

conducta y coordinarnos con ellos. ¿Por qué la estrategia de la racionalización o de la simulación parecen funcionar bien para explicar o predecir la conducta de los demás, si no existen tales estados mentales? Es cierto que hay ejemplos de teorías falsas que logran explicaciones y predicciones afortunadas; estas deben ser sustituidas por teorías más ajustadas y precisas, cuando se dispone de ellas. Pero no es posible tener teorías eliminacionistas o reductivistas que logren explicar la subjetividad y las experiencias en primera persona, precisamente porque adoptan perspectivas objetivizantes, de tercera persona.

2.2. El argumento epistemológico contra la prueba de los estados mentales

Esta última razón (el éxito de las atribuciones psicológicas como medio de interpretación y predicción de la conducta) sirve también para refutar el argumento epistemológico contra el carácter genuino de la prueba de los estados mentales. Simplemente, no parece cierto que no podamos conocer los estados mentales de los demás, aunque no tengamos un acceso directo a ellos. Más sólida es la versión del argumento que no niega que, en general, podemos hacer atribuciones satisfactorias de estados mentales, pero afirma que estas mismas atribuciones son demasiado débiles para superar criterios de suficiencia probatoria o estándares de prueba jurídicos. Paredes Castañón, por ejemplo, ha puesto de manifiesto las dificultades metodológicas que plantea la prueba del dolo en el proceso penal, que le parecen insuperables hoy en día (PAREDES CASTAÑÓN, 2001). Por ejemplo, a su juicio sería necesario dejar de lado la psicología popular, basada en generalizaciones “de sentido común” (el entrecomillado es suyo) y sustituirla por una psicología de carácter científico. Ello es así, entre otras razones, porque las atribuciones a partir de estas generalizaciones empíricas no son “demostraciones” de los hechos psicológicos y los indicios que suelen ser usados por los tribunales dejan abierta la posibilidad de que el dolo atribuido realmente no se haya dado (o viceversa: que el dolo no atribuido sí se hubiera producido en realidad). Sin embargo, no faltándole razón en sus propuestas constructivas, hay que advertir, por un lado, que la prueba nunca es tan concluyente como para poder considerarse una “demostración absoluta” que excluya totalmente el resto de posibilidades alternativas (GONZÁLEZ LAGIER, 2020) y, por otro lado, que si la prueba de los estados mentales nunca puede superar un estándar de prueba aceptable, entonces se complica también la prueba de cualquier acto humano, porque la prueba de la intención es necesaria (a) para determinar que se trata realmente de una acción, y no de un acto reflejo o involuntario (en el sentido de no dirigido por los estados mentales proposicionales del sujeto) y (b) para identificar qué acción realizó el sujeto (GONZÁLEZ LAGIER, 2013). Nada de esto sería posible si las generalizaciones empíricas “de sentido común” o alguna otra estrategia que sea factible en el proceso judicial no fueran mínimamente fiables.

2.3. El argumento ideológico contra la prueba de los estados mentales

De acuerdo con lo que he llamado el argumento ideológico, la prueba de los estados mentales *reales* no es necesaria o siquiera conveniente en el marco del proceso. Hay varias maneras de sostener esto: se puede negar que la verdad, en general, sea una finalidad relevante de la prueba, sosteniendo alguna concepción no cognoscitivista de la misma. Esta idea ha sido discutida ya en muchas ocasiones, y no insistiré aquí sobre la importancia de una teoría fundamentalmente cognoscitivista de la prueba judicial (TARUFFO, 2011; GASCÓN, 2004; FERRER, 2021). Otra forma de sostener el argumento ideológico es negar la relevancia, en particular, de la verdad sobre los estados mentales. Esto suele hacerse vinculando la finalidad de la sanción penal a un objetivo distinto del reproche o la retribución, como la prevención general. Si la finalidad de la pena —como sugiere Jakobs— es asegurar a la sociedad que la norma violada sigue vigente, o restaurar su vigencia frente al hecho que la lesiona (Jakobs 2001), no puede permitirse el caso de que la sociedad considere que una norma ha sido violada intencionalmente y no se siga la pena pertinente, con independencia de que hubiera intención real o no de violar la norma, de donde se infiere que lo importante no es la intención real del agente (o cualquier otro estado subjetivo), sino cómo va a interpretar la sociedad su conducta (“por lo que se refiere al dolo y a la imprudencia, estos, como hechos puramente psíquicos, carecen de interés a los efectos de la atribución”, escribe en JAKOBS, 1989: 639). En palabras de Laurenzo Copello (refiriéndose a Puppe y a Jakobs): «Lo importante, por consiguiente, de cara a determinar si se produce una contradicción con la norma jurídico penal, no son las actitudes internas del sujeto, sino lo que exprese la conducta desde la perspectiva de la comunicación entre seres inteligentes» (LAURENZO COPELLO, 1999: 225).

