

0 1048 3

29.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE FILOSOFIA Y LETRAS
INSTITUTO DE INVESTIGACIONES FILOSOFICAS

La Teoría de la Prueba Estadística de Deborah Mayo y
Su Relación con la Filosofía de la Experimentación

Tesis que para optar por el título de: Maestro en Filosofía de la Ciencia
Presenta: Huang Xiang
Director de Tesis: Dr. Sergio Fernando Martínez Muñoz

México, D.F. 1998

TESIS CON
FALLA DE ORIGEN

267069



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS.

En el curso de mi formación y de esta tarea en particular he recibido la ayuda de varias personas que de una manera u otra colaboraron en la realización de este trabajo. Sin su apoyo la culminación de esta investigación no hubiera sido posible. A todos ellos mi más profundo agradecimiento.

Sin embargo, de manera especial quisiera agradecerle a las siguientes personas. Al Dr. Sergio Martínez Muñoz por haberme ayudado, a través de sus excelentes cursos y su enseñanza personal, a la formulación de esta investigación. Al Dr. Carlos López Beltrán por su disposición, sugerencias y críticas y, finalmente, les agradezco su paciente dedicación para superar los escollos de una lengua tan ajena a la mía. A la Dra. Edna Suárez, al Dr. Ambrosio Velazco y al Mtro. Mario Casanueva por sus útiles comentarios. A Inés y Alicia Pazos quienes siempre estuvieron dispuestas a tenderme a la mano a pesar de sus múltiples ocupaciones. A mi amigo Alejandro Alcalde por auxiliarme en mis problemas con esta lengua. Asimismo, tengo una deuda enorme con el Instituto de Investigaciones Filosóficas y la Facultad de Filosofía y Letras de la UNAM que me ofrecieron una sólida educación. Por último, sin el apoyo financiero de la Fundación UNAM esta labor no hubiera sido posible.

ÍNDICE

INTRODUCCIÓN: HACIA UNA FILOSOFÍA DE LA EXPERIMENTACIÓN.....	1
1. LA FILOSOFÍA DE LA EXPERIMENTACIÓN DE DEBORAH MAYO.....	11
1.1 Los tres propósitos de Mayo.....	11
1.2 Los modelos de las investigaciones experimentales.....	14
(1.2.1) Los modelos de los datos de Suppes.....	15
(1.2.2) El modelo primario.....	16
(1.2.3) El modelo experimental.....	17
(1.2.4) El modelo de los datos.....	19
(1.2.5) La jerarquía de los modelos experimentales.....	21
1.3 El uso de la estadística en los experimentos.....	22
2. LA LIMITACIÓN DEL MODELO PRIMARIO.....	33
2.1 "La experimentación tiene vida propia".....	33
2.2 Los experimentos no estadísticos.....	40
2.3 La tradición basada en "imágenes".....	46
3. LAS DIFICULTADES PARA MODELAR LA INFERENCIA CAUSAL.....	54
3.1 La estadística del error y los métodos para modelar inferencias causales.....	54
3.2 El problema de las causas de referencia.....	59
3.3 La explicación causal de Cartwright.....	63
3.4 La explicación causal de Lipton.....	71
4. EL CÁLCULO DE LA SEVERIDAD Y LA SELECCIÓN DEL NIVEL DE SIGNIFICACIÓN.....	76
4.1 Hacia una inferencia objetiva.....	76
4.2 El nivel de significación y el tamaño de las muestras.....	82
4.3 Selección del nivel de significación.....	86
CONCLUSIÓN.....	94
REFERENCIAS BIBLIOGRÁFICAS.....	100

Introducción: Hacia Una Filosofía de la Experimentación

Actualmente los estudios filosóficos sobre la experimentación se han convertido en un tema central en la filosofía de la ciencia. Esto es consecuencia del rechazo a dos principios del positivismo lógico. El primero es que los positivistas consideraban que los hechos observacionales sirven para controlar el desarrollo de las teorías. Pero, puesto que estos hechos observacionales son cosas seguras y no problemáticas, no se les presta ninguna atención a las prácticas experimentales en que se producen los enunciados de los hechos observacionales. El segundo consiste en que, de acuerdo a la estructura de la ciencia que los positivistas lógicos defienden, las teorías se relacionan con los hechos observacionales por medio de conceptos de explicación y confirmación que pueden caracterizarse en término de una relación lógica aceptada universalmente en todas las investigaciones científicas, de manera que la racionalidad científica puede tener una base segura basada en estos conceptos. Por ello, la realización de un experimento desde el punto de vista de la racionalidad científica, sólo puede servir para verificar o falsificar una teoría por medio de esta relación lógica.

