

Excelencia científica: la construcción de la ciencia a través de su evaluación. La Comisión Sectorial de Investigación Científica (CSIC), Uruguay*

*Amílcar Davyt y Léa Velho***

El presente artículo estudia el proceso social de construcción de la excelencia científica a partir del análisis de los procesos de evaluación de proyectos realizados en la Comisión Sectorial de Investigación Científica del Uruguay. Se investigan los procesos de construcción de consenso y especialmente la influencia del criterio de relevancia social en la definición de la excelencia. El trabajo se establece a partir del marco de referencia provisto por la sociología de la ciencia y la tecnología de orientación constructivista.

1. Introducción

La palabra *excelencia*, habitualmente acompañada de los adjetivos “científica” o “académica”, se escucha cotidianamente en ámbitos vinculados a la ciencia: la excelencia se busca en propuestas y en resultados de investigación. En la comunidad científica, en el gobierno, en el sector productivo y aun en la población en general se habla de criterios de evaluación que buscan la excelencia científica, y que legitiman y justifican la distribución de recompensas y decisiones de todo tipo. Sin embargo, la definición del concepto es pocas veces explícita y frecuentemente omitida al realizar evaluaciones o discutir criterios.¹

* Este artículo es una síntesis de la Tesis de Maestría del primero de los autores, realizada bajo orientación del segundo en el Departamento de Política Científica y Tecnológica de la Universidad Estadual de Campinas, Brasil. Se agradece a los integrantes actuales y pasados de la Unidad Académica de la Comisión Sectorial de Investigaciones Científicas por brindar todas las facilidades posibles para la realización del trabajo, a los entrevistados por dedicar su tiempo, información y opinión, e institucionalmente a la propia CSIC. A todos ellos eximimos de responsabilidad por el resultado final.

** Departamento de Política Científica y Tecnológica de la Universidad Estadual de Campinas, Brasil.

¹ Pocos han explorado la noción de excelencia. En 1960, R. Merton comenzaba así su análisis sobre la evaluación en ciencia: “Muchos de nosotros estamos convencidos de que *sabemos* lo que queremos decir con *excelencia* y preferiríamos que no se nos pidiera una explicación. Actuamos como si creyéramos que una inspección minuciosa de la idea de excelencia pudiera disolverla en nada”

Esta cuestión conceptual es el tema central de este trabajo: buscamos analizar y explicar el significado del concepto que representa el término excelencia, y su relación con la relevancia, dentro de los procesos de evaluación científica realizados por los investigadores. Dentro de las nuevas perspectivas en el estudio de la ciencia delineadas por los enfoques post-kuhnianos, según las cuales el contenido de la ciencia puede analizarse sociológicamente, se ha explorado la “construcción” del conocimiento científico en lo que se considera el *locus* privilegiado donde ocurre: el laboratorio.² Sin embargo, no sólo allí la ciencia es construida: los procesos de evaluación –para su publicación o financiamiento– son locales de construcción de la ciencia tal vez tan importantes como el laboratorio. Para explorar, entonces, el significado de la excelencia en ciencia, nos pareció necesario entrar en el *locus* donde se determina lo que es académicamente excelente: los procesos de evaluación de las actividades científicas.

Estos mecanismos son parte importante del propio proceso de construcción del conocimiento. Una manera casi directa de “dar forma a la dirección y al contenido del trabajo científico” (Chubin y Hackett, 1990, p. 8) es a través de su financiamiento. Sólo el proyecto financiado tiene posibilidades de dar lugar a nuevo conocimiento científico. Las propuestas rechazadas –académicamente o por falta de recursos– no se dirigen a ninguna parte (a no ser que sean financiadas por otro organismo). El financiamiento es un determinante clave no sólo de la dirección de la ciencia, sino también de su contenido: cuando es financiado sólo uno de dos proyectos que buscan resolver el mismo problema por vías teórico-metodológicas diferentes, una respuesta posible es de hecho privilegiada, impidiendo el desarrollo de su alternativa.

“La ciencia puede verse como el producto de múltiples realidades; sólo una de ellas es generalmente aceptada en un momento dado” (Chubin y Hackett, 1990, p. 8), en parte porque es financiado el inves-

(Merton, 1973 [1960], p. 422). Años después, en un intento por conjugar sociología de la ciencia y psicología se concluye que: “la noción de excelencia científica continúa siendo elusiva; es una entidad no fácilmente identificable o rápidamente medible” (Zuckerman, 1987, p. 7).

² Hay muchos trabajos en esta perspectiva: *Laboratory Life: The Social Construction of Scientific Facts* de Latour y Woolgar (1979) es el más conocido. También pueden verse Zenzen y Restivo (1980), Knorr-Cetina (1981), Law y Williams (1982), Lynch (1982). Los fundamentos de la idea de que estudiando el laboratorio se pueden entender procesos que ocurren “fuera” de él están en Latour (1983).

tigador o la propuesta de investigación (con cierta metodología o en una línea teórica específica). Ésta puede dar lugar a algunas de esas múltiples realidades, pero no a otras. En un proceso sin fin, evaluación tras evaluación, algunas de las posibilidades van “muriendo”, dejando de “existir”, y otras “sobreviven”. En los laboratorios se toman pequeñas decisiones cotidianas respecto a cómo interpretar los datos obtenidos o cómo usar una metodología. Pero quizás las grandes direcciones, las líneas generales de una u otra área, son en parte definidas en el *locus* de la decisión del financiamiento, en la arena de negociación que vincula el mundo de las actividades científicas con el mundo de los recursos. Y la decisión sobre qué financiar es tomada, tradicionalmente, sobre la base de la opinión de los pares que, a veces, simplemente apuntan lo que es *excelente* (o “tiene mérito”) y, por lo tanto, merece ser financiado; en otras, comparan la “grandeza de las excelencias” e indican cuál debe tener prioridad a la luz de las circunstancias. Dada la centralidad del juicio de los pares en la construcción de la excelencia científica, se presenta a continuación una discusión sucinta sobre algunos de sus aspectos.

2. La evaluación de las actividades científicas

La evaluación en ciencia es casi sinónimo de revisión por pares, es decir, aquélla realizada a través de la opinión sobre el trabajo de una persona por otras con calificaciones similares. Siendo una práctica iniciada con la revisión de manuscritos enviados para publicación, es hoy también utilizada por las agencias financiadoras para distribuir recursos.³

La evaluación por pares ha recibido variadas críticas, referidas a los mecanismos implementados y a sus principios y fundamentos. Como respuesta, estudios sobre los procedimientos de diversas agencias fueron desarrollados en los sesenta y setenta en los Estados Unidos y luego en otros países. Aunque los estudios no han sido capaces de terminar con las críticas, generando aún más controversia, ciertos

³ Según Rip (1994, pp. 7-8), el sistema de pares se originó en las propias agencias de fomento y, a diferencia de la revisión de artículos científicos, generó una interacción entre la comunidad científica y los organismos gubernamentales, que fue gradualmente aceptada, “capturada” por los científicos y transformada en parte del sistema de recompensas de la ciencia.

cambios han sido implementados en algunas agencias.⁴ Sin embargo, estos cambios no han dado aún respuestas satisfactorias a preguntas como: “¿Qué concepto de objetivos y qué criterios de mérito vamos usar en la evaluación de la investigación? ¿Vamos a restringirnos al concepto de excelencia científica, conforme es definido por la comunidad científica en determinado momento?” (Cozzens, 1990, p. 293).

Algunos han respondido proponiendo usar dos tipos de criterios –internos y externos– para seleccionar proyectos de investigación (por ejemplo, Ziman, 1994, pp. 100-101). Los primeros refieren a atributos que sólo pueden ser apreciados por especialistas de la misma disciplina; los externos incluyen resultados esperados de tipo científico, tecnológico o social. En realidad, tanto los cambios implementados en los procedimientos como las sugerencias de autores destacan la importancia de obtener la opinión de no-especialistas en relación con la relevancia de la investigación. Esto, ya propuesto por Weinberg más de 30 años atrás,⁵ no ha sido aún adoptado por la mayoría de las agencias de fomento a la investigación del mundo.

La necesidad de considerar la relevancia para la sociedad de la investigación realizada ha sido también enfatizada en América Latina por intelectuales que, en los sesenta y setenta, ya analizaban las políticas de cyt implementadas en la región (Herrera, 1973; Sábato y Botana, 1968; Varsavsky, 1969). Uno de los puntos destacados estaba constituido por lo que sostenían era un proceso de “alienación” de la comunidad científica local: los científicos distribuían fondos de investigación sobre la base de criterios científicos configurados en el exterior, sin tomar en cuenta la aplicabilidad de los resultados a las necesidades locales.

Sea como sea, es posible preguntar si la relevancia no está, de algún modo, incorporada en la opinión de los propios pares. ¿Las relaciones cotidianas de los investigadores con su contexto no influyen en sus juicios sobre mérito científico? ¿Es posible separar las valoracio-

⁴ Por ejemplo, la NSF incluyó la relevancia como criterio explícito en su proceso de evaluación; según Chubin y Hackett, “[...] el cambio en la NSF de evaluación *por pares* a evaluación *de mérito* –esta última mezclando opiniones de pares sobre mérito técnico con asesoramientos de utilidad práctica– señalan un nuevo punto de balance en el equilibrio dinámico entre conocimiento y utilidad” (1990, p. 159, en cursiva en el original).

⁵ Weinberg (1963, 1964) criticó el “universo cerrado” de la evaluación, argumentando que la opinión de los pares tiene como referencia reglas creadas por la propia comunidad disciplinaria, sin cuestionar su validez, y proponiendo obtener opiniones de personas de otras disciplinas –sobre el impacto en las demás áreas– y de fuera de la comunidad científica –sobre el impacto en el conjunto de la sociedad–.

nes sobre las proposiciones lógicas (excelencia científica) de las opiniones sobre oportunidad (relevancia de la investigación)?

Muchos estudios han sido realizados sobre diversos aspectos del sistema de evaluación por pares, algunos centrados en procedimientos adoptados por revistas científicas (Crane, 1967; Mahoney, 1977; Peters y Ceci, 1982; Cicchetti, 1982), otros de asignación de fondos a proyectos (Gillespie *et al.*, 1985; Hackett, 1987; Sonnert, 1995); algunos fueron encargados por agencias de financiamiento u otros órganos gubernamentales (Hensler, 1976; NIH Grants Peer Review Study Team, 1976; Cole, Rubin y Cole, 1977; Cole, Cole y Simon, 1981; McCullough, 1989; Cogan, 1986; Boden, 1990; Sirilli, y Meliciani, 1994). Asimismo, la metodología y las fuentes de información utilizadas también han sido muy variadas. Prácticamente todo este esfuerzo de investigación tuvo lugar en los Estados Unidos y en otros países avanzados; América Latina ha contribuido muy poco en investigación empírica en la materia.⁶ Sea como sea, ninguno de los muchos estudios, algunos de ellos ya clásicos, analizó el problema de la excelencia y los mecanismos de evaluación desde esta perspectiva: su papel como parte del proceso de construcción del conocimiento científico.