Sin embargo, esta manera de pensar está legitimando que se atribuyan consecuencias penales al sujeto al margen de que efectivamente se den en la realidad algunos requisitos (los elementos subjetivos del delito) que exigen las normas penales. Podemos preguntarnos por qué pensamos que tiene sentido atribuir responsabilidad a los seres humanos, y no a los animales. Parece plausible sostener que la dignidad especial que le atribuimos al ser humano, que constituye el fundamento de sus derechos y deberes, descansa —al menos, en buena parte— en la creencia de que este posee ciertas características mentales que lo diferencian de otros seres que, o bien no las tienen, o las tienen en un grado significativamente distinto. Por ejemplo, se ha sugerido por varios filósofos que un presupuesto de la responsabilidad por nuestras acciones es que seamos conscientes de lo que estamos haciendo intencionalmente, que conozcamos nuestras propias creencias, deseos e intenciones (MOYA, 1996). Esta misma dignidad exige tener en cuenta los estados mentales reales de los sujetos a la hora de adscribirles responsabilidad. Por el contrario, la objetivización de los elementos subjetivos del delito (y, por tanto, de su prueba) va en la dirección de una responsabilidad objetiva difícil de justificar en el ámbito penal.

2.4. Entonces... ¿prueba o imputación?

Como hemos visto, una consecuencia de sostener la inexistencia de la mente, la imposibilidad de conocerla o su irrelevancia es la imputación de los estados mentales a partir de criterios normativos, construidos legal o jurisprudencialmente. Se habla por ello de que la prueba de los estados mentales consiste en adscribirlos a partir de ciertas reglas de imputación o presunciones (entendidas como reglas que establecen la obligación de aceptar como probado determinado hecho psicológico si se dan ciertas circunstancias). Pero también podría pensarse que el recurso a estas reglas no es consecuencia de que los estados mentales sean ficciones, sino la causa de que lo sean. Esto es, que el hecho de que la jurisprudencia haya establecido reglas para la valoración de la prueba de estos hechos, facilitando así su determinación, es ya un argumento suficiente para negar el carácter epistemológico y orientado hacia la verdad de estas inferencias. No es que los hechos psicológicos sean hechos normativos y por eso deban probarse de esa manera, es que es el hecho de que se determinen por medio de normas lo que los convierte (para el Derecho) en “hechos normativos” (o calificaciones jurídicas).

No obstante, en mi opinión, que un hecho se pruebe por medio de presunciones (reglas de valoración normativamente guiadas, indicios-tipo o como queramos llamarlas) no nos autoriza a decir sin más que al Derecho no le interesa si en realidad ocurrió o no ni lo convierte en un hecho normativo o institucional. Debemos tener en cuenta que hay tipos distintos de presunciones: Algunas presunciones (que desde la filosofía podrían considerarse las genuinas presunciones) pueden estar respaldadas por la observación de una asociación regular entre hechos (en cuyo caso son similares a máximas de experiencia, pero revestidas de autoridad normativa); otras, vienen justificadas como medios de proteger algún valor o principio que se considera relevante (por ejemplo, el de seguridad, el de protección de los intereses de la parte más débil, etc.). Las primeras, por tanto, tienen un fundamento epistemológico; las segundas, un fundamento normativo. Unas y otras, a su vez, pueden ser derrotables (*iuris tantum*) o no (*iuris et de iure*). Si una presunción no está respaldada por una regularidad empírica y no admite prueba en contrario, podríamos decir que efectivamente está prescindiendo de la realidad y podemos verla como una regla constitutiva que crea una ficción normativa. Pero si se limita a unificar criterios en torno a una regla de la experiencia confirmada o plausible y es derrotable (*iuris tantum*), entonces todavía mantiene un grado importante de “sensibilidad hacia la verdad” o hacia la dimensión empírica de estos hechos.

Las presunciones que la jurisprudencia ha ido introduciendo para la prueba de los hechos psicológicos (a) no cubren todos los supuestos (hay muchos casos en los que es necesario probar el propósito de un agente que no están resueltos por estas reglas y, por tanto, el juez debe valorar los indicios libremente), (b) por lo general, son derrotables y (c) en mi opinión, se basan en máximas de experiencia, en razonamientos