Desde los años sesenta los anti-positivistas han criticado con vigor estos dos principios del positivismo. Según los anti-positivistas, los hechos observacionales no son neutrales y sin problemas, sino que están cargados de teoría; y por otro lado, el reconocimiento de los sociólogos e historiadores de la ciencia de que muchas investigaciones científicas simplemente no obedecen las reglas de racionalidad universal que los positivistas plantean. Eso implica que la relación entre las teorías y los experimentos no puede restringirse a una formulación meramente lógica, sino que existen procedimientos muy complejos que tienen lugar en las prácticas experimentales, procedimientos que van desde la obtención de los datos por medio del diseño experimental hasta la utilización de estos datos como evidencia para la validez de las hipótesis.

El rechazo de estos dos principios positivistas requiere que, a fin de entender filosóficamente la dinámica de las investigaciones científicas, los estudios de la filosofía de la ciencia deban enfocarse a las cuestiones sobre las prácticas científicas reales. Por ello, los estudios sobre los experimentos han tomado un lugar preeminente en la filosofía reciente de la ciencia.

Hasta hoy en día, han habido dos direcciones en los estudios sobre los problemas de la experimentación. Una es la de los estudios sociológicos e históricos, y la otra la del nuevo experimentalismo.¹ Los sociólogos han hecho muchos estudios de caso sobre las prácticas experimentales en que usan métodos como, por ejemplo, entrevistas con los científicos tanto teóricos como experimentales. Estos sociólogos elaboran la idea anti-positivista de que las observaciones están cargadas de teorías, recurriendo a la tesis de que los hechos científicos se constituyen

socialmente. Muchos sociólogos se han dedicado a los estudios sobre la construcción social de los hechos científicos (*social construction of scientific facts*). Bruno Latour y Steve Woolgar (1979), Donald MacKenzie (1981) y Andrew Pickering (1984) usan explícitamente la palabra “construcción” para el título o subtítulo de su libro. Otro tema central de estudios sociológicos es una consecuencia del primero: puesto que los hechos observacionales no se consideran una base segura y sin problema para la inferencia científica, se piensa que las controversias científicas se resuelven no con algún criterio objetivo, sino que son el resultado de negociaciones sociales. Es decir, lo que llamamos la racionalidad científica es sólo una cubierta ideológica para designar el proceso de toma de decisión que está determinado por circunstancias sociales e históricas. Por ejemplo, Harry Collins insiste en que una estricta replicación experimental no es alcanzable (1975 y 1985, pp.33-46) y los juicios de los científicos durante el procedimiento de los experimentos, en su caso el experimento sobre la radiación gravitatoria, no obedecen a razones puramente cognitivas sino son productos sociales (1981 y 1994). Trevor Pinch trata de mostrar, por medio de su estudio sobre el experimento de neutrino solar, que los científicos, para llegar a un consenso, negociando entre diferentes intereses sociales desde diferentes campos de estudios, y que estas negociaciones seguirán hasta que el problema de neutrino solar se resuelva (1980,1981 y 1985a,b).

Por su lado los nuevos experimentalistas comparten con los anti-positivistas y sociólogos la crítica hacia los dos principios positivistas anteriormente mencionados, pero consideran que los estudios sociológicos sobre las prácticas experimentales abren la puerta al escepticismo y al relativismo, y que por tanto no son plausibles. La relación entre teoría y observación, según los nuevos experimentalistas, no es simplemente una relación lógica como los positivistas plantean, y los experimentos tampoco están dominados por una relación de este tipo, sino que se constituyen por medio de una serie de actividades diferentes (tales como el diseño experimental, la calibración de instrumentos, la estimación correcta sobre los posibles resultados y posibles errores, etc.) de donde puede obtener los datos. Pero, el rechazo de estos dos principios de las prácticas experimentales no implican, como los sociólogos proponen, que no existen, *tout court*, buenas razones en la racionalidad científica, sino sólo negociaciones sociales. Cuando por primera vez introduce el concepto “nuevo experimentalismo”, Ackermann señala que la intención de este corriente es readaptar una intuición tradicional, que habían sido rechazada por los anti-positivistas y los sociólogos de la ciencia, la idea de que “experimental observation and theoretical conjecture should, somehow, have a symmetrical status with respect to scientific development, either being capable of producing a temporary fixed point onto which progress can be hinged.” (Ackermann,1989, pp.185-186) Es decir, los hechos observacionales obtenidos a partir de procedimientos experimentales

complejos imponen restricciones objetivas sobre las teorías científicas por medio de inferencias científicas. Hacking usa una metáfora para mostrar más claramente la diferencia entre los sociólogos de la ciencia y los nuevos experimentalistas. Al contestar la pregunta de que ¿es la ciencia invención o descubrimiento? los nuevos experimentalistas comparten la respuesta de R.B Braithwaite en su libro *Scientific Explanation*: la gente propone un sistema de hipótesis y la naturaleza dispone su verdad y falsedad. Pero para los sociólogos, la segunda parte de la respuesta sería: también es la gente quien dispone su verdad y falsedad. (Hacking,1988, p,150) Ahora, la tarea de los nuevos experimentalistas es mostrarnos cómo la naturaleza dispone la verdad y falsedad de los teorías que la gente propone sin caer en los supuestos típicos de la tradición positivista. Se requiere una filosofía de la experimentación cuyo propósito radica en explicar cómo las prácticas científicas pueden en cierto sentido legitimar los datos experimentales como una fuente de los hechos observacionales, y cuáles son y cómo se justifican los patrones de razonamiento objetivo que los científicos usan en la inferencia científica. Hasta hoy día, muchos filósofos se han dedicado a esta tarea, y han propuesto diferentes enfoques y diferentes maneras para fundamentar una filosofía de la experimentación. Ackermann señala que, dentro de los nuevos experimentalistas, existen dos corrientes que tratan de cumplir esta tarea:

If the notion of experimentation is to be central to a new philosophy of science, there are at least two possibilities to consider: that the notion of experimental legitimation could be exploited towards the end of providing a new experimental foundationalism, and that the notion of providing a dialectical account of scientific progress in which either theory or experiment could provide the temporary support and constraint for the tentative advance of the other given the concrete set of scientific practices available at a specific point in time. (Ackermann,1989,p,186)

La teoría de Allan Franklin representa la primera corriente. Franklin trata de mostrar que se pueden lograr conocimientos científicos por medio de la legitimación de los resultados experimentales, porque existe una epistemología de la experimentación. Esta provee a los científicos de creencias razonables sobre los procedimientos de la selección o confirmación de las diferentes teorías, y sobre las manipulaciones experimentales, lo que le permite distinguir los hechos de las construcciones espurias.² Esta epistemología incluye varias estrategias usadas por los científicos en las prácticas de laboratorio:

- (1) la comprobación y la calibración experimentales, en que los aparatos reproducen los fenómenos como los científicos han planteado;
- (2) la reproducción de los efectos espurios cuya presencia ha sido anticipada;

- (3) la intervención en que los experimentadores pueden manipular los objetos bajo estudio;
- (4) el uso de los diferentes experimentos o diferentes instrumentos para confirmaciones independientes;
- (5) la eliminación de posibles errores e hipótesis alternativas;
- (6) el uso de los propios resultados para apoyar la validez de su existencia, mostrando que es imposible que estos resultados sean construcciones espurias;
- (7) el uso de una teoría independientemente corroborada para explicar los resultados;
- (8) el uso de los aparatos que se basan sobre una teoría bien corroborada, y finalmente
- (9) el uso de los argumentos estadísticos. (Franklin, 1986, pp. 166-184; 1990, pp. 103-116)

Franklin no niega que los elementos sociales y causales pueden influir en la selección de teorías y en el juicio sobre los resultados, pero insiste en que, por medio de estas nueve estrategias, la práctica científica, aunque falible, en última instancia depende de las evidencias legitimadas por las buenas razones que estas nueve estrategias garantizan. Además, Franklin cree que estas nueve estrategias se pueden representar en términos de la teoría bayesiana sobre la inferencia científica. (Franklin, 1986, pp. 113-119; 1990, pp. 99-103) Este intento de Franklin de reducir todos los razonamientos experimentales a una teoría de la metodología estadística es distintiva de un nuevo funcionalismo experimental, que se ha convertido en una corriente del nuevo experimentalismo.

La segunda corriente es la que sostiene que la objetividad de la inferencia científica se da localmente por medio de una relación dialéctica entre las teorías y los experimentos en contextos específicos. Es decir, la evidencia sí impone restricciones sobre las inferencias científicas, pero no por medio de una relación puramente lógica, sino por los procedimientos específicos, históricamente constituidos y justificados, para obtener los datos y los conocimientos experimentales en experimentos específicos. Hacking y Galison toman esta dirección.

Hacking ha insistido en que el experimento no es simple una consecuencia de teorías, y que la relación entre teoría y experimento no se puede simplificar reduciéndola a una relación lógica general. Los experimentos y las teorías se conectan en las prácticas científicas por medio de procedimientos complicados y de maneras muy diversas. (Hacking, 1983, pp. 177-194) Una filosofía de la experimentación debe tomar en cuenta este aspecto importante. Pero la complejidad de la relación entre teoría y experimento no significa que la evidencia no genera restricciones objetivas sobre las teorías. Por un lado, Hacking señala que la observación que obtienen los científicos de los experimentos no es un conjunto de enunciados de la experiencia, sino que depende crucialmente de una habilidad, pues incluye las intervenciones de estos científicos como parte de prácticas aprendidas (Hacking, 1983, pp. 195-210). Por otro lado, mediante un estudio sobre el uso de los microscopios en

las investigaciones científicas, Hacking muestra que los científicos legitiman la objetividad de lo que ven en los microscopios por medio de estas intervenciones.