Este artículo lidia precisamente con esta cuestión. Se intenta realizar esta contribución, es decir, mostrar empíricamente cómo ocurre este proceso de construcción de la ciencia, a partir de una implementación institucional específica de prácticas de evaluación: la Comisión Sectorial de Investigación Científica, órgano que desarrolla programas de fomento a la investigación en la Universidad de la República del Uruguay. Atacando el problema a partir del estudio de esta situación concreta, se pretende hacer derivaciones para casos similares. Sin embargo, es preciso destacar que es una situación en la que se comienza a institucionalizar una práctica científica. En este sentido, es pertinente recordar la distinción, ya clásica, entre las condiciones de los países desarrollados y científicamente centrales de aquéllas de los países periféricos a la ciencia *mainstream*, en los cuales se pretende desarrollar una infraestructura adecuada para la ciencia. El estudio de las condiciones particulares de las prácticas de evaluación científica en estos países puede ser un insumo fundamental para las discusiones sobre el tema y contribuir a cambios en los sistemas.

⁶ Encontramos sólo un estudio empírico sobre evaluación por pares en Latinoamérica: Bonilla *et al.* (1995). Otros artículos sobre el tema son reflexiones generales sobre los procedimientos, como Loría y Loría (1996); Nicoletti (1985).

3. La CSIC y sus procesos de evaluación

Luego del régimen militar (1973-1985) el sistema científico y tecnológico uruguayo comienza a ser reconstruido lentamente. En 1990 las acciones desarrolladas por la Universidad de la República para apoyar las actividades científico-tecnológicas culminaron en la constitución de la Comisión Sectorial de Investigación Científica (en adelante CSIC). Ésta tiene la capacidad de realizar y coordinar actividades y programas de fomento integral de la investigación y se ha constituido en la agencia financiadora de las actividades de investigación en el interior de la universidad. Ésta, por su parte, fue, hasta no hace más de 15 años, la única institución de educación superior del país y aún es la única donde se realiza investigación en todas las áreas del conocimiento; es responsable de más del 60% de las actividades científicas desarrolladas en el país.⁷ Y, aunque existen fondos provenientes del Consejo Nacional de Investigaciones Científicas y Tecnológicas, los recursos controlados por la CSIC son significativos.⁸ Por lo tanto, el análisis de los procesos de evaluación realizados por la CSIC implica el estudio de gran parte del sistema científico-tecnológico local.

Confluyen en la CSIC todas las grandes áreas científico-tecnológicas en la tarea de juzgar el desempeño de investigadores y evaluar la calidad de propuestas científicas, a fin de otorgarles —o no— determinadas recompensas. Sus procedimientos son similares, en general, a los de cualquier otra agencia de fomento a la investigación. La integración de la Comisión es mixta, en respuesta a diferentes exigencias. Por un lado, ella tiene, de acuerdo con la tradición de universidad cogobernada, representantes de los diferentes órdenes universitarios y del rector; por otro lado, incluye, siguiendo el patrón internacional de agencias de fomento, a representantes de las diferentes áreas del conocimiento.⁹

⁷ La última información disponible afirmaba que 58% de las actividades de I+D del país eran llevadas a cabo por la Universidad de la República (Argenti *et al.*, 1987). Estimaciones más recientes destacan un aumento de esta concentración de actividades (Saráchaga, 1997).

⁸ Desde 1992 a 1996 el CONICYT distribuyó cerca de u\$s 10 millones por año (Bortagaray y Sutz, 1996). La cantidad asignada por la CSIC fue de u\$s 7 millones por año en el mismo período, distribuidos exclusivamente dentro de la universidad.

⁹ Estos representantes son “delegados” de subgrupos de facultades e institutos, simplemente llamados áreas en la jerga universitaria: a) Agraria: agronomía y veterinaria; b) Básica: ciencias y química; c) Salud: medicina, odontología y enfermería; d) Social: ciencias sociales, derecho, economía, humanidades y educación, psicología y bellas artes; e) Tecnológica: arquitectura e ingeniería.

La csic desarrolla tres tipos de programas, con diferentes modalidades de evaluación, coordinados por subcomisiones específicas: a) suplemento salarial por desempeño integral en la universidad (régimen de dedicación total); b) propuestas de investigación y desarrollo (proyectos de I+D y proyectos de vinculación con el sector productivo) y c) diversas actividades de apoyo y formación de recursos humanos. En los dos primeros tipos de programas el proceso de evaluación implica el uso de pares, categorías y *rankings*. En las subcomisiones encargadas de la evaluación de proyectos se ha requerido –de forma diferente– la opinión de dos evaluadores.¹⁰ Aunque la decisión final sobre el financiamiento está en manos de la Comisión, ésta habitualmente sigue las recomendaciones de las subcomisiones.

En nuestro análisis utilizamos exclusivamente los procedimientos usados por las subcomisiones que evalúan los proyectos de I+D y los proyectos de vinculación con el sector productivo. Algunas de sus características son las siguientes:

Los proyectos de I+D son el tipo tradicional de propuestas presentadas por investigadores individuales. Los proponentes sólo tienen que especificar en qué área del conocimiento están compitiendo, correspondiente a una de las cinco de csic.¹¹ Estos llamados comenzaron en 1991, pero desde 1992 en adelante fueron realizados cada dos años. Hasta ahora se han realizado cuatro procesos de evaluación completos: 1991, 1992, 1994 y 1996. Los procedimientos variaron un poco en cada llamado, pero todos incluyeron dos evaluadores externos –pares en sentido estricto– seleccionados por la subcomisión, la cual a su vez aprueba el fallo final. En 1994 se estableció un sistema de paneles, uno para cada una de las cinco áreas, que recogen las evaluaciones de los pares y, junto con sus propias opiniones, construyen los *rankings*, ratificados por la subcomisión y por csic por medio de discusiones conjuntas.

El Programa de Vinculación con el Sector Productivo, también creado en 1991, define un proyecto que difiere del tradicional en el

¹⁰ Para mayores detalles respecto de los procedimientos, véase Davyt y Yarzabal (1991), y Mujica (1997).

¹¹ Aunque ellas coinciden nominalmente con las áreas de csic, es necesario aclarar que: Agraria incluye investigaciones vinculadas a agronomía o veterinaria, “provenientes de” las ciencias básicas –esto es, no las que implican un componente social importante–; algo similar ocurre con Salud; Básica incluye las así llamadas ciencias “duras”; Social, tanto las fundamentales como las aplicadas, y también las humanidades y artes; Tecnológica está relacionada con ingeniería, incluyendo parte de la biotecnología –aquella vinculada a ingeniería química– y de arquitectura –en sus componentes técnicos.

sentido de que debe “dirigirse de forma inequívoca al tratamiento de problemas de interés directo de los sectores productivos”.¹² El programa apoya dos tipos de proyectos: a) aquellos en que existe una participación directa del sector productivo, con una contraparte financiera; y b) aquellos de iniciativa exclusiva de grupos universitarios, que fomenten y consoliden la capacidad de relacionarse con el sector productivo; éstos implican sólo una declaración de interés de este sector. Aquí, los proponentes no se clasifican en ninguna de las áreas clásicas del conocimiento de CSIC.

Las características de este programa determinaron la necesidad de diferencias en el modo de evaluación en relación con el de los proyectos de I+D. Aquí se introducen “criterios de valoración que se suman a aquéllos, imprescindibles, relacionados a la calidad académica”.¹³ Para esto, cada proyecto es sometido a una doble evaluación: una de un par en sentido estricto que opina sobre los aspectos académicos —o excelencia—. Otro evaluador realiza un análisis tecno-económico del proyecto, considerando su pertinencia, su factibilidad y el impacto de sus contribuciones al desarrollo de empresas o sectores productivos específicos —o relevancia—.¹⁴ La subcomisión ha tenido la palabra final sobre el financiamiento, en los tres llamados realizados hasta el momento: 1992, 1993 y 1996.

Intentamos explorar el significado del concepto de excelencia científica y su vinculación con la relevancia en el marco de la implementación de estos mecanismos en la práctica cotidiana de la evaluación de la CSIC. Analizando lo ocurrido en los procesos de evaluación de las subcomisiones de CSIC buscamos identificar y explicitar el carácter construido, contextualizado y contingente de la excelencia, y, como consecuencia, su estrecha vinculación con la relevancia. A continuación damos cuenta de esa exploración. Con fines analíticos, distinguimos lo que llamamos “momentos” del proceso de evaluación. Es decir, fases en las que los actores llamados a participar, con funciones específicas, son diferentes. Así, identificamos tres momentos: 1) la selección de los pares adecuados por parte de la subcomisión; 2) la ac-

¹² Informe de actividades de la subcomisión a la CSIC de 1993, s/f, p. 2.

¹³ Informe de la subcomisión a la CSIC de 1993, s/f, p. 2.

¹⁴ Este evaluador es un profesional que trabaja en el sector productivo y tiene conocimientos sobre estudios de factibilidad técnica y económica en empresas. Ambos evaluadores deben trabajar dentro del país, “debido a la especificidad de la problemática colocada” (Burgueño y Mujica, 1996, p. 224.)

tuación de los evaluadores seleccionados; y 3) la toma de decisiones final de la subcomisión en relación con los candidatos a ser financiados.

4. Seleccionando a los evaluadores

El primer momento del proceso de evaluación es la selección de los pares adecuados para evaluar la propuesta por primera vez, emitiendo una “evaluación”. En el caso de csic, esta selección está en manos de los miembros de las subcomisiones. Se entrevistaron algunos miembros, actuales o pasados, de las subcomisiones,¹⁵ a fin de identificar los elementos presentes en este proceso, especialmente los criterios para la selección de los pares.

Según los entrevistados, la selección de los evaluadores *ad hoc* se inicia con la búsqueda de un especialista en la disciplina y, si es posible, en la propia línea de trabajo del aspirante. En este sentido, la definición de evaluador, de par, es bastante estricta.¹⁶ Otro criterio previo, definido por los integrantes de las subcomisiones, es la necesidad de que el árbitro elegido esté “lejos” del proponente bajo evaluación. O sea: “cuanto más alejados estén los evaluadores de los evaluados y menos los conozcan, mejor”. Puesto que parece difícil que el especialista no utilice, de una forma u otra, su conocimiento “extra” al de la propuesta o el currículum, parece adecuado, para alcanzar mayor imparcialidad y objetividad, buscar el mayor “desconocimiento” posible entre evaluado y evaluador, “a menos que haya sido imprescindible, o que no nos hayamos dado cuenta”. Estas palabras reflejan la siguiente idea:

La comisión, en el entendido de que el evaluador tiene que estar libre de cualquier tipo de presiones, de la misma manera que no elige amigos ni enemigos, no elige conocidos. [...] *La evaluación* que se hace del nivel científico, factibilidad del proyecto, aplicación de resultados, *no pasa por una ecuación estricta que termine con un único resultado*, entonces cuanto menos se conozca a alguien, es más imparcial el jui-

¹⁵ Fueron entrevistados siete investigadores. Tres de ellos trabajan en química, dos en biomedicina, uno en agronomía y uno en sociología. Dos de ellos han participado como delegados del área Básica, dos del área Tecnológica y uno de Agraria, Salud y Social, respectivamente.

