

Evaluación, transparencia y democracia

Eulalia Pérez Sedeño (eulalia.perez@fecyt.es)
Departamento de Ciencia, Tecnología y Sociedad
Instituto de Filosofía, CSIC, España

173

La responsabilidad individual del científico ha decrecido, en parte, por la organización de la ciencia del siglo veinte. El carácter social de la ciencia y la responsabilidad que ello trajo aparejado pueden ser leídos a través del análisis del sistema de arbitraje mediante pares. En este artículo la autora indaga en la articulación de dicho sistema para develar sus resortes fundamentales y plantear alternativas de evaluación más abiertas.

Palabras clave: ciencia, sistemas de arbitraje, evaluación por pares, democracia.

Scientists' individual responsibility has decreased, partly, because of the 20th century science organization. The social character of science and the responsibility that this has entailed can be read through an analysis of the peer review system. In this paper the author makes a critical review of this system to unveil its main features and to propose opener evaluation alternatives.

Keywords: science, evaluation systems, peer review, democracy.

Desde el final de la Segunda Guerra Mundial, evaluar la importancia de la investigación científica, y lo que se invierte en ella, ha sido una cuestión fundamental para los políticos (aunque tal preocupación llegara décadas más tarde a España en particular). Lo cierto es que, dado que la seguridad militar ya no es la razón fundamental para establecer prioridades, ha habido una cierta exigencia de un nuevo contrato entre la ciencia y la sociedad que establezca una relación de trabajo más estrecha entre la academia, la industria y los laboratorios nacionales, y que cree un entorno favorable a la comercialización de tecnologías innovadoras que se generen a través de los fondos públicos de investigación.

Diversos adelantos tecnológicos se lograron gracias a un cambio radical que se había producido en la forma de hacer ciencia y que la había convertido en una organización de numerosos científicos que contribuían con diversos grados de pericia (*expertise*) en diversos campos a un proyecto común. El modelo era el Proyecto Manhattan para física, en el que físicos, ingenieros, matemáticos, etc. colaboraban para un mismo fin, y que sirvió de incubadora para la electrónica, la ciencia computacional y las tecnologías militares y aeroespaciales desarrolladas durante la guerra fría (Kotz, 1995; Lenoir y Hays, 2000). En el proyecto Manhattan también había un proyecto para biomedicina, pues los principales oficiales médicos comenzaron a planificar cómo adaptar al mundo posbélico el trabajo que habían hecho bajo condiciones de extrema seguridad. No sólo contemplaron cómo continuar las prometedoras investigaciones que habían iniciado durante la guerra, sino que querían llevarlas más allá de los confines militares y adaptarlas al mundo civil, estableciendo disciplinas médicas y académicas en física médica, biofísica y medicina nuclear; también tenían programas que iniciaron para entrenar a personal médico en el uso de materiales radioactivos, así como un programa estratégico para subvencionar el desarrollo de instrumentación biomédica, radio-farmacéutica y con radio-isótopos como parte de un esfuerzo por crear la infraestructura de una industria nuclear biomédica auto-sostenible. Estos programas afectaban, pues, a la biología y a la medicina que se ocupaban, entre otras cosas, del diagnóstico y control de los efectos producidos por la exposición a radiaciones de los materiales utilizados en los experimentos y operaciones, así como de la toxicidad química, y que puso a trabajar juntos a biólogos, radiólogos, médicos, físicos nucleares etc.).¹ El nuevo proyecto Manhattan, en potencia, para la biología es la iniciativa del Genoma Humano que ha servido de modelo para el nuevo contrato entre la ciencia y la sociedad, por las posibilidades de estimular el trabajo en colaboración entre físicos, ingenieros, científicos computacionales y genéticos en los laboratorios, de modo que sus resultados sean comercialmente viables (Lenoir y Hays, 2000).