En "The Self-Vindication of the Laboratory Sciences"(1992), Hacking expresa sistemáticamente la idea de que la legitimación de la objetividad de la evidencia sólo se da localmente por medio de las relaciones dialécticas entre las teorías, los instrumentos y las manipulaciones. Hacking toma las ciencias de laboratorio como el objeto de estudio y lo clasifica en tres áreas: (1) las ideas, que incluyen las teorías globales y locales, teorías en el contexto de fondo, teorías para los instrumentos; (2) las cosas (*things*) que consisten en todas las sustancias materiales que los científicos usan; y finalmente (3) las marcas (*marks*) y las manipulaciones de las marcas, tales como la generación de los datos, los análisis estadísticos de los datos y las interpretaciones sobre estos datos. Hacking señala que estas tres áreas se ajustan mutuamente y se constituyen en sistema holista de autojustificación (*self-vindication*), en caso de que los datos producidos por los instrumentos dentro de este sistema se puedan explicar coherentemente con las teorías en el mismo sistema. Por ejemplo, la mecánica newtoniana es un sistema holista de este tipo, en que sus propios aparatos producen los fenómenos que se justifican por la teoría mecánica newtoniana como verdadera sólo adentro de este sistema. Para Hacking, el progreso de la ciencia depende de la formulación, el desarrollo y la mutación de muchos sistemas de autojustificación de este tipo. La formulación de estos sistemas pueden ser socialmente construidos. Pero, una vez que se ha formulado un sistema, o dicho en la terminología de Hacking, se ha constituido una ciencia madura y estable (Hacking, 1992, p.55), las evidencias producidas adentro de este sistema tendrá las restricciones objetivas sobre las hipótesis bajo prueba adentro del mismo sistema. Como consecuencia, la legitimación de la objetividad de evidencia, para Hacking, no es garantizada por medio de algunas metodologías o reglas aceptadas universalmente, sino que es local y relativamente dada dentro de las relaciones dialécticas entre las ideas, las cosas y las manipulaciones de las marcas.

Peter Galison (1983) ha señalado que son incorrectas, tanto la idea de los positivistas de que observaciones necesariamente están controladas por teorías, como la idea de los sociólogos de que las observaciones están necesariamente cargadas de teorías. De acuerdo a los positivistas, los experimentos no son problemáticos para la filosofía de la ciencia, porque un experimento termina simplemente cuando han habido suficientes observaciones relevantes para ciertas teorías. Bajo la segunda idea, los experimentos son buenos ejemplos del procedimiento de la construcción social de conocimiento científico, porque un experimento termina cuando los científicos han llegado a un consenso por medio de negociaciones entre intereses sociales. Según Galison, ambas ideas olvidan un

tema crucial para la filosofía de la ciencia: la manera cómo se construyen y se legitiman los argumentos que los científicos usan para convertir las experiencias en los hechos científicos.

Los científicos sí tienen buenas razones para separar los hechos reales de las construcciones espurias. Pero esto no es por medio de la relación lógica entre teoría y observación, sino que por medio de las manipulaciones y las intervenciones de los científicos. Por ejemplo, en el experimento de la atracción gravitatoria, Henry Cavendish, a fin de evitar la influencia de la variación de temperatura, diseñó cuidadosamente los aparatos dentro de una habitación cerrada, controló el procedimiento experimental por control remoto y observó el resultado por un telescopio. Para conocer si el campo magnético distorsionaría el resultado, Cavendish sometió sus aparatos a un campo magnético artificial y midió la diferencia que estos aparatos señalaron. Además, los cálculos cuidadosos también sirvieron a Cavendish para eliminar ciertas construcciones espurias. (Galison, 1983, pp. 2-4)

En su libro "How Experiments End", a través de los estudios históricos sobre los experimentos de alta-energía realizados en la segunda parte de este siglo, Galison trata de describir el procedimiento por medio del cual las evidencias experimentales se hacen convincentes. Galison clasifica dos tipos de instrumentos: los productores de imágenes y los contadores de partículas. (Galison, 1983, pp. 248-249) Y esta diferencia está asociada con diferentes tipos de argumentación. Los experimentos que usan el primer tipo de instrumento se terminan generalmente cuando aparecen los "sucesos nítidos" (*golden events*). Es decir, cuando se obtienen imágenes técnicamente bien hechas y de sucesos cuya existencia ha sido conjeturada por ciertas teorías. Por otro lado los experimentos que usan el segundo tipo de instrumento se terminan cuando los científicos, usando generalmente análisis estadísticos sobre los datos, llegan a la decisión de que los resultados son suficientemente significativos para apoyar una conclusión.³ El que un experimento se termine, no significa que su resultado sea conclusivo. Pero sí se puede terminar por buenas razones, aunque estas razones no son reconocidas universalmente, sino son diferentes en distintos experimentos específicos.