¹⁶ Solamente cuando se hace difícil encontrar estrictamente tal par, se busca en un círculo un poco más amplio, pero intentando siempre no sobrepasar los límites de la disciplina, ya que “si no, no es un evaluador”. En el lenguaje cotidiano de la csic, el término “evaluador” es utilizado como sinónimo de revisor, árbitro o *referee*.

cio que uno pueda hacer. [...] Las desavenencias, rivalidades que existen en el país hacen que sea difícil; la comisión que dirige la evaluación tendría que conocer bastante quiénes son los enemigos y quiénes son los amigos. Eso lo aprendimos (las cursivas son nuestras).

Los entrevistados dieron opiniones diferentes en relación con las dificultades del evaluador para excluir los asuntos personales, en particular, su comportamiento frente a las relaciones de amistad. Para algunos, eso puede hacer que el nivel de exigencia sea menor: “Un amigo es un amigo, y es difícil que uno lo pueda juzgar bajándole demasiado el martillo, es decir, no hay una rigurosidad igual cuando es enemigo que amigo”. Pero otro expresó que: “Yo confío que la mayoría de los investigadores agarran el proyecto de un amigo y lo juzgan tal vez un poco más severamente que el proyecto de un enemigo”. Esta última, sin embargo, es opinión minoritaria y la búsqueda de evaluadores que no conozcan al evaluado es la política que prevalece. Esto, aliado a las características de la comunidad y del país en cuanto a tamaño, ha inducido a las subcomisiones a buscar evaluadores fuera del país.¹⁷ Los científicos extranjeros conseguirían, según los entrevistados, evaluar con mayor “objetividad”.

Uruguay tiene una ventaja comparativa importante: por ser tan chiquito, mandamos los proyectos a evaluar fuera, entonces las influencias subjetivas, aunque existen, son menores. Existen, porque vos tenés colaboración con los de afuera, porque muchas veces llegan a uruguayos en el exterior, entonces están involucrados subjetivamente también. Existen, pero menos que si evaluás en el mismo ambiente.

Esta política, aunque compartida en términos generales, admite varias consideraciones específicas al ser concretada. Una de las primeras diferencias se establece entre áreas del conocimiento; este criterio de extrema separación entre evaluado y evaluador parece ser mucho más evidente en áreas básicas que en las aplicadas. Tratándose de ciencia básica, el concepto de universalidad de las metodologías y de las ideas, en fin, de paradigmas, presente en el raciocinio de los investigadores, hace que sea totalmente adecuado consultar a un científico del exterior: “Porque el tipo se va a fijar cuáles son las ideas

¹⁷ En el caso de proyectos de I+D, por ejemplo en el llamado de 1992, cerca del 85% de los evaluadores fueron del exterior, principalmente de Chile, la Argentina y el Brasil; en 1994, el porcentaje fue menor, 65%, principalmente de la Argentina, Chile y el Brasil.

que vos estás planteando en el proyecto y con qué método vas a abordar el problema, eso es universal”.

La consulta a pares extranjeros puede ser un problema en las áreas más aplicadas, de proyectos tecnológicos o de desarrollo, cuyos objetivos se vinculan a la resolución de cuestiones locales. En este caso, la necesidad de contextualizar los temas de investigación, y la evaluación consecuente, es más importante que el criterio de evitar la subjetividad.¹⁸

En cosas que sean muy locales, es muy difícil. Alguien puede decir: “a quién le interesa tal cosa”; sí, al Uruguay le puede interesar eso, no le interesa al mundo, pero al Uruguay le puede interesar.

Yo soy un defensor de los evaluadores nacionales en las áreas tecnológicas, no creo que un tipo de afuera tenga más experiencia para evaluar un proyecto. [...] A pesar de ser de I+D, tenían un componente tecnológico muy grande; enviados para afuera, el tipo ni siquiera entendió de lo que se trataba el tema, y no porque no fuera capaz. Hoy acá vivimos una cierta realidad de la producción y el tipo ni siquiera se puede imaginar, en otro lado, cómo es que funciona la producción en nuestro país. De repente en la industria no es tan así, pero en agronomía hay particularidades muy específicas...

Este problema, que parece claro tratándose de proyectos tecnológicos, también surge en otras áreas, ya que el contexto no sólo se refiere a la realidad productiva, sino también a otros elementos. Así, el conocimiento de la situación que rodea al investigador dentro del propio ámbito académico parece ser un atributo fundamental de un juicio equilibrado. Y aquí aparece, entonces, una contradicción, una tensión, entre la necesidad de estar lejos del evaluado para no introducir elementos subjetivos, y la necesidad de conocer el ambiente donde se mueve el proponente del proyecto. De esto dan cuenta las siguientes opiniones:

Ése es uno de los problemas cuando uno usa un evaluador externo: si el evaluador externo no conoce la situación ni el grupo, en el país, se le van a presentar dudas que no tendría si lo conociera.

Yo creo que los evaluadores extranjeros ahí están con el *handicap* de que no conocen el medio. Me dicen, por ejemplo: este proyecto en inmunología es demasiado ambicioso; y están mirando desde Chile. Si yo estoy aquí, digo: conozco ese grupo de inmunología y el proyecto es factible. Si yo miro de lejos, digo: este proyecto es un disparate, por-

¹⁸ El problema de la falta de información sobre el contexto por parte de los expertos extranjeros es también discutido por Stolte-Heiskanen (1986).

que es imposible que puedan hacer todo lo que dicen. Entonces, *uno de los problemas del juicio, en la evaluación, viene por la falta de conocimiento de con qué medios se cuenta en un lugar*. El otro problema son las vinculaciones con las que se cuenta; puede ocurrir que se presente un joven como director del proyecto porque el director del grupo ya tiene demasiados proyectos, y entonces ahí se dice: este señor no tiene condiciones de obtener una información bibliográfica que necesitaría. Y resulta que el grupo, que es muy importante, tiene acceso, porque el proyecto es sólo un pedazo de aquello que el grupo hace. Entonces, nuevamente, el evaluador externo no sabe con qué contactos puede contar ese grupo, con qué apoyos externos. *Sin conocer el ambiente es muy difícil evaluar* (las cursivas son nuestras).

A pesar del extremo cuidado en la elección del evaluador, algunos entrevistados destacan otro factor que aparece en esta fase de la evaluación: los elementos fortuitos, que son inevitables: “Yo lo he dicho muchas veces en la csic, hoy la variable más importante que determina el éxito es el evaluador que *le tocó en suerte*”. La posibilidad de corregir eso depende de saber a priori cómo “funciona” cada evaluador, según la siguiente opinión:

La suerte de tener una persona que lo conocía [en la comisión] o la suerte del evaluador. Y sí, eso existe, eso es inevitable. Eso es parte del proceso. Hay evaluadores que son durísimos, y bueno, la comisión tiene que ir aprendiendo a conocer los evaluadores. En este caso, no se les puede dar consejos a los evaluadores, pero... si uno los conoce, sabe cómo hacerlo. Conozco varios evaluadores que ya sé: éste es durísimo, si dijo que era regular, éste es un buen proyecto.

En síntesis, el primer paso de la operacionalización del proceso de evaluación tiene sus secretos. Uno de los objetivos centrales en la selección de los pares es alcanzar un equilibrio entre dos elementos: lo que podríamos denominar “desconocimiento evaluador-evaluado”, que evitaría las subjetividades personales, y el “conocimiento evaluador-contexto”, que evitaría la descontextualización de los juicios. Esto último no sólo aparece con referencia a proyectos aplicados, sino también a aquéllos más académicos. De esta forma, la excelencia parece no tener existencia independiente: sólo puede ser evaluada en un determinado contexto, en determinadas condiciones, incorporando cierta dosis de aleatoriedad.

Así, es posible identificar los diferentes factores que ya en este primer momento comienzan a dar forma al concepto de excelencia. En la

propia selección de los evaluadores, los integrantes de las subcomisiones inciden en la constitución de una excelencia particular, específica. A través de su conocimiento de las características personales del par, determinan en parte el resultado final de la evaluación, pero también lo hacen a través de su “desconocimiento” de la forma de actuación del evaluador, permitiendo la incidencia del azar como factor determinante del éxito de las propuestas. Las condiciones y contingencias del primer momento de la evaluación colaboran en la configuración de la excelencia, constituyendo así el primer paso de su proceso de construcción.

5. La opinión de los evaluadores: divergencias y discrepancias

El segundo momento en el proceso de evaluación de las actividades de investigación es la propia actuación de los evaluadores, es decir, la emisión de la primera opinión respecto de la calidad de la propuesta. En las ocasiones en que existen dos o más evaluadores para cada propuesta, se ha encontrado que los resultados presentan diversos grados de divergencia o desacuerdo entre sí. Aunque las explicaciones para ello varían significativamente, es lugar común el reconocimiento de la existencia de esas divergencias.¹⁹

Analizamos la existencia de divergencias entre evaluadores en los procesos de evaluación realizados por las subcomisiones de proyectos de I+D y de vinculación con el sector productivo. En los primeros, dos pares estrictos buscaban calificar su excelencia. En los segundos, se espera encontrar discrepancias importantes entre opiniones, ya que los dos evaluadores —el par y el experto tecno-económico— buscan evaluar características diferentes: excelencia y relevancia, respectivamente.

5.1. Las divergencias en el proceso de evaluación de proyectos de I+D

A efectos de ilustrar la existencia de divergencias entre las opiniones de evaluadores sobre la misma propuesta, diseñamos un ejercicio simple, usando los resultados de las evaluaciones de proyectos de I+D presentados en 1992 y 1994.²⁰ Primero seleccionamos las propuestas

¹⁹ En general, las causas de las divergencias no han sido claramente establecidas: puede decirse que aquí también existe un importante desacuerdo. Las variadas explicaciones son discutidas más adelante.

²⁰ Se decidió no utilizar los datos de 1991 debido a que fue el primer intento de usar revisión por pares y los procedimientos variaron desde 1992 en adelante.

que fueron evaluadas por dos árbitros, lo que resultó en 157 proyectos en 1992 y 137 en 1994.²¹ Luego observamos el contenido de las evaluaciones y las comparamos. El trabajo de los pares, de acuerdo con las instrucciones de csic, comprende tres aspectos: un puntaje de 0 a 5 sobre doce ítems acerca de la calidad de la propuesta y del proponente;²² un fallo o conclusión final sobre el proyecto –recomendado, no recomendado y recomendado con modificaciones–; y comentarios sustantivos o “apreciación global”. Un número final, también en un rango de 0 a 5, con dos lugares decimales, es calculado para cada evaluación.

Comparamos, entonces, los puntajes finales atribuidos por cada evaluador a la misma propuesta. Luego construimos las siguientes categorías:

- categoría 1 - *coincidencia*: diferencia entre los puntajes atribuidos menor que 1;
- categoría 2 - *divergencia pequeña*: diferencia mayor o igual a 1 y menor que 2;
- categoría 3 - *divergencia grande*: diferencia mayor o igual que 2.