174

¹ Pero los estudios fisiológicos y los primeros atisbos de que podía ser efectivo un programa de investigación médico utilizando isótopos radioactivos selectivamente localizadores para destruir las células cancerígenas son anteriores al proyecto Manhattan. Desde 1923, por lo menos, se estaba experimentando con torio-B, agua pesada y otra serie de isótopos radioactivos, primero en Manchester, pero luego en Berkeley, y en el MIT. Sin embargo, lo más importante es que hacia 1943, cuando el proyecto Manhattan estaba en marcha, ya había una serie de éxitos importantes y varias líneas bien demarcadas de investigación y de terapias, abiertas por unos pocos equipos cruciales de biofísicos y médicos nucleares.

Pero el nacimiento de la BS ha provocado otros cambios que afectan a qué conocimiento se produce y cómo, el contexto en que se prosigue, su forma de organización, el sistema de recompensas que usa y los mecanismos que lo controlan. Todas estas son características sociales que están bien articuladas en las ciencias paradigmáticas (física, química y biología) aunque algo menos en las ciencias sociales y las humanidades. El paso a esta nueva forma de producción del conocimiento se caracteriza por diversos aspectos. En lo que J. Ziman (2000) ha denominado la ciencia académica o la también denominada ciencia en Modo 1 (Gibbons et al, 1994) los problemas se plantean dentro de la estructura disciplinar, mientras que en la postacadémica es en el contexto de aplicación, que exige cada vez más estructuras transdisciplinares (y que provoca que constantemente surjan nuevas disciplinas). En la ciencia académica o, lo que es igual, en el Modo 1, hay homogeneidad de intereses, instituciones (principalmente universidades u organismos estatales) y actividades, pero en la postacadémica hay heterogeneidad de intereses, de instituciones participantes (empresas, etc.) y de actividades. Si en la ciencia académica la estructura es jerárquica y autoritaria, en la postacadémica es más abierta heterogénea y transitoria, con mayor interacción entre múltiples actores, que tienen mayor responsabilidad social, frente a la responsabilidad individual típica de la ciencia académica. Esto conduce a la última característica diferenciadora, aunque no sea menos importante: en el Modo 1 es la propia comunidad científica la que evalúa los resultados, mientras que en el Modo 2 hay una mayor gama de mecanismos de control en los cuales tienen cabida otros intereses, valores, etcétera.

La responsabilidad individual del científico se ha visto diluida, en parte, por la organización de la ciencia del siglo XX. En dicha organización hay que distinguir, al menos, entre tres funciones separadas: la de producir conocimientos, dar un opinión experta y tomar decisiones. El papel del científico es avanzar en el estado del conocimiento; el del experto es hacer ver o aclarar las cuestiones proporcionando la información relevante y estar preparado para entrar en el dominio de la controversia; finalmente, las personas que toman las decisiones políticas tienen que sopesar la información y tomar decisiones, aceptando el control público y la posibilidad de ser relevado de su cargo, por lo menos en los regímenes democráticos. Aunque podemos separar conceptualmente estas tres funciones, a la vez que en ellas podemos determinar áreas diferentes de debate y distintos estilos de intervención, se solapan de diversos modos: los científicos son expertos a la vez que, en muchas ocasiones, tienen puestos políticos de toma de decisiones y a la inversa; y lo es en la medida en que lo que dice el experto ya no es una cuestión de objetividad científica sino que ahí entran en conflicto intereses, valores y sentimientos muy fuertes.

175

El carácter social de la ciencia y de la responsabilidad se evidencia perfectamente en el sistema que certifica el conocimiento: el sistema de arbitraje mediante iguales, pares o colegas. Dicho sistema de evaluación recibe cada vez más atención y críticas, debido, sobre todo, a determinados casos en que parece fracasar por la intervención de valores que supuestamente no deberían estar presentes y que suponen el “fracaso” del sistema.

Se entiende por “evaluación por pares, iguales o colegas” el principio de que la

investigación debe ser evaluada por personas que se rigen por la confianza y respeto mutuo y que son reconocidas socialmente como expertas en un campo determinado de conocimiento. Este tipo de evaluación se ha utilizado para lograr distintos fines asociados a la producción de conocimiento, fundamentalmente para determinar qué investigaciones se financian y cuáles no, y qué se publica como resultado de la investigación.