Por todo ello, Galison describe las relaciones entre teorías, experimentos y hechos científicos en un experimento como una dialéctica que difiere tanto de las versiones positivistas como de las de los sociólogos. Según Galison, las actividades experimentales están limitadas por muchas restricciones (*constraints*) teóricas y experimentales, las cuales ayudan a los científicos a llegar a adoptar creencias y acciones razonables. Estas restricciones no son aceptadas y usadas universalmente; pertenecen más bien a diferentes niveles y son aceptadas y usadas según las diferentes etapas de la experimentación. Estas etapas son: la del diseño experimental, la de generación de los datos, la de análisis de los datos y la de la interpretación de los resultados. Bajo

estas restricciones, los resultados experimentales, o los hechos científicos, pueden ser estables en ciertas circunstancias. Es decir, cuando las teorías cambian, los resultados pueden mantenerse sin cambiar su validez bajo ciertas restricciones. Eso significa que la experimentación tiene cierto nivel de autonomía que es importante para el desarrollo de la ciencia. (Galison, 1983, pp. 244-255)

Deborah Mayo ha desarrollado un nuevo proyecto para fundamentar la filosofía de la experimentación. Ella trata de hacerlo por medio de encontrar las reglas metodológicas que guían las actividades científicas. Mayo está de acuerdo con la idea de que las teorías no están directamente conectada con los experimentos por una relación puramente lógica, sino que esta relación es muy complicada y tiene que ver con las prácticas reales. Por tanto, para poder estudiar esta relación, Mayo propone un enfoque “por partes”, que es un método que requiere que se clasifique el procedimiento de las investigaciones científicas en varias partes locales, y que se estudien los problemas específicos en cada una de éstas. Ella considera que las investigaciones se pueden clasificar en los siguientes pasos: (1) determinar las hipótesis locales a partir de las teorías globales; (2) diseñar un experimento para probar las hipótesis locales; y (3) analizar los resultados a fin de conseguir un juicio sobre la aceptación o el rechazo de las hipótesis a prueba. Mayo refina este enfoque al proponer una jerarquía de los modelos que representan los procedimientos experimentales.

Mayo se une a quienes sostienen que los científicos aceptan o rechazan una hipótesis en un experimento no por causas o intereses sociales, sino por razones bien justificadas. Ella cree que los nuevos experimentalistas han hecho excelentes estudios para comprender cómo es que los científicos pueden, por medio de las manipulaciones y intervenciones, tener los resultados experimentales convincentes. Pero en su opinión no han logrado un estudio sistemático aceptable sobre los razonamientos que usan los científicos. Por eso, ella insiste en que una filosofía de la experimentación debe plantear una epistemología sistemática del experimento que sea capaz de explicar los razonamientos usados y los conocimientos conseguidos en los experimentos. Mayo propone un proyecto que se puede entender de dos maneras: una intuitiva y otra formal. La manera intuitiva es la que ella llama argumentación a partir de los errores: la idea básica es que los científicos saben que los errores no ocurren sólo cuando realizan procedimientos de investigación, y que son muy posibles detectar esos errores, pero no pueden encontrarlos. Es por eso que realizan estos procedimientos de investigación guiados por diferentes reglas metodológicas, en diferentes pasos experimentales representados en una jerarquía de modelos que Mayo propone. La manera formal consiste en afirmar que estas reglas metodológicas son representables por medio de la teoría de probabilidad no-bayesiana, específicamente por la estadística diseñada por Neyman y Pearson para modelar los experimentos, y que es llamada por Mayo la “estadística del error”. Mayo cree que esta

estadística del error permite un criterio objetivo para explicar sistemáticamente las buenas razones que los científicos usan en los experimentos, y para resolver los problemas sobre evidencia y inferencia científica.

Recordemos la distinción de Ackermann entre las dos corrientes del nuevo experimentalismo: la de fundamentalismo de Franklin y la de interpretación dialéctica defendida por Hacking y Galison. El proyecto de Mayo es un intento del primer tipo, porque intenta fundamentar la objetividad de la inferencia científica en la inferencia estadística, por medio de la reducción de todas las reglas metodológicas usadas en los experimentos a la estadística del error. Como hemos visto, el proyecto de Franklin es un intento muy similar, si bien él interpreta la inferencia experimental de una manera más directa, y acepta la teoría bayesiana para representar la base razonable de la ciencia.

En esta tesis, haré un estudio sobre la plausibilidad del proyecto de Mayo en la filosofía de la experimentación. Estoy convencido de que este proyecto ha mostrado una base muy importante para la filosofía de la experimentación, a saber, el uso de la argumentación estadística en los experimentos. Este uso de la argumentación estadística sirve como un instrumento muy importante en la inferencia científica. Pero lo que no me parece plausible es la intención de reducir todos los aspectos epistemológicamente significativos de la experimentación a la estadística del error. Me dedicaré sobre todo a mostrar las razones que hacen implausible esta tesis de Mayo. Trataré de mostrar, por un lado, que en las prácticas científicas, hay muchos experimentos que no son reducibles a lo estadístico. Esto implica que el proyecto de Mayo no es suficiente para fundamentar una filosofía de la experimentación, porque ha dejado un gran parte de los otros tipos de experimentos sin explicar. Por otro lado, mostraré que la inferencia estadística también tiene su limitación, y por tanto, tampoco es suficiente para garantizar la objetividad de toda la inferencia científica, como Mayo pretende. A través de estas críticas sobre el proyecto de Deborah Mayo, voy a defender la idea de que la dirección fundamentalista que Mayo y Franklin adoptar, no es una dirección correcta para hacer la filosofía de la experimentación, puesto que las prácticas científicas son mucho más complicadas de lo que tal tipo de fundamentación puede ofrecer.