que en última instancia tienen una justificación empírica. Esto último explica que todas estas reglas nos parezcan razonables —no arbitrarias— desde el punto de vista de su eficacia para descubrir (por supuesto, dentro de un margen de error) los estados mentales que realmente tenía el sujeto en los casos a los que se aplican (cosa que no tendría por qué suceder si el objetivo fuera una imputación dirigida a satisfacer otros fines). Así ocurre, por ejemplo, con los criterios establecidos para la distinción entre el *animus necandi* y el *animus laedendi* (“entre otros: a) los antecedentes de hecho y las relaciones entre el autor y la víctima; b) la clase de arma utilizada; c) la zona o zonas del cuerpo a que se dirige la agresión; d) el número de golpes inferidos; e) las palabras que acompañaron al ataque; f) las condiciones de lugar, tiempo y circunstancias conexas o concomitantes con la acción; g) la causa o motivación de la misma; h) la entidad y gravedad de las heridas causadas” STS, 2.^a, de 31-1-2011); o con los criterios para dar por probado el requisito, en el delito de receptación, del conocimiento del origen delictivo del bien comprado (la irregularidad de las circunstancias de la compra o modo de adquisición; la mediación de un precio vil o ínfimo, desproporcionado con el valor real de los objetos adquiridos; la clandestinidad de la adquisición; la inverosimilitud de las explicaciones aportadas para justificar la tenencia de los bienes sustraídos; la personalidad del adquirente acusado y de los vendedores o transmitentes de los bienes; la adquisición fuera de los cauces ordinarios del comercio. STS, 2.^a, de 29-1-2000); o, por poner un último ejemplo, con el criterio (entre muchos otros) de la posesión de una cantidad de droga notablemente superior a la que un consumidor emplea en 5 días para la atribución de la intención de traficar con sustancias tóxicas (entre otras, STS de 5-6-1997). En todos estos casos, las atribuciones se basan en máximas de experiencia y generalizaciones empíricas, y no en criterios sin fundamento epistemológico.

Todo esto no quiere decir, no obstante, que no haya nada normativo en la prueba de los estados mentales. El Derecho usa o presupone necesariamente conceptos (también conceptos relativos a hechos psíquicos: el concepto de intención, de conocimiento, de emoción, etc.) que pueden tener un componente normativo importante. Estos conceptos tienen una función clasificatoria: tratan de “poner orden” en una realidad (en este caso, psicológica) que, antes de ellos, no está completamente estructurada. Al construir estos conceptos, los juristas deben tener en cuenta dos tipos de criterios de corrección: (1) deben tener en cuenta que la realidad objetiva tiene en sí misma algún grado de organización o estructura que los conceptos deben tratar de reflejar y (2) deben diseñarse de manera que satisfagan en la mayor medida posible su finalidad. Al fin y al cabo, los conceptos son en buena medida herramientas, y su finalidad, en el caso del Derecho, es permitir una aplicación adecuada de las normas y la adscripción correcta de responsabilidad de acuerdo con las mismas. Esto hace que los conceptos de acción, intención, causalidad, etc. que usan los juristas tengan un fuerte componente normativo en el momento en que se construyen. Ahora bien, una vez fijados estos conceptos, la comprobación de si ha ocurrido en la realidad un hecho subsumible en ese concepto no tiene por qué ser una ficción, sino que ya requiere “una mirada” a la realidad empírica (GONZÁLEZ LAGIER, 2007).

2.5. La prueba de los estados mentales con contenido representacional

¿Qué argumentos y qué criterios pueden usarse en el Derecho para la prueba de los hechos psíquicos? Como hemos visto, la filosofía de la mente (con su descripción de la estrategia de la racionalización, de las teorías basadas en “reglas de psicología popular” y la estrategia de la empatía/imaginación) también puede ofrecer algunas orientaciones generales útiles para responder a esta pregunta. Sin embargo, hay que tener en cuenta una diferencia importante entre los contextos que estudia la filosofía de la mente y el que nos interesa como juristas: Las teorías de atribución psicológica que hemos visto tratan de describir cómo atribuimos estados mentales, pero no están preocupadas por el problema de la justificación de esas atribuciones, esto es, por dar razones aceptables intersubjetivamente que muestren que esas atribuciones son correctas. La prueba judicial, por el contrario, requiere la motivación o justificación de la atribución de los hechos psicológicos, lo que, como sabemos, remite a un tipo especial de razones. Esto descarta el valor probatorio directo de las estrategias de atribución en primera persona (la teoría de la simulación), por un lado porque se conciben como estrategias no inferenciales y, por otro lado, porque el hecho de que el juez piense que tendría uno u otro estado mental en la situación en la que se encuentra el sujeto al que atribuye ese mismo estado mental no es, por sí sola, una razón justificativa (salvo que se “objetive” en alguna regla de experiencia compartida y generalizada). No obstante, que carezca de fuerza justificativa no le resta importancia en el “contexto de descubrimiento”, esto es, en el proceso de formación de la hipótesis del juez.

Teniendo esto en cuenta, puede avanzarse un poco más en la prueba de los estados mentales si se parte de la distinción entre las actitudes proposicionales y los estados mentales de carácter fenomenológico. Comencemos por los primeros. En este caso es plenamente aplicable la estrategia racionalizadora.