Este tipo de evaluación por pares es compleja y tiene múltiples funciones que a veces se entremezclan entre sí: por ejemplo, estandariza la conducta investigadora a la vez que ejerce el control de calidad sobre dicha investigación; refuerza la conformidad ideológica entre los investigadores y establece quién entra en la categoría de investigador; por otro lado, recompensa la investigación realizada, influye en la dirección de las investigaciones futuras, asume la responsabilidad colectiva de los resultados investigadores y maximiza su impacto.

Los pares son, en principio, personas que tratan las ideas u opiniones de los otros con el mismo respeto que tratan las propias. Pero además, son personas que pueden funcionar recíprocamente, bien como evaluadores o como evaluados. Estos pares pueden ser de dos tipos: absolutos o relativos. El par absoluto es una persona que trabaja en el mismo campo que quien envía una propuesta de financiación o un artículo para ser publicado. En cambio, el par relativo no trabaja en el mismo campo, pero conoce bastante sobre él y es un potencial beneficiario o crítico de ese trabajo. Ambos tipos son esenciales para el funcionamiento del proceso de evaluación y plantean sus propios problemas.

176

La evaluación por pares parece seguir dos modelos. El primero sería un mecanismo para certificar que la actividad investigadora es conocimiento y acreditar a las investigaciones que han producido conocimiento científico, a la vez que evita que el conocimiento básico se contamine con errores (y que al público le lleguen investigaciones que no son sólidas). Pero también es un mecanismo para distribuir de manera eficiente y equitativa los recursos escasos que son necesarios y disponibles para llevar a cabo la investigación. El primer modelo parece conllevar que la revisión por pares es sólo un *input* entre varios para tomar una decisión final, pero no el mecanismo por el cual se toma la decisión, ya sea ésta financiar una investigación o publicar un artículo resultado de una investigación. En cambio, el segundo modelo implica que la evaluación por pares sirve para tomar la decisión final. El primer modelo parece ajustarse a las publicaciones en revistas, mientras que el último encaja bien con las decisiones de financiación, en las que los financiadores no tienen un conocimiento del campo o campos que se están financiando (o, al menos, no de todos ellos). El primer modelo considera al par como un testigo experto mientras que en el último es más un jurado.

Desafortunadamente, muchas veces los evaluadores o revisores no reciben instrucciones que especifiquen de manera clara cuáles de los dos papeles deben desempeñar. Si se les pide que den un veredicto sobre una propuesta o un proyecto que debe ser financiado, entonces tendrá algún sentido conocer el número de competidores, los fondos disponibles para financiar las propuestas y cualquier otra

constricción que pueda influir en las decisiones del organismo financiador. Por otro lado, si los revisores están ofreciendo simplemente un juicio experto, entonces la entidad financiadora debe aclarar los tipos de juicio que considera útiles y cuáles consideraría inútiles.

El concepto de “par” implica igualdad de estatus, aunque, como diría Orwell, algunos pares son “más iguales” que otros. Los pares que trabajan en un determinado campo en el que son bien conocidos, están institucionalmente bien situados y desean revisar a sus colegas, ejercen un influjo desproporcionado sobre el sistema, en gran parte debido a la inercia institucional, pues los editores y los financiadores confían en quienes han hecho el trabajo en el pasado a la vez que les suelen pedir que den nombres de otros potenciales evaluadores.

Los estudios empíricos de procesos de revisión por pares tienden a mostrar que, de todas las formas de discriminación que hay en el sistema, las más difíciles de erradicar son las que se basan en el sexismo y en el nepotismo (Wenneras y Wold, 1997), es decir las que se basan en redes personales y prejuicios contra las mujeres.

Desde un punto de vista sociológico, el nepotismo es una forma de provincialismo que actúa de manera doble contra las mujeres: si yo trabajo en el mismo campo que X y no he tenido contacto con X o cualquiera de los que tienen contacto con X, entonces X no es digna de tener contacto conmigo. Por eso en la práctica el par absoluto está sesgado jerárquicamente hacia los pares mayores y mejor conectados, que suelen ser hombres en abrumadora mayoría.