A.C Crombie (1994) señala que en la tradición científica occidental, la estadística es sólo un estilo de argumentación entre los otros que los científicos usan en las investigaciones científicas. A través de un estudio histórico, Crombie clasifica seis tipos de estilo de argumentación en la historia de la ciencia europea: (1) la postulación de las ciencias matemáticas axiomáticas; (2) la investigación experimental y la medición; (3) el modelo hipotético; (4) la comparación taxonómica; (5) la estadística sobre población, y finalmente (6) la descripción histórica y genética. (Crombie,1994,pp84-85) Estos seis estilos son totalmente independientes y aparecieron en diferentes

épocas bajo diferentes circunstancias culturales. Según Crombie, no hay una jerarquía de la importancia entre estos seis estilos. Los científicos los consideran como un depósito de los diferentes instrumentos para la investigación y los usan en las prácticas científicas dependiente de las situaciones específicas.

Por supuesto, la clasificación de Crombie no es la única manera de clasificar los razonamientos científicos. Crombie la hace sólo bajo un punto de vista histórico y antropológico, sometiendo las actividades científicas al contexto de toda la cultural occidental. Si tomamos otras perspectivas, no es imposible que tengamos otro tipo de clasificación. Sin embargo, Crombie y los filósofos de la experimentación de la segunda corriente están de acuerdo en que, en cualquier tipo de clasificación, no es plausible encontrar sólo una argumentación definitiva que sirva para explicar todas las prácticas científicas. Además, en las prácticas científicas, la utilización de un estilo de argumentación no se realiza necesariamente por medio de una manera fija y unívoca. Los científicos pueden usar el mismo estilo de argumentación de diferentes maneras en diferentes circunstancias.

Si tomamos en serio esta diversidad y complejidad de los diferentes estilos de argumentación, me parece que la primera corriente de la filosofía de la experimentación debe ser rechazada. Tanto la estadística bayesiana defendida por Franklin como la no-bayesiana defendida por Mayo, sirven nada más como un estilo de argumentación entre otros usados en las investigaciones científicas, y no pueden considerarse como una fundamentación para todos los campos científicos. A lo largo de la tesis, a través de una crítica del proyecto de Mayo voy a substanciar esta idea.

En el primer capítulo, haré una breve reseña de la filosofía de la experimentación de Mayo. En el capítulo 2, mostraré que muchos experimentos en la historia de la ciencia no usan la argumentación de la estadística de error. Por tanto, el proyecto de Mayo no es suficiente para una filosofía de la experimentación. En los capítulos 3 y 4, trataré de mostrar que la estadística del error no es suficiente para garantizar una objetividad en las inferencias científicas como Mayo plantea. En la conclusión, analizaré por qué la primera corriente en la filosofía de la experimentación defendida por Mayo y Franklin no es plausible; y usando el concepto de Sergio Martínez de estructura heurística, sugeriré posibles maneras de enfrentar las críticas en capítulos anteriores que permiten hacer una filosofía de la experimentación cercana a la segunda corriente. Estas sugerencias son muy esquemáticas pero espero desarrollarlas más a fondo en mi tesis de doctorado.

Notas

1. Sobre los sociólogos me refiero a tres escuelas principales: la escuela de Edinburgo encabezada por B. Barnes y D. Bloor, se enfocan principalmente a estudiar la ciencia como práctica social, en que los conocimientos científicos se formulan dentro de un grupo por medio de la determinación de las relaciones complejas de los "intereses" sociales de

este grupo; la escuela de Bath presentada por H. Collins y T. Pinch, entre otros, quienes se enfocan, a través de los estudios de los casos específicos, a las negociaciones sociales que juegan el papel decisivo en la formulación de los consensos entre los científicos ; y la corriente encabezada por B. Latour, S. Woolgar y M. Callon, la que considera que las actividades científicas son muy similares a los conflictos políticos. Sobre los historiadores me refiero a los quienes estudian las influencias sociales en los acontecimientos en la historia de la ciencia, por ejemplo, S. Shapin y S. Schaffer sobre la controversia entre Boyle y Hobbes, T. Porter y D. Mackenzie sobre el nacimiento de la idea estadística, etc. La fórmula “nuevo experimentalismo” fue usada primera vez por Ackermann (1989) para referir a los estudios de Hacking (1983), Galison (1983), Franklin (1986) y él mismo (1985).