Es posible decir que las diferencias menores que 1 serían esperables, debido a pequeñas diferencias en el juicio y en la apreciación de la calidad existente en la propuesta;²³ la categoría 2 también podría ser aceptable, principalmente en los casos de áreas de “bajo consenso”.²⁴ Lo que parece difícil de explicar son las diferencias grandes, en este caso mayores que 2 puntos, de los 5 posibles.

²¹ 72% de los proyectos en 1992 y 62% en 1994 recibieron respuestas de los dos evaluadores.

²² Algunos ítems se relacionan con calidad científica (“claridad de objetivos”, “diseño y metodología”, “entrenamiento y experiencia del proponente y del equipo”, “bibliografía adecuada”, etc.) otros con impacto (“relevancia para el área”, “aplicabilidad potencial”).

²³ De acuerdo con Merton y sus seguidores la explicación para estas pequeñas diferencias puede residir simplemente en la distinta percepción de la adecuación de la propuesta a los criterios establecidos o en la ambigüedad de tales criterios; véase Zuckerman y Merton (1973 [1971]). Este tipo de causas para las divergencias puede ser identificado con la perspectiva teórico-epistemológica que implica la existencia de un atributo “real” –la calidad científica– respecto del cual distintos observadores –los pares– pueden tener percepciones un poco diferentes. Una explicación así justificaría sólo las pequeñas diferencias en los juicios sobre un objeto que tiene calidad en sí mismo.

²⁴ Aunque el estudio de Zuckerman y Merton (1973 [1971]) estaba enfocado en física, también menciona otras áreas del conocimiento que podrían mostrar mayores divergencias entre pares. Los autores atribuyen las diferencias entre áreas al grado de institucionalización del campo, o sea, al grado de acuerdo o consenso respecto de los patrones de conocimiento adecuado. La misma explicación para divergencias en las opiniones entre pares es ofrecida por autores que analizaron el proceso de evaluación de la NSF (Cole, Rubin y Cole, 1977).

La distribución de proyectos de cada llamado en diferentes categorías se observa en las figuras 1 y 2. Puede verse que hay diferencias difíciles de explicar, respectivamente en 15% y 20% de las propuestas; la coincidencia no alcanza 50% en ninguno de los dos llamados. La evaluación del llamado de 1994 implicó mayores divergencias.

Figura 1. Distribución del total de proyectos en las diferentes categorías en el llamado de I+D de 1992

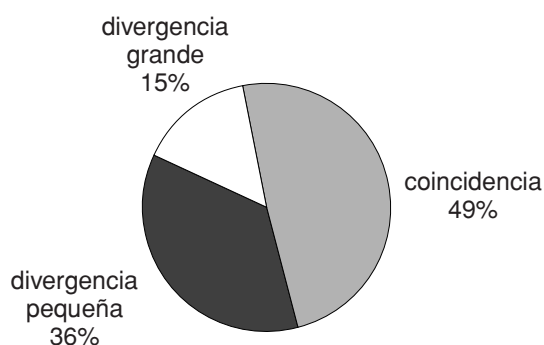
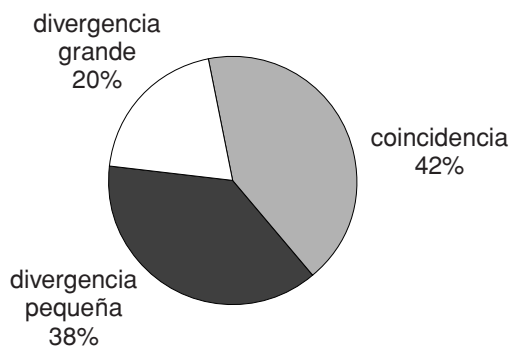
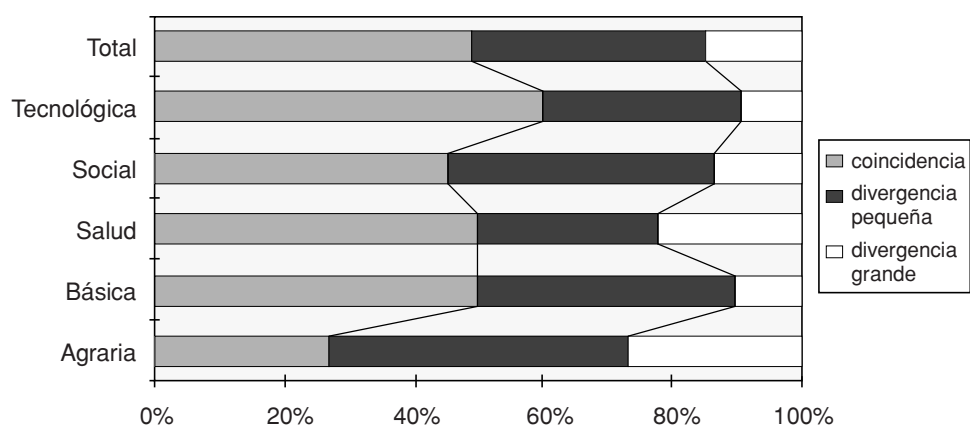


Figura 2. Distribución del total de proyectos en las diferentes categorías en el llamado de I+D de 1994



El área Agraria en 1992 y Salud en 1994 tuvieron el menor nivel de coincidencia, lo que era esperable de acuerdo con la hipótesis según la cual el grado de consenso entre colegas es menor en las áreas aplicadas.²⁵ Parece ser más “difícil” recibir opiniones iguales o similares de dos evaluadores en proyectos utilitarios, de los cuales se espera resultados de impacto a nivel social o productivo. Sin embargo, el área Tecnológica no se comporta así: en ambos llamados, sobrepasó el 50% de coincidencia, siendo la mayor de todas las áreas. Estas diferencias en el grado de divergencia entre áreas de conocimiento pueden ser apreciadas en las figuras 3 y 4.

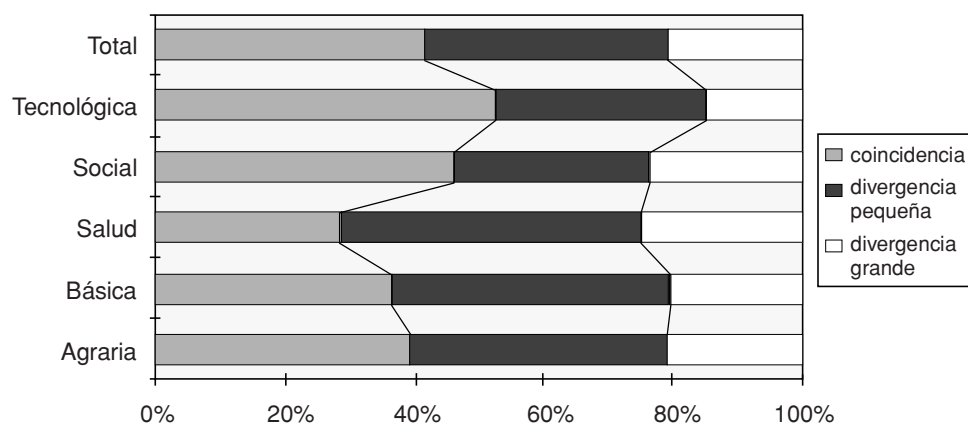
Figura 3. Distribución de categorías de acuerdo con el área de conocimiento en el llamado de I+D de 1992



Para apreciar el impacto que estas diferencias en la opinión de los evaluadores puede tener en el destino de un proyecto diseñamos otro ejercicio. Utilizamos el puntaje final dado por cada evaluador para construir dos *rankings* diferentes de proyectos para cada llamado: de 1 a 157 para 1992 y de 1 a 137 para 1994, en orden descendente de prioridad. Supusimos que el puntaje dado por los evaluadores deter-

²⁵ En los trabajos que han analizado las diferencias entre áreas se destacan dos tendencias respecto del grado decreciente de consenso: de las áreas básicas a las aplicadas y de las áreas “du-ras” a las “blandas”; véase, por ejemplo, Price (1969), Lindsey (1978), Abt (1992), Prpic (1994).

Figura 4. Distribución de categorías de acuerdo con el área de conocimiento en el llamado de I+D de 1994



minaba directamente el financiamiento²⁶ y que había fondos suficientes para aprobar la mitad de los proyectos en cada llamado.²⁷ La idea fue encontrar “reversiones”, o sea: propuestas que serían financiadas de acuerdo con la opinión de un evaluador pero no con la del otro.²⁸

En 1992, la línea de corte estaría en el proyecto 78 de acuerdo con el evaluador 1. De éstos, 46 serían también financiados por el evaluador 2, o sea, 59%; los restantes 41% no tendrían financiamiento. En 1994, la línea de corte estaría en la propuesta 68, resultando en 40 proyectos también financiados por el evaluador 2 (59%), mientras que 41% no tendrían éxito.²⁹ Esta tasa de reversión de más de 40% significa que el trabajo de evaluación de dos pares disminuyó en un quinto la posibilidad de que la propuesta fuese financiada al azar, o sea, el

²⁶ Éste no fue el caso de CSIC, como vimos anteriormente.

²⁷ En realidad, en el llamado de 1992 sólo el 35% de los proyectos fue financiado. El porcentaje correspondiente para 1994 fue 54%.

²⁸ Este término (del inglés *reversals*) fue acuñado por Cole *et al.* (1981) en un ejercicio similar desarrollado con datos de la National Science Foundation, USA.

²⁹ Puede esperarse cerca de 50% de reversiones si la ubicación de una propuesta por debajo o por encima de la línea de corte se determina arrojando una moneda (tendencia para grandes números). En el otro extremo, una coincidencia exacta de los dos *rankings* resultaría en una tasa de reversión de 0%. De este modo, las reversiones pueden ocurrir entre 0 y 50%.

azar sería responsable de un 80% del éxito de un proyecto.³⁰ Esto está muy cerca de la tasa de reversión esperada si el segundo “evaluador” fuera una moneda. Sea como sea, a fin de disminuir la influencia del azar a la mitad, sería necesario multiplicar por 8 el número de evaluadores utilizados por CSIC (de 2 a 14-17). Dado el tamaño de la comunidad científica local, el hecho de que los proyectos están escritos en español –y por lo tanto la comunidad científica internacional capaz de dar opinión está relativamente reducida– y la carga humana y financiera implicada en tal procedimiento, parece que el incremento en el número de evaluadores es una imposibilidad para la CSIC. De cualquier forma, ¿es deseable? Con el mayor esfuerzo extra, la existencia del azar podría ser reducido de 80% a 50%. ¿Cómo juega la “excelencia” de una propuesta dentro de este proceso de decisión?