177

El sexismo refiere a comportamientos individuales o colectivos que perpetúan la dominación de los varones y la subordinación de las mujeres. Es un producto de una sociedad androcéntrica que valora más lo masculino. El sexismo institucional refiere a la manera, muchas veces oculta, en que el sexismo se incrusta en las instituciones sociales, operando con la intención, expresa o no, de dañar a las mujeres. Esa discriminación puede ser directa o indirecta. La directa pretende producir consecuencias negativas para el grupo subordinado. La indirecta es más difícil de erradicar porque está constituida por patrones de comportamiento, políticas y prácticas que, aunque se consideran neutrales, producen consecuencias negativas no deseadas. A veces, la discriminación indirecta es producto de discriminaciones previas que han conformado las políticas, las prácticas, la composición de la comunidad o de la organización. Por ejemplo, “la regla de la antigüedad” salvaguarda la posición de los varones blancos, dado que debido a la discriminación directa previa, ellos dominan en las instituciones. Luego se establecen otros requisitos aparentemente neutrales que impiden de *facto* que las mujeres entren. La discriminación indirecta, además, puede producir discriminación “de efecto colateral”, cuando la discriminación en un área conduce a la discriminación indirecta en otra. Por ejemplo, si las mujeres son discriminadas en la educación de modo que no consiguen la misma educación que los varones, luego serán discriminadas económicamente, al no poder competir en igualdad de condiciones por el trabajo. Por ese motivo son necesarias las denominadas acciones positivas, esto es, acciones

que pretenden compensar situaciones de desigualdad preexistentes: dado que si la ley trata como igual lo desigual, perpetúa la desigualdad.

Wenneras y Wold (1997), en un excelente trabajo publicado en *Nature*, analizaron las evaluaciones de los proyectos presentados al *Medical Research Council* sueco en 1995. En él analizaban los diversos ítems puntuables: el proyecto y la competencia científica del evaluado de acuerdo con el ISI, que es sumamente objetivable (es decir, número de publicaciones, número de publicaciones de primera autoría, impacto total de las publicaciones, impacto de las de primera autoría, citas totales en el año anterior y en las de los artículos de primera autoría). Su análisis buscaba averiguar si influía en la decisión de los evaluadores el sexo de la persona evaluada, su nacionalidad, titulación básica, área de conocimiento, el comité de evaluación, la experiencia postdoctoral, la relación profesional o personal con un miembro del tribunal, las cartas de recomendación y la universidad de pertenencia. Lo que descubrieron fue que a los varones se les puntuaba 2,6 veces mejor el CV que a las mujeres, y que también influía la relación o afiliación con alguno de los evaluadores.² El *Medical Research Council* analizó la siguiente convocatoria y descubrieron que se repetía lo mismo y que, además, a aquellas mujeres que obtenían financiación se les daba menos dinero.

178

Los pares relativos también pueden estar sometidos a prejuicios considerables. Por ejemplo, cierta evidencia sugiere que las investigadoras feministas raras veces hacen de pares de hombres, a pesar de que sus intereses se solapan sustancialmente, mientras que a los hombres se les pide periódicamente evaluar la investigación feminista debido a ese solapamiento (y se les pide por lo general que revisen propuestas que están más lejos de su campo que lo que se les pide a las mujeres). Otro problema adicional de los pares relativos es que hacer ese peritaje presupone un “mapa del campo” que permite que las especialidades se dividan en dominios separados, como naciones-estados, cada uno con su soberanía territorial. Esta metáfora cartográfica, utilizada por S. Fuller (2002), permite distinguir entre “regiones” e “imperios” que posibilitan que quienes están en ciertas especialidades tengan dominio y poder intelectual que va más allá de su propia área de conocimiento.