2. Franklin no pretende elaborar directamente una teoría de la racionalidad científica, sino mostrar la plausibilidad de que la ciencia es razonable (*reasonable*). Dice: “...I will use the term reasonable, rather than rational, in this discussion because I do not have a complete theory of rationality. I do believe, however, that when you observe evidence entailed by a hypothesis it is rational to strengthen your belief in that hypothesis.” (Franklin, 1990, p.99, nota, 2)

3. En su libro de 1997, Galison ha estudiado en mayor grado y en una manera más detallada los conflictos y las cooperaciones entre los dos tipos de tradiciones durante el desarrollo de la física moderna. En el capítulo 2, analizaré una vez más la significación filosófica de esta distinción entre dos tipos de tradición basada en dos diferentes tipos de instrumento.

Capítulo 1 La Filosofía de la Experimentación de Deborah Mayo

1.1 Los tres propósitos de Mayo

Es sabido que para la fundamentación de la estadística se han desarrollado en tres escuelas principales: el bayesianismo, el conductismo de Neyman-Pearson-Wald y finalmente el enfoque de Fisher. Los últimos dos comparten un rechazo a la metodología bayesiana, pero están en profundo desacuerdo sobre la naturaleza del criterio apropiado para decidir sobre la plausibilidad de una inferencia estadística. (Hacking, 1965, pp, 80-3 y 92-111; Mayo, pp, 379-84) La polémica ha generado mucho escepticismo en la filosofía de la ciencia respecto a la estadística no bayesiana. Además, han habido otras objeciones importantes a la estadística no bayesiana. Los sociólogos y los psicólogos se han quejado del abuso de la prueba de Neyman-Pearson-Wald cometido en los estudios de la ciencia social y casi unánimemente han tenido una actitud negativa hacia las pretendidas interpretaciones objetivistas del método estadístico. (Morrison y Henkel, 1970, pp, v-xiii) Esto ha contribuido a que el enfoque bayesiano haya sido aceptado por los filósofos de la ciencia de manera casi unánime. Un primer propósito de Mayo es hacer ver que hay un núcleo importante de ideas en la tradición no bayesiana que es importante recuperar. Ella arguye que en la práctica científica la metodología de la estadística del error no sólo es muy útil sino es el núcleo del razonamiento científico, y que el desprecio de esta metodología por los filósofos y científicos sociales constituye una actitud muy criticable. Por otro lado, Mayo elabora su propuesta no bayesiana en término de un enfoque metodológico que ella llama "enfoque por partes" (*piecemeal approach*). Un enfoque es "por partes" en la medida en que el método requiere la separación del proceso por medio del cual aprendemos de los errores experimentales en partes que deben de ser tratados diferencialmente. El modelo de Mayo de la estadística del error pretende dar respuesta a tres temas centrales e íntimamente ligados en la filosofía de la ciencia: (1) modelar el proceso de la inferencia científica, a saber, el proceso de conseguir los datos, el de hacer las inferencias y tomar las decisiones, y el de confirmar o probar las hipótesis; (2) resolver los problemas de la objetividad de la observación, la infradeterminación y la tesis de Duhem; y (3) proveer un criterio para seleccionar los principios metodológicos que nos permiten juzgar la racionalidad de las investigaciones científicas. (Mayo, p, xiii y 70) De lograrse esto, la estadística no-bayesiana evitaría el problema de la determinación de la probabilidad anterior (*prior probability*) que ha sido fuente de serias críticas al bayesianismo. Para el bayesianismo la evidencia e confirma la hipótesis H para el sujeto S en la medida en que el grado de creencia de S sobre H con la presencia de e es mayor que si e no se toma en cuenta, esto es $(Pr(H/e) > Pr(H))$. El

grado de creencia de un sujeto se representa por la asignación de probabilidad subjetiva. Llamamos probabilidad anterior al grado de creencia de un sujeto sobre la base de H cuando no se toma en cuenta e , y cuando se toma en cuenta e la llamamos probabilidad posterior. El teorema de Bayes como resultado matemático es impecable, sin embargo, uno de los inconvenientes que Mayo ve en la interpretación bayesiana tradicional es que la asignación de la probabilidad anterior depende puramente de la creencia personal, y que los científicos en la misma situación pueden tener creencias muy diferentes. En caso extremo, algunos científicos pueden creer que esta situación es irrelevante y asignarle una probabilidad cero, con lo que el teorema bayesiano simplemente no funciona. Mayo insiste en que el proceso de asignación de una probabilidad anterior introduce elementos subjetivos y arbitrarios, de manera que no puede proveernos un modelo satisfactorio de la inferencia inductiva (Mayo, p,x)¹.