Estos datos y preguntas fueron discutidos con miembros de las subcomisiones. Primero se preguntó acerca de las posibles razones de las grandes divergencias entre las opiniones de los dos evaluadores. Algunas respuestas implicaron incredulidad, tanto en “nuestras cuentas” como en aspectos internos del puntaje, ya que sostenían que la evaluación de ciertos aspectos puede dar sólo un resultado. Éste es el argumento principal:

Puede haber discrepancias en las evaluaciones a ese nivel, dentro de determinados aspectos de la evaluación, pero no en otros. Si vos decís: la bibliografía está al día, la idea es original, las metodologías que se plantean son racionales, en ese tipo de cosas no puede haber variación, *salvo que los evaluadores no se hayan tomado realmente la cosa en serio*. Donde sí puede haber diferencias es en si es aplicable, o en si esta cosa es transferible, o esta cosa se puede hacer en el plazo que se puede hacer, o se dispone de los medios para poder hacerlo, ahí puede ser que haya discrepancia. Pero no en los otros. Entonces yo no creo que alcance con medir la diferencia del puntaje final, sino ver en qué parte de esos puntajes se diferencia [...]. Creo yo que los ítems en los que puede haber diferencias, si los evaluadores trabajaron bien, son todos aquellos que empiezan a rayar con la subjetividad, pero no con la objetividad de la evaluación de un proyecto (las cursivas son nuestras).

³⁰ Cole *et al.* (1981, p. 882) encontraron tasas de reversión entre 24 y 30% en varias disciplinas. En este caso, el primer *ranking* fue construido con los puntajes recibidos por los 4 o 5 evaluadores seleccionados por la NSF para cada proyecto; el segundo *ranking* derivó de los puntajes atribuidos por otros 10 o 12 evaluadores, elegidos por los autores, para las mismas propuestas. A pesar de tal elevado número de evaluaciones, los autores concluyeron que el azar determina cerca de la mitad de las posibilidades de éxito de un proyecto.

La existencia de evaluadores “buenos” o “malos” fue la réplica favorita para las divergencias, ya que: “es muy difícil que de un proyecto horrible alguien diga que es muy bueno, eso es difícil. Ahí lo que es claramente malo es el evaluador”. Esto último podía llegar, ocasionalmente, a la “extrema falta de seriedad”, permitiendo que las relaciones personales con “amigos o enemigos” interfiriesen en la evaluación:

A mí me dio la impresión de que en algunos proyectos, de alguna manera, no sé cómo, hubo alguna influencia del evaluado en la selección del evaluador. Tuve esa sensación, después conocí a la gente y pensé: ¡qué casualidad! éste es amigo de este otro... Pasó con uno que era tan desproporcionadamente elogioso, yo no sabía la relación, después entré a averiguar; elogiaba tanto, extraordinario y pensé: ¿qué pasa acá?... Empezamos a escarbar... ¡inclusive comentaba cosas que no estaban en el proyecto! ¿y éste cómo sabe? los elogios eran sobre información que no surgía del proyecto.

A mí me ha pasado. Ya sé que si proyectos nuestros caen en manos de determinada gente, tienen menos probabilidades que en otras. Y no es gente con la que competimos. Te voy a decir algo para mostrar una cosa que la gente no piensa. Nosotros somos bastante exitosos como grupo [...] eso genera celos profesionales. xx me odia. Simplemente porque yo logré hacer una cosa con éxito [...] Entonces, proyecto mío que llega a las manos de xx, le baja la caña; y yo lo he podido verificar, aunque él no sabe. Ha cometido torpezas, como poner datos en las evaluaciones que sólo él podía tener. Dos más dos son cuatro. [...] Así como hay de un lado hay del otro. No quiere decir que todo el mundo es hijo de la madre, pero existen esas cosas [y] pasa absolutamente en todos lados [...] es un problema claro de la evaluación por pares, porque uno asume que los pares son perfectos, y no es verdad, los pares son como cualquiera de nosotros.

Algunos evaluadores, sin embargo, no son malos o llevados por intereses personales, serían sólo “más o menos” rigurosos en cuanto a lo que esperan de un buen proyecto o simplemente expresarían diferente la misma apreciación. En sus propias palabras:

Y en otros [evaluadores] te dabas cuenta de que había evaluaciones que eran muy bajas, que tenía la escala de 1 a 5 mucho más abajo que el de al lado; entonces vos mirabas los proyectos y veías que uno te impresionaba mejor, que el evaluador había sido muy duro en una cosa muy objetiva como era la formación previa o el equipo de investigación.

Yo recuerdo evaluadores que los números eran completamente distintos, pero cuando leías lo que decían, eran cosas muy parecidas. Eran diferentes criterios de poner números que tenían, pero estaban

diciendo casi lo mismo. Uno era muy severo al poner los números y el otro era muy condescendiente. Cuando escriben dicen las críticas reales que tiene el trabajo. [...] Resulta que a veces hay diferencias muy grandes [en los números] y en el fondo las diferencias son menores.

Así, las divergencias de opinión son atribuidas a “errores” humanos: a la incapacidad de los evaluadores de hacer su trabajo seriamente, a su disposición a favorecer a un amigo o castigar a un enemigo; a su exceso o falta de rigor. La excelencia está ahí, en los proyectos y en los proponentes: los evaluadores no son capaces (o no quieren) verla adecuadamente. En consecuencia, no hay nada que cambiar en el sistema, sólo cambiar los evaluadores —o averiguar sus razones y hacer los descuentos—. Apenas algunos entrevistados reconocieron que la noción de lo que es excelente en un proyecto puede diferir genuinamente de una persona a otra.³¹ La idea fue así expresada:

Cada evaluador le da pesos distintos a distintas cosas. Eso es lógico, y es así. [...] Si hay cinco evaluadores que evaluaron, cada uno, 3 proyectos, y vos te fijás, los 3 proyectos evaluados por el mismo evaluador tienen una cierta coherencia; pero cuando comparás estos 3 con estos otros 3, ahí falta coherencia, porque el peso relativo que le da cada evaluador a las distintas cosas, es diferente. Y además es sano que así sea, porque cada uno le presta más atención a determinadas cosas, entonces en el conjunto tú te podés formar una idea de cuáles son los puntos críticos que hay que tomar en cuenta para rankearlos a todos juntos (las cursivas son nuestras).

El hecho de que las divergencias fueran mayores en las áreas Agraria y Salud no encontró demasiada explicación en los entrevistados, más allá de la idea de que esto era esperable en las áreas más aplicadas. Sin embargo, un resultado no esperado —las relativamente bajas divergencias en las ciencias sociales comparadas con las básicas— llamó la atención de varios entrevistados, que afirmaron:

Yo creo que nosotros, los cientistas sociales, somos mejores evaluadores que los cientistas duros. [...] Porque los cientistas sociales tenemos una estructuración del campo tan baja, que tenemos que sistemáticamente estar construyendo una intersubjetividad; en cambio, los científicos duros dicen: publicó en tal revista, esto es tantos puntos; no

³¹ Algunos autores se han centrado en la diferente demanda —o nivel de severidad— de cada árbitro, o sea, cómo cada persona traslada su opinión a un número o categoría, o a “diferencias de opinión reales y legítimas entre especialistas respecto de lo que es o debería ser buena ciencia” (Cole *et al.*, 1981, p. 885).

se preocupan por el contenido; en tanto que evaluadores, sistemáticamente están delegando las evaluaciones a otros. Yo creo que por ahí tiene que haber una explicación fuerte.

Yo hubiera esperado mucha discrepancia en Sociales, porque son más opinables [...] Yo creo cada día más que no es posible tener leyes causa-efecto en la sociedad, entonces todo es mucho más opinable, sería mucho más lógico que ahí haya más discrepancias. En las ciencias básicas pienso que las discrepancias muchas veces surgen por el embanderamiento de la gente con corrientes. Por ejemplo: el famoso tema de los zoólogos... los biólogos modernos, por decir de alguna forma, sienten cierta repugnancia innata contra los zoólogos descriptivos. Entonces, sin argumentos objetivos “patentes”, tienden a “sacarlos del medio”. [...] Esas cosas pasan, entonces vos tenés gente que tiene puesta la bandera de la biología molecular y entiende que por eso es como Peñarol y Nacional, si soy de Peñarol estoy en contra de Nacional. Eso pasa mucho. [...] Se legitiman con cosas académicas. [...] En la Argentina hay, en biología molecular, dos grupos que tienen una competencia terrible entre ellos. Tú le mandás a evaluar cosas, olvidándote de eso ... ¿pero qué pasa? cuando le mandás un proyecto que está más en el área de lo que hace el competidor, están predispuestos a mandarlo al diablo. [...] Realmente hay escuelas también, aunque se trate de no decir, porque no tienen, como en las ciencias sociales, un fundamento legitimado. Acá son cosas impresentables, pero están. [...] Acá queda mal decir que hay corrientes, pero hay. Hay varias estrategias distintas de encarar problemas, y que compiten entre ellas y que se pelean cuando publican, entonces también se pelean cuando evalúan.

En suma, las divergencias encontradas en las opiniones de los pares parecen ser muy importantes para ser justificadas apenas como pequeñas diferencias en el juicio respecto de algo que existe por sí mismo: la calidad de la propuesta de investigación. Más aún: las divergencias no son mayores en todas las áreas aplicadas, como era esperado en base a la literatura. Lo que tanto los datos como las entrevistas parecen indicar es que la excelencia es percibida de maneras diferentes por el evaluador aislado –el par– que, en el proceso de dar su opinión sobre una propuesta, tiene como referencia sólo su propio repertorio intelectual –éste, claramente construido socialmente–.

5.2. Las discrepancias entre excelencia y relevancia en el proceso de evaluación de proyectos de vinculación con el sector productivo

La subcomisión de este programa busca distinguir dos atributos que los proyectos deben tener: excelencia y relevancia. Para ello se-

lecciona dos tipos diferentes de evaluadores: un par estricto y un experto en análisis tecno-económico, los cuales deben responder a una serie de preguntas y ofrecer un fallo o conclusión sobre la recomendación —o no— de la propuesta. El formulario de evaluación no incluye puntajes y hay diferencias en las preguntas a los dos evaluadores: ambos deben opinar sobre la metodología y la adecuación de los recursos solicitados, pero sólo el evaluador tecno-económico es consultado acerca de la importancia del problema como condicionante para el desarrollo socioeconómico y del impacto del proyecto en la producción.³²

Utilizamos los resultados del llamado a proyectos de 1993,³³ un total de 43 proyectos que tuvieron las dos evaluaciones. En este caso las propuestas no fueron clasificadas por áreas de conocimiento.³⁴ Buscamos discrepancias en la opinión de los dos evaluadores en los fallos y en el contenido de las evaluaciones. Clasificamos los fallos en tres grupos: a) recomendado, b) no recomendado y c) recomendado con modificaciones, de los cuales derivamos categorías con diferentes grados de discrepancia:

- categoría 1 -*coincidencia*: misma opinión en ambas evaluaciones;
- categoría 2 -*discrepancia pequeña*: una evaluación recomienda “con modificaciones”, la otra apenas recomienda o no recomienda;
- categoría 3 -*discrepancia total*: una evaluación implica recomendación y la otra no.

La distribución de los proyectos en las categorías se observa en la Figura 5.