El uso de la evaluación por pares para valorar propuestas de financiación es una de las pocas ocasiones en que las comunidades científicas están cerca del autogobierno. Por analogía con los cuerpos legislativos, podemos distinguir entre la representación de pares unicameral y bicameral. En el enfoque unicameral, un único conjunto de pares decide sobre el mérito relativo y absoluto de las propuestas. Un ejemplo típico son las comisiones, en las que cada una de las personas que

² Por ejemplo, se establecieron cinco categorías en función del impacto total de las publicaciones: la de 0-19, 20-39, 40-59, 60-99, más de 99. Pues bien, a una mujer con un impacto total de publicaciones de más de 99 se le puntuaba por debajo de un varón que tuviera entre 20-39 de impacto total, es decir, quedaba en la segunda categoría, en vez de en la quinta en la que estaban sus colegas varones.

interviene en ellas evalúa privadamente las propuestas y luego se juntan para comparar sus puntuaciones y decidir el orden de las propuestas. Este sistema es interesante y a la vez peligroso porque permite la posibilidad de que los pares reinterpreten o cambien sus puntuaciones cuando se reúnen. Según otro estudio de Weneras y Wold (1999), cuando el *Medical Research Council* sueco aceptó las reformas que ellas propusieron, el sesgo de género se redujo espectacularmente, pues los evaluadores tenían que transmitir electrónicamente sus puntuaciones de todas las propuestas antes de decidir entre todos cuáles debían ser financiadas. Sin embargo, el nepotismo -que, como hemos visto, también juega en contra de las mujeres- ha resultado más difícil de manejar, aunque se prohíba a los evaluadores interesados participar en las decisiones de financiación.

En el enfoque bicameral, cada a propuesta se envía primero a varios evaluadores, cada uno de los cuales no sabe nada de las otras propuestas ni de las identidades de los otros revisores o evaluadores. Se supone que estos pares establecen el mérito absoluto de la propuesta. Una vez que ellos han hecho su trabajo, el mérito relativo de las propuestas se establece en un procedimiento semejante al del enfoque bicameral, excepto en que los pares aquí, un conjunto distinto de personas, no sólo puntúan las propuestas, sino que asignan pesos relativos a las puntuaciones expresadas por los pares de la primera Cámara. ¿Qué principios deben determinar la composición del panel o comisión?, ¿en qué medida deben intervenir cuestiones de género, edad, categoría, etnia, especialidad, región geográfica, afiliación institucional?

179

Si aceptamos el principio de que la revisión por pares debe juzgar más el trabajo que el trabajador, la evaluación por pares en las revistas académicas ha evolucionado cada vez más hacia la consideración de los manuscritos de una forma anónima. Por ello cabe preguntarse si los evaluadores de proyectos de investigación, becas, etc., necesitan saber las identidades exactas de quienes las proponen, por si es posible, y deseable, considerar las propuestas al menos en parte de manera anónima (de hecho, en algunos países se evalúa primero el proyecto de forma anónima y luego se evalúa el currículum del investigador principal y del resto de los investigadores participantes, bien sea por otras personas, bien por la o las mismas, pero una vez evaluado lo anterior). También se puede plantear la misma cuestión a la inversa y recibir una respuesta diferente: ¿deben conocer los investigadores que envían sus propuestas y su trabajo para ser evaluado quiénes los evalúan? Por ejemplo, en Suecia, los investigadores cuyas propuestas han sido revisadas por organismos financiadores estatales tienen derecho, por ley, al texto completo de las evaluaciones.

En la evaluación por pares surgen problemas éticos que, normalmente, tienen que ver con cuestiones de prioridad. En cierto sentido, la evaluación ideal por pares es aquella que hace alguien que trabaja en el mismo campo que el investigador cuya obra o proyecto se está revisando. Sin embargo, ese evaluador o evaluadora está también en una posición ideal para terminar el proyecto propuesto, no financiando la propuesta o impidiendo la publicación y apropiándose luego de ideas o técnicas para los propios propósitos del evaluador. Incluso un evaluador inteligente puede

apropiarse de algo tan simple como la bibliografía que aparece en el proyecto o artículo de modo que pueda lograr rápidamente un nivel de competencia comparable con el del investigador que está evaluando. De ambas cosas tenemos diversos ejemplos, lo que ha puesto en cuestión el sistema de evaluación.