Hay una estrecha conexión entre lo que llamamos una acción intencional y lo que podemos llamar racionalidad estratégica o racionalidad de medio a fin. Actuar intencionalmente es actuar con un propósito, persiguiendo un fin. Pero solo podemos decir seriamente que perseguimos un fin si sabemos cómo obtenerlo, o si confiamos en llegar a saberlo; esto es, si sabemos qué hemos de hacer para lograr ese fin. Esto establece una fuerte conexión conceptual entre (1) las creencias acerca del mundo y de las relaciones de medio a fin, (2) las intenciones y deseos y (3) las acciones. Esta relación se puede expresar con lo que se ha llamado el Principio de Racionalidad Mínima (PRM) (GONZÁLEZ LAGIER, 2003 y 2004). El PRM puede enunciarse de la siguiente manera: «Si un agente actúa intencionalmente, tiende a realizar la acción que en ese momento cree más adecuada para lograr el fin que persigue».

El PRM es un principio emparentado con el principio de racionalidad o al de caridad (puede verse como una consecuencia del mismo), que hemos visto antes

como presupuesto de lo que hemos llamado la estrategia racionalizadora de atribución psicológica, pero en lugar de referirse a una racionalidad objetiva, pone el acento en lo que el sujeto cree, a la luz de su evaluación del contexto en un momento determinado, que es lo más racional o adecuado. Se trata, por tanto, de una racionalidad mínima o imperfecta, que no tiene por qué coincidir con lo que el sujeto consideraría racional en una situación “menos tensa” (es decir, sin urgencia por decidir) o con lo que sería “objetivamente racional”. El PRM tiene, en mi opinión, un fundamento, por un lado, empírico y, por otro, conceptual. Tiene un fundamento empírico porque este principio, o alguno semejante, es el que parece estar detrás de nuestras atribuciones de actitudes proposicionales que, como ya hemos visto, parecen ser en buena medida exitosas. Tiene también un fundamento conceptual en la medida en que hemos construido la noción de intención de manera que si un sujeto no elige en un momento determinado el medio que le parece más adecuado para lograr un fin concreto, no diríamos que tiene la intención de conseguir ese fin (no es infrecuente que una correlación empírica fundamente un concepto: por ejemplo, la constatación de que el calor dilata los metales fundamenta una definición de metal que incluye la propiedad de ser dilatado por el calor).

Si se acepta el PRM, entonces la prueba de la intención podría verse como una inferencia del siguiente tipo:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C
- (3) S hizo A en las circunstancias C

-
- (4) S hizo A con la intención de conseguir F

Veamos un ejemplo: Alfredo y Herminio viven en huertas colindantes. Una tarde, Alfredo llama a grandes voces a Herminio, con el que anteriormente había mantenido múltiples y acaloradas discusiones. Al asomarse este a la puerta de su casa entablan una fuerte discusión sobre la propiedad de una vereda que separa ambas fincas. La discusión es interrumpida por el yerno de Herminio, quien asiéndole del brazo lo mete en la casa. A continuación, Alfredo entra en su propia casa, coge una escopeta IS, modelo PR, del calibre 12 y vuelve a salir, retando a Herminio a gritos; le sigue su hermana Arsenia, quien, agarrándole del brazo, trata de disuadirle. Cuando Herminio se asoma de nuevo a la puerta, Alfredo le dispara un cartucho de perdigones desde una distancia de unos 12 metros. En el momento en que Alfredo dispara, Arsenia tira de él hacia atrás para desviar la trayectoria del disparo. Los perdigones impactan en un muro de piedra que está detrás de Herminio (a un metro de altura), sin que este resulte lesionado. Se calcula que la trayectoria del disparo se ha desviado 70 cm. respecto de la que hubiera matado o herido a Herminio. Estos hechos son calificados por el tribunal que los juzga como un delito de homicidio en grado de tentativa, lo

que significa que se da por probado (a) que la acción de Alfredo es intencional y (b) que su intención es la de matar a Herminio y no meramente lesionarle (en cuyo caso se le podría haber condenado de un delito de lesiones en grado de tentativa).

¿Cómo sabemos que Alfredo disparó con la intención de matar a Herminio, y no meramente de lesionarle? La respuesta es que, de acuerdo con el Principio de Racionalidad Mínima, los sujetos procuran realizar la acción que, a su juicio, tiende a asegurarles el resultado que pretenden y la acción de Alfredo fue instrumentalmente adecuada para matar a Herminio (como sabemos por el calibre de la escopeta usada, el lugar del cuerpo de Herminio hacia donde apuntó, la distancia, etc.).

El razonamiento sería el siguiente:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
 - (2) Alfredo creía que la manera más adecuada de matar a Herminio consistía en disparar en las circunstancias C
 - (3) Alfredo disparó en las circunstancias C
-
- (4) Alfredo disparó con el fin de matar a Herminio (esto es, tenía la intención de matar a Herminio).