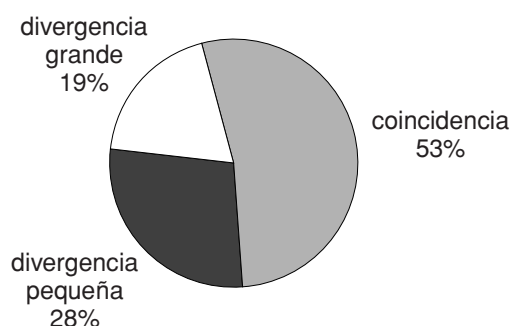
Las discrepancias máximas entre las dos evaluaciones (cerca de 20%), y por lo tanto en los dos conceptos —excelencia y relevancia—, no sobrepasan las divergencias encontradas entre dos pares que califican el mismo atributo —excelencia—. La coincidencia entre evaluadores sobre los aspectos puramente académicos de una propuesta de I+D parece ser tan frecuente como la convergencia de evaluaciones sobre dos atributos diferentes —uno interno a la lógica de la ciencia y otro externo al medio académico—.

³² Otros aspectos a ser calificados por el par se refieren al interés académico del problema, a la capacidad del equipo y su posible fortalecimiento, al aprovechamiento del conocimiento creado y a las posibilidades de transferirlo al sector productivo. Aquí ya hay importantes diferencias en relación con el formulario usado para los proyectos de I+D, a pesar de estar dirigido al mismo tipo de evaluador.

³³ Se decidió usar sólo este llamado debido a dos razones: el correspondiente a 1992 fue la primera experiencia y el número de proyectos presentados fue muy pequeño.

³⁴ Las características del llamado mismo hacen esta clasificación inapropiada.

**Figura 5. Distribución de las discrepancias entre evaluadores
en el total de proyectos del llamado SP de 1993**



Por otra parte, en nuestro análisis del contenido de las evaluaciones, percibimos que las respuestas de los evaluadores no muchas veces siguieron el camino solicitado en los formularios. En varias oportunidades fue difícil distinguir el texto del par de aquél del evaluador tecno-económico, a no ser por la indicación que encabezaba la hoja. A veces, el par centraba su análisis en los aspectos de aplicabilidad, dando mucho menos atención a los académicos. Por su parte, el evaluador “profesional” también, en varios casos, desarrollaba su opinión académica respecto de los aspectos de calidad. Es posible decir, luego de este análisis, que los límites entre excelencia y relevancia son mucho más difusos en la práctica, en la evaluación de casos concretos, de lo que es pensado comúnmente.

En suma, parece que la separación entre excelencia y relevancia no es tan clara cuando se trata de evaluar proyectos tecnológicos. La introducción explícita de criterios de aplicabilidad y utilidad debería haber implicado diferencias mayores en la opinión de los evaluadores si cada uno se restringiera al atributo para el que fue consultado. La imposibilidad de determinar distinciones claras entre lo “interno” —excelencia— y lo “externo” —relevancia— de la actividad científica, entre los factores cognitivos y los sociales, parece ser la base de las dificultades de los evaluadores en concentrarse en uno solo de esos aspectos.

6. La toma de decisión final

El tercer momento de los procesos de evaluación coordinados por CSIC es la toma de decisión final en relación con los candidatos a ser financiados. El análisis desarrollado aquí está basado en una comparación entre las listas elaboradas a partir de las opiniones de los evaluadores y las listas finales de proyectos efectivamente financiados. El objetivo es explorar la extensión en la cual las subcomisiones usan o alteran, en sus decisiones, los *rankings* contruidos a partir de los puntajes y conclusiones de los evaluadores.

Para ello construimos listas de los proyectos de los llamados de I+D de 1992 y 1994, por área de conocimiento, basados en la media matemática de los puntajes adjudicados por cada par. El porcentaje de cambio –o reversión– surge de la comparación entre esta lista y el resultado final de CSIC: ¿cuántos proyectos efectivamente financiados no lo serían si la decisión fuera determinada por el *ranking* matemático surgido de los puntajes de los evaluadores? Si la coincidencia fuera total, el número sería cero, y sería cien si el cambio fuera absoluto. El Cuadro 1 presenta los resultados.

Cuadro 1. Diferencias entre el *ranking* de los pares y la lista final de CSIC en los llamados de I+D de 1992 y 1994 (en porcentajes)

Llamado	1992	1994
Área Agraria	20	0
Área Básica	14	21
Área Salud	11	13
Área Social	0	16
Área Tecnológica	44	18
Total	15	17

Los porcentajes totales de cambio efectuados por la subcomisión no difieren mucho de un llamado al siguiente, a pesar de que el mecanismo utilizado para llegar a la resolución fue diferente: solamente una subcomisión en 1992 y paneles para cada área en 1994. Más aún, las diferencias entre la subcomisión y los evaluadores fue com-

parable a aquellas divergencias existentes entre las opiniones de los propios pares.

Hubo diferencias significativas en las tasas de reversión de cada área de conocimiento: desde Social, que no cambia absolutamente nada del *ranking*, hasta Tecnológica, con 44% de reversión en 1992. En el llamado de 1994, las diferencias entre áreas fueron un poco menores, a pesar de la existencia de distintos paneles que discutieron por separado la elaboración de las listas. Llama la atención aquí el hecho de que el área Básica fue la que tuvo la mayor tasa de reversión, mientras que en el área Agraria (uno de los campos más aplicados) el panel siguió exactamente las recomendaciones de los pares.

Con los proyectos de vinculación con el sector productivo no es posible realizar un ejercicio similar. Sólo observamos que de los 29 proyectos finalmente financiados en 1993 (de los 43 con dos evaluaciones completas), dos tenían una recomendación negativa del par académico y uno un fallo negativo del evaluador profesional.

Con estos datos en mano entrevistamos a miembros de csic, de las subcomisiones y de los paneles de área de 1994, en un intento de entender cómo fueron tomadas las decisiones finales, cómo las opiniones de los evaluadores nutrían las decisiones y cuál era el papel jugado por la excelencia y la relevancia en el proceso. Los cambios en las recomendaciones de los evaluadores son vistos como necesarios en el marco del ajuste de una opinión aislada en una valoración colectiva:

En el llamado anterior, '92, yo participé en el final, ahí ya se miraban los proyectos, no sólo era el puntaje pelado. [...] Yo me acuerdo de haber trabajado y ajustado puntajes, en función de lo que leíamos en el proyecto. En el último llamado la cosa fue un poquito más organizada, y se crearon esas comisiones asesoras, que agarraron el paquete de proyectos [...] empezaron a mirar las evaluaciones con el proyecto. Ahí, yo participé en la de Salud, miramos los proyectos a fondo, confrontamos lo que decía el proyecto con las evaluaciones.

La comisión que está acá es la que va a mirar los proyectos y los juicios de los evaluadores. Nosotros ajustamos los juicios de los evaluadores, todas las veces. [...] vienen los evaluadores y decimos: no, a éste se le fue la mano de blando, y a éste se le fue la mano de duro.

Los aspectos a ser tomados en cuenta para este “trabajo de ajuste”, apuntados por los entrevistados, no son llamados “científicos”, sino situados en el ámbito contextual: es necesario que los tomadores de decisión conozcan la situación de los candidatos, de sus laboratorios, de sus equipos, de su formación previa; en síntesis, aspectos de

su historia y desarrollo que no están en los papeles. En palabras de los entrevistados:

[El par extranjero] no lo puede saber, porque *la historia social de la ciencia de cada país es distinta y el momento actual que está viviendo cada país es distinto*. [...] Entonces ellos no pueden evaluar igual que uno [...] y vos sí podés uniformizar, siempre que la parte científica esté bien, o sea, los tipos te dicen: de estos 3, éste es mejor que los otros dos científicamente, y eso es lo que a vos te sirve. Entonces comparo eso con estas otras cosas que yo sé, y que él no sabe. Y eso es lo que a mí me parece que puede hacer el panel. Uniformizar los criterios sobre la base no de lo científico puramente [...] en todas las agencias financiadoras no se incluyen sólo elementos puramente científicos, sino elementos como ése, de evaluar la formación del tipo; pero la formación se evalúa en un contexto social e histórico, no es la misma. [...] Para nosotros son personas que conocemos, sabemos su historia, sabemos el laboratorio en el que está; y eso es una ventaja comparativa de la pequeñez de Uruguay. Siempre que mantengamos la evaluación externa, o sea, que podamos creer en lo que nos dicen los evaluadores del aspecto científico, y nosotros poder hacer política en este tipo de cosas, en contextualizar la inversión, o sea, la apuesta (cursivas nuestras).

Así, los elementos “científicos” quedan reducidos a los asuntos que el par puede evaluar porque no conoce el contexto del investigador y del proyecto. Y ese conocimiento del contexto permite hacer política “en el buen sentido de la palabra”, según un entrevistado.³⁵

Los ajustes entre elementos “científicos” y “contextuales” parecen ser más aceptables en algunas áreas que en otras. Ésta por lo menos fue la explicación para el 0% de reversiones en el área Social en 1992.³⁶ Un entrevistado de esta área ni siquiera terminó de definir los problemas que podían surgir si la subcomisión cambiaba las recomendaciones de

³⁵ Aquí aparece el miedo de que elementos y consideraciones políticas puedan influir sobre los juicios científicos; el entrevistado aún crea una nueva categoría, aunque siempre separada de la “parte científica”: “Nosotros sabemos cómo son esos laboratorios donde van a trabajar los tipos, que estos otros no los conocen. [...] Vos podés uniformizar algunos criterios, que te permiten seleccionar entre los que por calidad científica están parejos. Ellos no pueden. Y es ése el tipo de cosas que sí influye, y en parte es política, pero es *política práctica*, no sé como decirle, es *pragmático*, son las cosas que tú conocés y el de afuera no puede conocer” (cursivas nuestras).

³⁶ En el llamado de 1994 sí ocurrieron cambios en el área Social (16%). Sin embargo, en este caso había mayor legitimidad para alterar las recomendaciones de los pares, ya que había un panel especial constituido por cientistas sociales, que tomó la decisión.

los pares: “Lo que pasa es que [...] tocar sociales puede ser un gran lío, terminamos...”. Un representante de otra área, aunque también miembro de la subcomisión en 1992, recordó que: “Sociales por temor a [...] que se metan las cuestiones políticas y los líos, las rencillas, tienen mucho miedo a eso. Es el área donde hay más de ese tipo de líos”.

En el área Tecnológica, por otro lado, la necesidad de contextualizar los proyectos fue una razón suficiente y legítima para seguir o no las recomendaciones de los evaluadores:

Yo estuve precisamente en la comisión tecnológica, que reordenó las cosas. El problema es que los evaluadores eran a veces muy especialistas, tengo un caso que es claro. La Facultad de Medicina, el Instituto de Higiene, hizo esfuerzos muy grandes para desarrollar algunas vacunas. Uno de los evaluadores era del Butantán. Entonces dijo, pensando en escala Butantán [...]: eso ya está sabido en Butantán. Entonces, un proyecto que iba a ser apoyo al desarrollo de una determinada disciplina, en otros lugares estaba tan desarrollado que esas técnicas eran conocidas. Entonces hubo que modificar eso, porque se dijo: en sueros antiofídicos sería bueno que se tuviera una técnica. Las cosas que se hacen en el Instituto de Higiene es bueno que se desarrollen aquí. Ahí se cambió la puntuación. [...] En tecnología, si pensamos que en lugar de importar se puede hacer, se puede utilizar para desarrollar un grupo de trabajo; el evaluador externo normalmente ve solamente el objeto a producir en lugar de ver la parte universitaria, tener un grupo entrenado en las técnicas. [...] En lo tecnológico es razonable que estemos intentando no sólo tener un producto sino tener un personal especializado. [...] Es necesario adaptarlo [al *ranking*] al Uruguay.