Para evitar este supuesto fracaso, algunos autores³ han propuesto ciertas medidas. Pero dicho fracaso se exagera y se basa en una concepción demasiado individualista de la construcción del conocimiento. El sistema de arbitraje es un aspecto más de esa construcción, pues, tras la publicación, los trabajos son sometidos a crítica mediante las cuales se redefinen ideas y técnicas. Así pues, el conocimiento científico es producto de la comunidad y trasciende la contribución de cualquier individuo (o sub-comunidad).

La nueva forma de producción del conocimiento exige, además, como ya hemos dicho, nuevas formas de evaluación y control en las que intervengan otros intereses, valores, etc., y que no pueden quedar en manos exclusivamente de la comunidad científica, sino que debemos incluir un sistema de “democracia deliberativa” en el que las decisiones se tomen no por imposición ni por suma de votos, sino a través de la deliberación. No sólo cuando hay desacuerdo, sino cuando las cuestiones científicas afectan a diversas comunidades o amplios sectores de la población, a la hora de tomar decisiones hay que valorar, previamente, las propuestas, el intercambio de argumentos y justificaciones para avalarlas, el acuerdo entre las partes acerca de qué compromisos adquiere cada uno para llevar a cabo lo que le corresponde. En resumen, conviene aumentar las negociaciones con los sectores afectados, no dejar la toma de decisiones sólo en manos de los expertos, y potenciar debates sobre diversos temas en la esfera pública.

180

Esa deliberación exige una transparencia que puede parecer extraña o irreal. Por un lado, es imposible conocer todos los pasos y todo lo que implica un experimento; por otro, resulta bastante complicado monitorizar cada etapa de la cadena de toma de decisiones, lo que significa que se corre el riesgo de que se tomen las decisiones sin que la gente, a través de sus representantes, esté implicada en ellas. No obstante, en las sociedades democráticas hay ocasiones en que se toman decisiones en secreto de forma deliberada, aunque suelen ser excepcionales. Pero además de esos contextos excepcionales, todavía hay decisiones sobre cuestiones tecnológicas importantes que se toman sin consultar con el parlamento o el público, incluso en Europa occidental (aunque, en esos casos, no está del todo ausente la transparencia porque es una exigencia de las sociedades democráticas que se dé publicidad a, y se discutan, las decisiones, incluso bajo circunstancias extraordinarias).

La exigencia de transparencia en nuestras sociedades democráticas viene por otro camino, porque los sistemas científico-tecnológicos son tan complejos que sólo los expertos los hacen funcionar o entienden su funcionamiento, lo que hace que mucha

³ Véase Fuller (2002), Pérez Sedeño (2006).

gente se sienta excluida de una supervisión efectiva democrática de esas actividades. El resultado es una paradoja que sólo una buena educación y difusión de la ciencia y la tecnología, en todos los niveles y edades, puede eliminar.

Bibliografía

FULLER, Steve (2002): *Knowledge Management Foundations*, Boston-Oxford, Butterworth Heinemann.

GIBBONS, Michael , et al (1994): *The New Production of Knowledge*, Londres, Sage.

LENOIR, Timothy, HAYS, Marguerite (2000): "The Manhattan Project for Biomedicina", en Sloan, Phillip (ed.)

PÉREZ SEDEÑO, Eulalia (2006): "Objetividad, evaluación por pares y valores", en J.F. Álvarez y R. Rodríguez Aramayo (eds.) *Disenso e incertidumbre. Un homenaje a Javier Muguerza*, Madrid, CSIC-Plaza y Valdés.

SLOAN, Phillip [ed.] (2000): *Controlling Our Destinies: Historical, Philosophical, Ethical, and Theological Perspectives on the Human Genome Project*, University of Notre Dame Press.

WENERAS, Christine, WOLD, Agnes (1997): "Sexism and nepotism in peer review", *Nature*, 387, pp. 321-343.

_____, (1999): "Bias and peer review of research proposals", en Smith, J. y Smith, R. (eds.), *Peer Review in Health Care*, British Medical Journal Publishing, pp. 77-87.

ZIMAN, John (2000): *Real Science*, Cambridge University Press. Traducción español: *La ciencia tal cual es* (2002): Madrid, CUP-Iberia.