Como puede apreciarse, una de las premisas que necesitamos para inferir la intención hace referencia a otro estado mental, las creencias del agente respecto de ciertas relaciones de medio a fin. Como ya vimos a propósito del principio de caridad, la estrategia racionalizadora requiere asumir que el agente al que atribuimos estados mentales tiene aproximadamente creencias razonables acerca de las relaciones de medio a fin. Esto puede justificarse a partir de reglas basadas en generalizaciones empíricas, como las siguientes, propuestas (entre otras) por Ramón Ragués (RAGUÉS, 1999: 521 y ss.):

a) La consideración de una persona como imputable lleva a atribuir a un sujeto todos aquellos conocimientos cuya ausencia solo se entiende posible en quien padece algún tipo de perturbación psíquica o sensorial, o en menores de edad.

b) El hecho de que una persona haya sido normalmente socializada hace que se le puedan atribuir todos aquellos conocimientos cuya ausencia solo se concibe en sujetos que no han tenido contacto con la civilización de la que se trate.

c) La circunstancia de que en un sujeto concurren determinadas características personales o de que ocupe una determinada posición social lleva a imputarle los conocimientos cuya ausencia haría impensable que tuviera esas características o que ocupara tal situación.

Por otra parte, la coherencia que la estrategia de la racionalización exige entre creencias, intenciones y acciones, tiene algunas consecuencias importantes para la prueba.

En primer lugar, también es posible inferir las creencias de un sujeto acerca de los medios para lograr un fin a partir del PRM si conocemos su intención y la acción realizada:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S hizo A en las circunstancias C
- (3) S hizo A con la intención de conseguir F

-
- (4) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C

Igualmente, es posible inferir (o predecir) que el sujeto realizó (realizará) cierta acción con el siguiente argumento:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C
- (3) S tenía la intención de conseguir F

-
- (4) S hizo A en las circunstancias C

En definitiva, si el agente es racional, podemos asumir las siguientes reglas de inferencia entre sus acciones, sus intenciones y sus creencias:

- a) Si conocemos su intención y sus creencias (acerca de las relaciones medio-fin adecuadas para satisfacer su intención) podemos inferir (explicar, predecir) su acción.
- b) Si conocemos la acción realizada y sus creencias (acerca de las relaciones medio-fin) podemos inferir su intención⁵.
- c) Si conocemos la acción realizada y su intención, podemos inferir sus creencias instrumentales (de medio a fin).

Igualmente, podemos usar las siguientes reglas (también basadas en criterios de racionalidad) como criterios de atribución (*prima facie*):

- a) Si un agente tiene la intención de hacer A, y cree que B es incompatible con A, no tiene la intención de hacer B (Principio de no contradicción de las intenciones).

⁵ Esto quizá arroja cierta luz sobre la discusión en la doctrina penal acerca de si el dolo es conocimiento o voluntad: no se trata en realidad de una alternativa, porque es conceptualmente imposible saber o ser consciente de que la acción x reúne los elementos del tipo penal, realizar esa acción y no tener la intención (voluntad) de dar lugar a una acción que reúne los elementos del tipo penal.

b) Si el agente cree que el resultado B se sigue necesariamente de la acción A, no puede tener la intención de A y no tener la intención de B (Principio de transmisión de la intención a las consecuencias necesarias o previsibles).

c) Si el agente tiene la intención de hacer B y cree que A es el único medio para dar lugar a B, tiene también la intención de hacer A (Principio de transmisión de la intención a los requisitos causales o convencionales).

2.6. La prueba de los estados mentales de carácter fenomenológico

En resumen, para la prueba de las actitudes proposicionales disponemos de una combinación de la estrategia racionalizadora y (especialmente para las creencias) de máximas de experiencia de sentido común (lo que los filósofos de la mente han llamado “teoría de la teoría” o “psicología popular”) que sirven como criterio de atribución psicológica. En la medida en que estas estrategias coinciden con las que usamos habitualmente para inferir los estados mentales de los demás, y en la medida en que tienen un éxito que podemos medir en nuestra capacidad de comunicarnos y entendernos (éxito cuya mejor explicación es que a menudo *acertamos* cuando atribuimos estados mentales, esto es, descubrimos una realidad), no se trata de criterios para una imputación en sentido estricto o fuerte (no sensible a la verdad). Dicho de otra manera, aunque se hable de “atribución” de estados mentales, los criterios tienen un fundamento empírico, están orientados a tratar de descubrir una realidad (psicológica) y son derrotables por una prueba en contrario de más peso. Exactamente como ocurre con la prueba de los hechos físicos.

Pero ¿qué ocurre con el resto de hechos psicológicos, con aquellos que carecen de un contenido proposicional, al menos en un sentido evidente? ¿Cómo se prueba el dolor, la somnolencia, la ansiedad, etc.? Es evidente que, en estos casos, no es aplicable la estrategia de la racionalización, dado que esta requiere estados mentales con contenido proposicional, y las sensaciones puras no lo tienen (otro caso, como veremos, es el de las emociones). Para este tipo de hechos, parece más adecuada la teoría de la empatía o simulación, de acuerdo con la cual el intérprete se imagina en el lugar del interpretado y le atribuye los estados mentales que cree que él tendría en esa situación o que él tendría si fuera el otro. No obstante, como ya he señalado, esta estrategia de atribución psicológica no permite obtener razones justificativas que sean útiles para motivar las decisiones sobre la prueba de este tipo de hechos, salvo que, de alguna manera, esas “intuiciones” del juez puedan respaldarse por medio de reglas de experiencia que correlacionen ciertos indicios con los estados mentales correspondientes. Estas reglas de experiencia pueden referirse a la relación entre ciertos hechos, que asumimos como causas, y los estados fenomenológicos, que asumimos como efectos (como, por ejemplo, cuando inferimos que alguien siente dolor porque hemos visto cómo se golpea un dedo con un martillo), o a la relación entre los estados