En suma, todas las divergencias entre pares —las diversas opiniones sobre los proyectos— desaparecen rápidamente en las reuniones de las subcomisiones o paneles. El consenso respecto de la excelencia es rápidamente alcanzado: en ese proceso de comparación, contextualización y negociación de criterios, los miembros de las subcomisiones llegan a acuerdos sobre metodologías adecuadas, temas convenientes en cada disciplina, investigadores que merecen fondos, propuestas que tienen calidad. Lo que tiene lugar dentro de las subcomisiones es, para el “mundo exterior” (incluyendo a los evaluados), una verdadera “caja negra”. Todo lo que surge de allí es un fallo: esta propuesta será financiada, ésta no; por lo tanto, la primera tiene calidad y la segunda no. Entonces, es posible decir que la calidad de las propuestas de investigación es construida localmente por los propios actores del proceso de evaluación; la excelencia no es un atributo que viene incluido en la propuesta; ella es definida de manera contingente en el momento de

la decisión. Y, como varían las contingencias en cada momento de la evaluación, varía también el concepto de excelencia.

6.1. La introducción explícita de la relevancia como criterio de evaluación

Consultamos también a los entrevistados su opinión sobre el rol de la relevancia de un proyecto en las decisiones de financiamiento. En otras palabras: ¿puede un proyecto relevante no muy bueno científicamente tener precedencia sobre uno no relevante y muy bueno científicamente? Las opiniones variaron significativamente; algunos rechazaron la idea:

Si hacíamos eso lo más probable es que favoreciéramos el desarrollo de mediocres, que subidos al carro de decir: esto es de interés nacional, propusieran proyectos que fueran una porquería científicamente, pero que podían tener financiación. Y el objetivo, más bien era otro; era hacer buena ciencia. [...] Durante muchos años, se nos metió el verso de que no se podía publicar, de que había que hacer investigación de interés nacional, de que lo que interesa acá no le interesa al norte, y qué sé yo, que era un cuento para no someterse a la evaluación de los pares. Y eso sigue subsistiendo en algunas áreas [...] A mí ese cuento me sigue sin gustar.

Otros, sin embargo, creen que, en el caso de los proyectos de Vinculación con el Sector Productivo, la relevancia debería ser más importante aún que la excelencia científica y, entonces, definir la prioridad de financiamiento. Es más: la opinión del evaluador tecno-económico no es tan importante, en estos casos, como la manifestación del sector productivo interesado en financiar la investigación (o simplemente en sus resultados):

Si alguien de afuera está interesado, un productor nacional, que trabaja en Uruguay, que está interesado en desarrollar algo, y hasta es capaz de aportar dinero para hacerlo, bueno, la universidad considera que [el proyecto] ya tiene un aval de “útil para el país”, de alguna manera, o útil para un sector del país, resuelve un problema tecnológico.

En realidad, aun la necesidad de una evaluación en esos casos es cuestionada, especialmente para los proyectos con contraparte empresarial:

Había una serie de proyectos donde había un interés del sector productivo, del mercado, y una cantidad de dinero que ese sector produc-

tivo ponía. Eso, más que una propuesta de investigación científico-tecnológica medianamente original, era una apuesta de políticas de la universidad hacia el apoyo de un sector productivo de un país que estaba medio muerto. Entonces, si se cumplía con todas las garantías, a ese proyecto no le podías decir nada, más que si aquello estaba medianamente bien planteado, o racionalmente bien planteado; la evaluación ya, prácticamente, no era ni necesaria. Nadie pone dinero en un proyecto si no tiene interés en ese proyecto.

La idea dominante, sin embargo, es que la excelencia científica y la relevancia social y económica tienen que estar juntas:

Cualquiera de los dos son descalificadoras. Si el tipo me dice que va a investigar la mejora genética de una producción animal que en Uruguay no existe y que no va a existir porque hay condiciones que la hacen inviable, por más que eso pueda generar un hermoso *paper*, yo no lo voy a financiar. [...] Por supuesto, yo creo que siempre habría que tratar de juntar las dos cosas, ¿no? Siempre que sea posible tratar de reunir lo básico, causal, con lo tecnológico. [...] lo hemos especificado en los llamados: es un mérito de los proyectos de Sector Productivo que estén integrados aspectos básicos.

Más allá de las diversas opiniones con respecto al rol de la relevancia en las decisiones, en la práctica la evaluación tecno-económica tuvo el mayor peso. Es a través de este diálogo con personas directamente involucradas en la producción que la subcomisión es capaz de legitimar sus decisiones: “[...] o sea [son] los profesionales que dicen: este tema es importante, éste no”.

A pesar de ello, no se han definido llamados específicos para temas considerados de relevancia. En otras palabras, la relevancia social y económica es buscada “espontáneamente” en las propuestas presentadas, y no a través de acciones específicas orientadas a ese fin. Evidencia de esto es el extremo cuidado que tiene la subcomisión en atender los temas específicos en proporción similar a su presencia en el total de proyectos. Sobre esto, un entrevistado afirmó:

Ahí termina como termina un poco [proyectos de] I+D. Que somos los reyes de los neoliberales. Nosotros decimos que no somos neoliberales, pero ajustamos la plata de acuerdo a la demanda del mercado, somos muy respetuosos del mercado en eso... Se respetan las proporciones de las diferentes áreas respecto de lo que se pide. [...] No hay grandes discusiones... Debería haberla, porque no hay razones para suponer que el nivel de excelencia de todas las áreas es igual y que el nivel de importancia nacional de todas las áreas es igual. ¡Oh! Casualidad que coincidiera exactamente con los porcentajes.

La respuesta pasiva a la demanda probablemente deriva de las dificultades de concretar las declaraciones de principios, tantas veces repetidas, sobre “prioridades nacionales”. Comenzando con el problema, no resuelto aún, de seleccionar propuestas entre diferentes áreas del conocimiento, los que toman las decisiones quedan sin alternativas.

Todavía queda otra etapa, que es cruzar los diferentes sectores, y ahí yo creo que le estamos tirando al bulto. O sea, si el área agronómica es más exigente o menos que la tecnológica, por ahora no hay nadie de un perfil tan generalista que pueda nivelar los criterios de evaluación. Y de repente lo que tiene que haber son decisiones políticas, decir: la CSIC va a dedicar el 50% al área agropecuaria, 30% al área industrial, etc. [...] Siempre serían decisiones políticas, entre un área y otra, no hay un nivelador técnico, salvo que sean cosas muy gruesas. [...] Yo creo que en otros países, a ese nivel las decisiones son básicamente políticas: el país está muy jugado a esto y las decisiones están en función... [de eso].

En síntesis, en este proceso de decisión específico, en el cual la relevancia es introducida explícitamente, otros actores, externos al medio académico, participan –directa o indirectamente– en la construcción de un nuevo concepto que puede llamarse “mérito”. Éste resulta de un proceso de negociación y, por lo tanto, está también construido localmente.

7. El papel de la evaluación de proyectos en la construcción del conocimiento científico y en la determinación de su excelencia

Una breve síntesis del proceso de evaluación coordinado por CSIC resalta varios elementos del papel de la evaluación en la construcción del conocimiento científico. En primer lugar, un factor clave en el proceso resulta ser la selección de los evaluadores; este paso comienza a decidir el destino del proyecto. Las dificultades para alcanzar un equilibrio entre las necesidades de “desconocimiento evaluador-evaluado” y “conocimiento evaluador-contexto” influyen en el resultado final. Así, en las opiniones inciden factores más allá del análisis de los componentes de investigación tradicionales –metodología, bibliografía, etcétera–.

La magnitud del desacuerdo entre pares, en todas las áreas,³⁷

³⁷ Como vimos, no fue posible encontrar más coincidencias en los juicios de árbitros de áreas duras que en los de disciplinas aplicadas, donde intervienen factores como relevancia y aplicabilidad. En este estudio, la hipótesis de mayor o menor consenso paradigmático no encuentra evidencia que la sostenga.

parece implicar que gran parte del fallo depende del evaluador —de sus conocimientos, de sus valores y de sus expectativas—. Si “los esfuerzos de los científicos son influidos por la cultura, por el entrenamiento especializado, por motivaciones diversas y por variadas habilidades intelectuales, además de intereses, valores, sesgos y prejuicios” (Chubin y Hackett, 1990, p. 8), sus opiniones sobre la excelencia de una propuesta de investigación también lo son. Las grandes divergencias entre pares respecto de una propuesta sólo pueden existir si esos elementos se combinan para dar forma al resultado. La objetividad del juicio de los pares es así puesta en cuestión: no parece haber nada especial en la forma en que se hace este juicio, nada que lo diferencie de otros que son realizados sobre las más diversas actividades y atributos humanos.

Cuando la decisión final sobre el financiamiento debe ser tomada por los miembros de las subcomisiones, en la práctica es difícil de separar la evaluación de la calidad o excelencia de una propuesta, de su pertinencia o relevancia. Así, las distinciones entre lo “interno” y lo “externo” de la actividad científica, entre los factores cognitivos y los sociales, son muy difíciles de definir y establecer en los procesos de evaluación, o simplemente no existen, siendo ellas mismas construcciones sociales.

El programa empírico del relativismo —PER— argumenta que la construcción del conocimiento pasa por una fase en la cual la multiplicidad de interpretaciones posibles desaparece a través de mecanismos de clausura, y emerge un consenso científico respecto de lo que es “verdad” (Pinch y Bijker, 1990, p. 410). No es nuestra intención aquí hacer una “traducción libre” ni una equivalencia de las etapas descritas por PER, pero la idea puede aplicarse a nuestro caso. En los procesos de evaluación, luego de un lento mecanismo de revisión por pares, en reuniones de los organismos o decisiones personales tomadas por los directores de programa emerge un consenso respecto de lo que es “excelencia”. Éste parece ser un proceso rápido y efectivo de clausura: sólo basta colocar “consensualmente” las propuestas en una lista y decidir (sobre la base del presupuesto disponible) dónde será la línea de corte, para determinar el futuro de investigadores y de líneas de investigación: reconocimiento y verdad al mismo tiempo.

Aun los componentes tradicionales de los proyectos y de los cv de los investigadores que parecen ser más “objetivos” —metodologías, publicaciones en revistas prestigiosas— son ellos mismos resultados de procesos de evaluación previos. Las publicaciones en revistas exigen un ritual de revisión que posiblemente no difiera mucho de aquéllos descritos en este trabajo. Se reciben propuestas —de textos—, se piden opiniones —de revisores—, se toman decisiones —por parte de editores.