mentales, que tomaríamos ahora como causas, y la conducta subsiguiente, que sería el efecto (como cuando inferimos que a alguien le duele el dedo porque vemos que se lo agarra fuertemente y oímos sus quejidos). Dicho de otra manera, el juez puede apelar (como, de hecho, hace) a la idea de que cualquier persona razonable tendría tales sensaciones en tales circunstancias, o actuaría de tal manera si tuviera esos estados mentales. Las apelaciones a lo que sentiría o haría una “persona razonable” o “persona media” pueden verse como la objetivización de la estrategia de la empatía (de manera que la “teoría de la simulación” acaba reconduciéndose a la “teoría de la teoría”). Pero —de nuevo— esta no es una estrategia para una mera imputación ciega a lo que de hecho ocurriera realmente, si las generalizaciones usadas tienen un buen respaldo inductivo.

2.7. La prueba de los estados mentales mixtos (las emociones)

Las emociones son un estado mental del que puede decirse que tiene simultáneamente contenido representacional y también un importante aspecto fenomenológico. En cierto sentido, las emociones pueden verse como una combinación de creencias (o algún estado mental cognitivo) y sensaciones (GONZÁLEZ LAGIER, 2009: Capítulo II). No siempre se ha visto de esa manera, sino que por influencia de la tradición cartesiana se han identificado solo con el aspecto sensitivo: las emociones, pensaban autores como Descartes o William James, son percepciones causadas por determinados cambios fisiológicos que experimentamos en ciertas circunstancias (GONZÁLEZ LAGIER, 2009: Capítulo I). Para esta concepción, las emociones quedan fuera del ámbito de la racionalidad (más bien, la impiden). Pero hoy en día ha habido una fuerte rehabilitación de la noción aristotélica de “emoción”, que presenta a las emociones como un tipo de juicio sobre el mundo, seguido de ciertos cambios fisiológicos, una sensación (la percepción de los mismos) y una tendencia a la acción. Un contexto típico en el que surge una emoción sería el siguiente: Los sujetos tienen un trasfondo de deseos, objetivos, proyectos, planes,... Ciertos acontecimientos pueden enjuiciarse como positivos (facilitadores) o negativos (obstáculos) en relación con tales objetivos; si el sujeto juzga un acontecimiento como un obstaculizador de su acción, se desencadena cierta sensación (en este caso, negativa, esto es, no placentera) y una tendencia a actuar de manera que se pueda sortear o resolver el obstáculo; por el contrario, si el acontecimiento es juzgado como algo que facilita el objetivo, la sensación será positiva (placentera) y se desencadenará una tendencia a actuar de manera que se aproveche esa oportunidad (GONZÁLEZ LAGIER, 2009: Capítulo II). Como puede apreciarse, en las emociones se vinculan creencias (los juicios evaluativos), deseos (los objetivos), sensaciones y tendencias a la acción. Esta conexión con deseos, creencias y acciones hace que las emociones mantengan una cierta relación con la racionalidad. Las emociones son racionales si están basadas en creencias o juicios de valor correctos o justificados, y son irracionales si no lo están. Igualmente,

las emociones son racionales si generan tendencias a la acción instrumentalmente adecuadas para resolver el obstáculo o para aprovechar la oportunidad que representa el acontecimiento que las dispara (sobre la relación entre racionalidad y emociones véase GONZÁLEZ LAGIER, 2009: Capítulo IV).

Lo anterior sugiere que la estrategia de la racionalización es aplicable también para la atribución de emociones. A partir de los objetivos de un agente, de sus creencias evaluativas y de su acción, podemos inferir las emociones que probablemente tuvo en determinadas circunstancias. De nuevo, resulta fundamental aquí la coherencia entre la acción y los estados mentales, hasta el punto de que la prueba de las emociones, como la de las intenciones o la de las creencias, puede verse como la búsqueda de una explicación coherente de la conducta del agente en la que sus estados mentales y su comportamiento se ajustan uno a otro y se explican mutuamente.

Pero, como hemos visto, la emoción tiene un componente, la sensación, que tomado aisladamente no tiene esta misma conexión con la racionalidad. En ocasiones, es este componente el que necesita ser probado, por ejemplo, cuando nos preguntamos si la emoción bajo la cual ha actuado el agente ha sido suficientemente intensa como para considerar que ha “nublado” su juicio y, por tanto, podría constituir un caso de atenuante de la culpabilidad. Para estos supuestos, la prueba ha de basarse en máximas de experiencia acerca de qué intensidad suelen tener las emociones generadas en unas u otras circunstancias, en suposiciones acerca de cómo un sujeto razonable se sentiría en tales circunstancias.

3. CONCLUSIONES

En las páginas anteriores he intentado ofrecer un recorrido por algunos de los temas de la filosofía de la mente que más relación tienen con el problema de la prueba de los estados mentales, especialmente aquellos que son útiles para entender qué son y para replantear la discusión acerca de si son “hechos” que deban probarse o el resultado de juicios normativos de imputación. Esto es, si forman parte de la *quaestio facti* o de la *quaestio iuris* de la decisión judicial. Esta discusión no me parece trivial, pues de ella dependen cuestiones como si la determinación de los estados mentales está sujeta a garantías procesales como la presunción de inocencia, si es posible o no (o por qué vía) el recurso ante tribunales superiores por una determinación incorrecta de los mismos o cuáles son los criterios adecuados para su atribución (si los estados mentales son hechos, el criterio de corrección de tales atribuciones será la verdad o falsedad de las mismas, pero si no son hechos, dependerá de otros objetivos que no tendrán un carácter epistemológico). En mi opinión, como hemos visto, la filosofía de la mente aporta argumentos sólidos para aceptar una concepción cognoscitivistica de la prueba de los estados mentales: Frente a los argumentos ontológicos que la rechazan es posible poner de manifiesto (a) que negar la existencia de estados mentales deja sin explicar un importante aspecto de la realidad y (b) que entre el materialismo eliminacionista y el dualismo cartesiano hay alternativas plausibles. Los argumentos

epistemológicos que aducen la imposibilidad de conocer los estados mentales de los demás con suficientes garantías, por su parte, parecen olvidar que los seres humanos poseen la habilidad de adscribir estados mentales de una manera aparentemente exitosa. Por último (aunque esto ya queda fuera de la filosofía de la mente), frente al argumento de que los fenómenos psicológicos no son jurídicamente relevantes, cabe oponer los problemas de una objetivización de la responsabilidad penal para una aplicación justa del Derecho.

Es importante también advertir que, como hemos visto, pese a esta discusión, los criterios que, *de hecho*, usan los jueces para atribuir estados mentales —al igual que los que propone la doctrina o las presunciones que establece el legislador o la jurisprudencia— normalmente resultan razonables si los estados mentales se consideran “hechos” y su determinación se considera “prueba” en el sentido habitual; esto es, se trata de criterios epistemológicamente razonables si nuestro objetivo es que las atribuciones psicológicas sean verdaderas.

Los filósofos de la mente han propuesto varias teorías para explicar cómo funciona la habilidad de las personas de hacer atribuciones psicológicas: las principales son la “teoría de la teoría” (recurso a generalizaciones o leyes), la estrategia de la racionalización (asumir la racionalidad de los agentes, de manera que la coherencia entre sus estados mentales y entre estos y su conducta permite realizar inferencias en ambas direcciones) y la “teoría de la simulación” (que pone de manifiesto la importancia de la empatía como recurso para la atribución psicológica). En realidad, más que como teorías enfrentadas parece más útil verlas como estrategias complementarias, con mayor o menor adecuación según el tipo de estado mental que se pretende conocer. No creo que estas teorías aporten criterios nuevos a los ya usados en el Derecho, pero tener en cuenta estas estrategias puede servir para darles un mayor fundamento y, quizá también, para entenderlos mejor y hacerlos más precisos.

BIBLIOGRAFÍA

- BRUNSTEINS, P. (2008), “Algunas reflexiones acerca de la simulación mental y la perspectiva de la primera persona”, en *Arete*, v. 20, n.1, Lima.
- (2011), “El rol de la empatía en la atribución mental”, en *Revista Argentina de Ciencias del Comportamiento*, vol. 3, núm.1, Córdoba.
- CASTRO, L.G. (2017) “¿Está presente el problema de las otras mentes en la filosofía de R. Descartes?”, en *Ingenium*, vol. 11, Madrid.
- CHURCHLAND, P. (1999), *Materia y conciencia*, Barcelona, Ed. Gedisa.
- DANCY, J. (1993), *Introducción a la epistemología contemporánea*, Madrid, Ed. Tecnos.
- DAVIDSON, D. (2003) *Subjetivo, intersubjetivo, objetivo*, Madrid, Ed. Cátedra.
- FERRER BELTRÁN, J. (2021), *Prueba sin convicción*, Madrid, Ed. Marcial Pons.
- GARCÍA CARPINTERO, M. (1995), “El funcionalismo”, en Fernando Broncano (ed.), *La mente humana*, Enciclopedia Iberoamericana de Filosofía, Madrid, ed. Trotta – Consejo Superior de Investigaciones Científicas.
- GASCÓN, M. (2004), *Los hechos en el Derecho. Bases argumentales de la prueba*, Madrid, Ed. Marcial Pons.