Los artículos científicos disputan espacio —de papel— de la misma manera que los proyectos compiten por recursos —en dinero—. Quizás la diferencia principal es que para la comunidad científica de países periféricos como los nuestros es muy difícil participar en la construcción de la excelencia de las revistas más leídas. Sin embargo, tal vez sea posible influir concientemente en la construcción de la excelencia local.

La excelencia, entonces, es un valor construido de manera contingente, moldeado por los propios actores del proceso de producción de conocimientos en los diversos momentos de la evaluación, y que se modifica en el tiempo y en el espacio; ella no tiene existencia como valor absoluto y universal. El tantas veces mencionado “patrón internacional de excelencia”, inmutable de una situación a otra, de un contexto a otro, es en realidad relativo y contextualizado. Afirmar esto no implica un juicio de valor sobre la validez de este patrón; obviamente él existe y es quien guía actualmente la producción de conocimientos originales. Pero esa “excelencia internacional” es apenas una “excelencia local” contextualizada, construida socialmente por determinados actores, que se “universaliza” a través de una sucesión continua de procesos de evaluación.

Evaluadores y tomadores de decisión determinan, junto con la calidad de cada propuesta, la dirección de la ciencia. Ellos son quienes obtienen los consensos respecto de las líneas de investigación, los objetivos y las metodologías adecuadas. En fin, ellos constituyen el “conjunto central” (Pinch y Bijker, 1990, p. 410), el grupo relevante en la decisión del camino que las diferentes especialidades y la ciencia local en su conjunto deben seguir.

Intentamos presentar aquí una visión o perspectiva diferente respecto de la evaluación en ciencia. Los mecanismos de revisión parecen ser esenciales en la construcción de consensos así como otros procesos de negociación de las verdades científicas, de importancia capital en el desarrollo de la ciencia como institución y en su relación con la sociedad. De esta manera, la principal función del sistema de pares ya no sería aquella directamente observable —distribuir recursos escasos—, sino un engranaje clave en los propios mecanismos de formación de consensos en ciencia. Centrarnos en los procesos de decisión de financiamiento nos permite explicar, en parte, las transiciones de estados de desacuerdo a momentos de consenso.

Así, en la arena de negociación que vincula la actividad científica y los recursos financieros —o sea, el *locus* de la evaluación de la investigación— también se da forma a la dirección y al contenido de la cien-

cia. Los proyectos e investigadores financiados son capaces de alcanzar consensos con respecto a algunas “realidades” o “verdades” determinadas; las otras posibilidades no llegan a existir. El futuro de la ciencia es, en parte, definido en el proceso de toma de decisiones del financiamiento. □

Bibliografía

- Abrams, P. (1991), “The Predictive Ability of Peer Review of Grant Proposals: The Case of Ecology and the US National Science Foundation”, *Social Studies of Science*, 21, pp. 111-132.
- Abt, H. (1992), “Publication Practices in various Sciences”, *Scientometrics*, 24 (3), pp. 441-447.
- Argenti, G., Filgueira, C. y Sutz, J. (1987), *Ciencia y Tecnología: Un diagnóstico de oportunidades*, Montevideo, CIESU.
- Barreiro, A. y Velho, L. (1997), “The Uruguayan Basic Scientists Migration and their Academic Articulations Around the PEDECIBA”, *Science, Technology and Society*, 2, pp. 261-284.
- Boden, M. (1990), “Report to the Advisory Board for the Research Councils, From the Working Group on Peer Review”, Londres.
- Bonilla, M., Herrera, R., Rius, J. y Yacamán, M. (1995), “Sistema de evaluación por pares en los proyectos de investigación y de fortalecimiento de la infraestructura científica y tecnológica”, *Ciencia y Desarrollo*, 122, pp. 8-17.
- Bortagaray, I. y Sutz, J. (1996), *Una aproximación primaria al Sistema Nacional de Investigación en Uruguay*, Montevideo, Trilce.
- Burgueño, G. y Mujica, A. (1996), “Relacionamiento entre la universidad y los sectores productivos: una experiencia reciente”, en Albornoz, M., Kreimer, P. y Glavich, E. (comps.), *Ciencia y Sociedad en América Latina*, Buenos Aires, Universidad Nacional de Quilmes, pp. 221-231.
- Chubin, D. y Hackett, E. (1990), *Peerless Science. Peer Review and U.S. Science Policy*, Albany, University of New York Press.
- Cicchetti, D. (1982), “We have Met the Enemy and He is Us”, *The Behavioral and Brain Sciences* 5, p. 255.
- Cogan, D. (1986), *Scientific Research Grants Review 1977/8-1981/2: A Theoretical Review and Statistical Analysis*, Dublin, National Board for Science and Technology.
- Cole, S., Cole, J. y Simon, G. (1981), “Chance and Consensus in Peer Review”, *Science* 214, pp. 881-886.
- Cole, S., Rubin, L. y Cole, J. (1977), “Peer Review in the NSF and the Support of Science”, *Scientific American* 237 (4), pp. 34-41.
- Cozzens, S. (1990), “Options for the Future of Research Evaluation”, en Cozzens, S., Healey, P., Rip, A. y Ziman, J. (eds.), *The Research System in Transition*, Dordrecht, Kluwer Academic Publishers, pp. 281-294.

- Crane, D. (1967), "The Gatekeepers of Science: Some Factors Affecting the Selection of Articles for Scientific Journals", *The American Sociologist*, noviembre, pp. 195-201.
- Davyt, A. y Yarzabal, L. (1991), "Programas centrales de apoyo a la investigación. Un decidido impulso a la creación científica", *Gaceta Universitaria* (Universidad de la República), v (2), pp. 28-31.
- Fölster, S. (1995), "The perils of peer review in economics and other sciences", *Journal of Evolutionary Economics* 5, pp. 43-57.
- Gillespie, G., Chubin, D. y Kurzon, G. (1985), "Experience With NIH Peer Review: Researchers Cynicism and Desire for Change", *Science, Technology and Human Values* 10 (3), pp. 44-53.
- Hackett, E. (1987), "Funding and Academic Research in the Life Sciences: Results of an Exploratory Study", *Science and Technology Studies* 5 (3/4), pp. 134-147.
- Hensler, D. (1976), "Perceptions of the National Science Foundation Peer Review Process: A Report on a Survey of NSF Reviewers and Applicants", preparado para el *Committee on Peer Review, National Science Board*, y el *Committee on Science and Technology, US House of Representatives*, Washington, D.C., NSF, pp. 77-33.
- Herrera, A. (1973), "Social Determinants of Science Policy in Latin America", en Cooper, C. (ed.), *Science, Technology and Development*, Londres, Frank Cass, pp. 19-37.
- Knorr-Cetina, K. D. (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Oxford, Pergamon Press.
- Latour, B. y Woolgar, S. (1979), *Laboratory Life: The Social Construction of Scientific Facts*, Londres, Sage.
- Latour, B. (1983), "Give me a Laboratory and I will Raise the World", en Knorr-Cetina, K. y Mulkay, M., *Science Observed. Perspectives on the Social Study of Science*, Londres, Sage, pp. 141-170.
- Law, L. y Williams, R. (1982), "Putting Facts Together: A Study of Scientific Persuasion", *Social Studies of Science*, 12 (4), pp. 535-558.
- Lindsey, D. (1978), *The Scientific Publication System in Social Science*, San Francisco, Jossey-Bass Publishers.
- Loría, A. y Loría, E. (1996), "Reflexiones en torno a la revisión por pares en revistas científicas", *Ciencia y Desarrollo*, 127, pp. 63-73.
- Lynch, M. (1982), *Art and Artifacts in Laboratory Science*, Londres, Routledge and Kegan Paul.
- Mahoney, M. (1977), "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System", *Cognitive Therapy and Research* 1 (2), pp. 161-175.
- McCullough, J. (1989), "First Comprehensive Survey of NSF Applicants Focuses on Their Concerns About Proposal Review", *Science, Technology and Human Values* 14 (1), pp. 78-88.
- Merton, R. (1973 [1960]), "Recognition and Excellence: Instructive Ambigui-

- ties”, en Merton, R., *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pp. 419-438.
- Mujica, A. (1997), “El fomento a la investigación: encuentros y desencuentros entre las políticas y sus resultados”, *Educación Superior y Sociedad* (CRE-SALC/UNESCO), vol. 8, No. 2.
 - Nicoletti, L. (1985), “Participação da Comunidade Científica na Política de cyT: o CNPq”, Texto 16, Brasília, CNPq.
 - NIH Grants Peer Review Study Team (1976), “Grants Peer Review: Report to the Director, NIH Phase I”, Washington, D.C.
 - Peters, D. y Ceci, S. (1982), “Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again”, *The Behavioral and Brain Sciences* 5, pp. 187-205.
 - Pinch, T. y Bijker, W. (1990), “The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other”, *Social Studies of Science*, 14, pp. 399-441.
 - Price, D. J. S. (1969), “The Structures of Publication in Science and Technology”, en Gruber, W. y Marquis, D. (eds.), *Factors in the Transfer of Technology*, Cambridge Massachussets, MIT Press, pp. 91-104.
 - Prpic, K. (1994), “The Socio-cognitive Frameworks of Scientific Productivity”, *Scientometrics* 31 (3), pp. 293-310.
 - Rip, A. (1994), “The Republic of Science in the 1990s”, *Higher Education*, 28 (1), pp. 3-23.
 - Sábato, J. y Botana, N. (1968), “La ciencia y la tecnología en el desarrollo futuro de América Latina”, *Revista de la Integración*, 3.
 - Saráchaga, D. (1997), *Ciencia y tecnología en Uruguay: una agenda hacia el futuro*, Montevideo, CONICYT/Trilce.
 - Sirilli, G. y Meliciani, V. (1994), “Research Evaluation at the National Research Council in Italy: a survey of decision-makers”, *Research Evaluation* 4 (2), pp. 75-88.
 - Sonnert, G. (1995), “What Makes a Good Scientist?: Determinants of Peer Evaluation among Biologists”, *Social Studies of Science* 25, pp. 35-55.
 - Stolte-Heiskanen, V. (1986), “Evaluation of Scientific Performance on the Periphery”, *Science and Public Policy* 13 (2), pp. 83-88.
 - Varsavsky, O. (1969), *Ciencia, política y cientificismo*, Buenos Aires, Centro Editor de América Latina.
 - Weinberg, A. (1964), “Criteria for Scientific Choice II: the Two Cultures”, *Minerva*, III (1), pp. 3-14.
 - ——— (1963), “Criteria for Scientific Choice”, *Minerva*, I (2), pp. 159-171.
 - Zenzen, M. y Restivo, S. (1980), “The Mysterious Morphology of Immiscible Liquids: A Study of Scientific Practice”, *Social Science Information* 21 (3), pp. 447-473.
 - Ziman, J. (1994), *Prometheus Bound: Science in a dynamic steady state*, Cambridge, Cambridge University Press.
 - Zuckerman, H. y Merton, R. (1973 [1971]), “Institutionalized Patterns of Eva-

uation in Science”, en Merton, R., *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, pp. 460-496.

- Zuckerman, H. (1987), “Foreword”, en Jackson, D. y Rushton, J., *Scientific Excellence. Origins and Assesment*, California, Sage, pp. 7-11.