- GOLDMAN, A. (1989), "Interpretation Psychologized", en *Mind & Language*, vol. 4, núm. 3
- GONZÁLEZ LAGIER, D. (2003), "Buenas razones, malas intenciones (sobre la atribución de intenciones)", en *Doxa. Cuadernos de Filosofía del Derecho*, núm. 26, Alicante.
- (2004), "La prueba de la intención y el Principio de Racionalidad Mínima", en *Jueces para la democracia*, núm. 50, Madrid.
- (2006), "La prueba de la intención y la explicación de la acción", en *Isegoría*, núm. 35, Madrid.
- (2007), "Hechos y conceptos", en *Cuadernos Electrónicos de Filosofía del Derecho*, núm. 15.
- (2009), *Emociones, responsabilidad y Derecho*, Madrid, Ed. Marcial Pons.
- GONZÁLEZ LAGIER, D. (2013), *Las paradojas de la acción*, Madrid, Ed. Marcial Pons.
- (2020), "¿Es posible formular un estándar de prueba preciso y objetivo? Algunas dudas desde un enfoque argumentativo de la prueba", en *Revista Telemática de Filosofía del Derecho*, núm. 23.
- HAI DT, J. (2012), «El perro emocional y su cola racional: Un enfoque intuicionista social del juicio moral, juicio moral», en A. Cortina (Ed.), *Guía Comares de Neurofilosofía práctica*, Granada, Ed. Comares.
- HAMPSHIRE, Stuart; HART, H.L.A. (1958), "Decision, Intencion and Certainty", en *Mind*, vol. LXVII, núm. 265, Oxford.
- IACOBONI, M.: (2013), *Las neuronas espejo. Empatía, neuropolítica, autismo, imitación o de cómo entendemos a los otros*, Ed. Katz
- JAKOBS, G. (1989), "Sobre la función de la parte subjetiva del delito en Derecho penal", en *Anuario de derecho penal y ciencias penales*, Tomo 42, Fasc/Mes 2, Madrid.
- (2001) "¿Qué protege el Derecho penal: bienes jurídicos o la vigencia de la norma?", en *Cuadernos de Doctrina y Jurisprudencia Penal*, vol. 7, núm. 11.
- KAHNEMAN, D. (2012), *Pensar rápido, pensar despacio*, Barcelona, Ed. Mondadori.
- KALPOKAS, D. (2019), "Algunas reflexiones en torno al conocimiento de las otras mentes", en Carolina Scotto, Fernando G. Rodríguez y Irene Audisio (compiladores), *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje*, Buenos Aires, Ed. Teseo.
- LAURENZO COPELLO, P. (1999), *Dolo y conocimiento*, Valencia, Ed. Tirant lo Blanch.
- MILL, J. S. (1867): *An Examination of Sir William Hamilton's Philosophy*, 3a ed., Londres, Longmans.
- MOYA, C. (1996), "El sujeto enunciado", en Manuel Cruz, *Tiempo de subjetividad*, Barcelona, Ed. Paidós.
- (2006), *Filosofía de la mente*, Universidad de Valencia.
- PAREDES CASTAÑÓN, J.M. (2001), "Problemas metodológicos en la prueba del dolo", en *Anuario de filosofía del derecho*, núm. 18.
- QUINTANILLA, P. (2004), "Comprender al otro es crear un espacio compartido. Caridad, empatía y triangulación", en *Ideas y valores*, núm. 125, Bogotá.
- (2019), *La comprensión del otro. Explicación, interpretación y racionalidad*, Lima, Pontificia Universidad Católica del Perú.
- RAGUÉS I VALLÉS, R. (1999), *El dolo y su prueba en el proceso penal*, Barcelona, J.M. Bosch Editor.
- (2002), "Consideraciones sobre la prueba del dolo", en *La Ley*, Año XXIII, núm. 5633, Madrid.
- RIZZOLATI, G; CRAIGHERO, L. (2004), "The Mirror Neuron-System", en *Annuary Review of Neuroscience*, 27.
- SEARLE, J. (1990), *Mentes, cerebros y ciencia*, Madrid, ed. Cátedra.
- (1992), *Intencionalidad*, Madrid, Ed. Tecnos.
- (1996), *El redescubrimiento de la mente*, Barcelona, Ed. Crítica.
- (2006), *La mente. Una breve introducción*, Barcelona, Ed. Norma.
- SKIDELSKY, L. (2011), "Autoconocimiento y Atribución de Estados Mentales en Teoría de la Mente", en *Revista Argentina de Ciencias del Comportamiento*, vol. 3, núm. 1, Córdoba.
- TARUFFO, M. (2011), *La prueba de los hechos*, Madrid, Ed. Trotta.
- WOLLHEIM, R. (2006), *Sobre las emociones*, Madrid, Ed. Antonio Machado Libros.