El segundo propósito de la obra de Mayo, radica en reconstruir la metodología racional por medio de la justificación de reglas objetivas para la discriminación entre hipótesis. Algunos empiristas lógicos buscaron reglas formalizables en una lógica inductiva, o en las teorías de confirmación, mientras que los popperianos intentaron fundamentar una metodología falsacionista sobre una lógica puramente deductiva. Los post-positivistas, después del ataque encabezado por Kuhn en los años 60s, dudaron seriamente de la existencia de estas reglas objetivas y universales. Kuhn y Feyerabend señalaron que la historia de la ciencia no parecía reflejar la existencia de las reglas metodológicas propuestas tanto por los positivistas lógicos como los falsacionistas. Para Kuhn, la elección entre diferentes teorías no siempre dependía de la lógica, ni de las experiencias neutrales representadas por enunciados básicos, ni de las reglas metodológicas, sino de los criterios propios adentro de un paradigma. Por otro lado, los psicólogos como P. Wason, P.N. Johnson-Laird, D. Kahneman y A. Tversky, entre otros, hicieron unas series de experimentos que muestran que el razonamiento humano no se rige por normas, lógicas de la manera simplista que se ha asumido. Esto es un resultado ha sido apoyado adicionalmente por estudios en sociología de la ciencia que exploran las maneras en las que el conocimiento científico se construye socialmente. Todos estos estudios cuestionan severamente dos principios centrales de la idea clásica de filosofía de la ciencia, a saber, la existencia de los fundamentos del conocimiento y la existencia de un método algorítmico para justificar la verdad de los enunciados. Sin embargo, esto no significa que la investigación científica sea una actividad puramente relativista que no responde a reglas objetivas. Muchos filósofos han tratado de encontrar las reglas objetivas en las metodologías científicas. Larry Laudan y Philip Kitcher son dos representantes bien conocidos de este tipo de empresa. Recientemente, Laudan propuso su proyecto del “naturalismo normativo”, en que la justificación de los reglas metodológicas

en las investigaciones científicas depende de la efectividad de estas reglas a fin de obtener ciertas metas cognitivas de los científicos. Por medio de esta manera, Laudan cree que su proyecto puede, con un enfoque naturalista, evitar los problemas del positivismo, y al mismo tiempo, mantener la normatividad de las actividades científicas. (Laudan,1996,chap.7&9) Kitcher está de acuerdo con el naturalismo y trata de describir el desarrollo científico colocándolo en contexto social. Mientras tanto, propuso una metodología eliminativista para defender la objetividad de la racionalidad científica. Según esta metodología, las evidencias positivas y los éxitos en predicciones no necesariamente confirman las teorías, sino que lo hacen sólo cuando pueden eliminar las posibles teorías alternativas. Kitcher intenta mostrar que en la historia de la ciencia, especialmente, en la revolución química y el desarrollo de la teoría de evolución biológica, los conocimientos científicos se avanzan por medio de esta metodología en la competencia de teorías alternativas. (Kitcher,1993,chap.7) Mayo figura entre ellos al declarar:

The challenge, if one is not to abandon the view that science is characterized by rational methods of hypothesis appraisal, is either to develop more adequate models of inductive inference, or else to find some new account of scientific rationality...I was led to pursue the first option---attempt to develop a more adequate account of inductive inference. (Mayo,p.ix)

Mayo estudia los usos de la estadística en los experimentos científicos y sostiene que una relación objetiva entre la evidencia y las hipótesis puede garantizarse por medio de una teoría de la inferencia estadística. Pero esta garantía no se da directamente sino sólo a través de un estudio por partes de los experimentos. Más adelante explicaré esta idea.

Un tercer propósito de Mayo es el de dar un fundamento filosófico a la experimentación. Mayo está de acuerdo con la idea central de los llamados nuevos experimentalistas: la experimentación tiene vida propia (Hacking, 1983, p.160). Ella entiende este lema de tres maneras:

There are three main senses in which the life of experiment may be independent of theories and theorizing, and each corresponds to an important theme brought out by the New Experimentalist work. First...asserts that the aims of experimental inquiry may be quite independent of testing, confirming or filling out some theory. Instead, actual experimental inquires focus on manifold local tasks: checking instruments, ruling out extraneous factors, getting accuracy estimates, distinguishing real effect from artifact, and estimating the effects of backgrounds factors. The second reading of the slogan asserts that experimental data may be justified indepently of theory, that experimental evidence need not be theory-laden in any way that invalidates its role in grounding

experimental arguments...A third reading of the slogan asserts that experimental knowledge may be retained despite changes of theory...This suggests that experimental knowledge may serve not only in adjudicating between rival theories, but also as a basis for progress in science. (Mayo,1994,pp,270-271)

Esto es, nos obliga a reconocer a que la teoría y el experimento tienen diferentes relaciones en diferentes ciencias y en diferentes estadios de desarrollo. Hasta hoy en día, la mayoría de estos nuevos experimentalistas han hecho muchos trabajos sobre los casos específicos y locales: la construcción y la verificación de los instrumentos, la manipulación y la técnica de los experimentadores, la estimación, la distinción entre el efecto real y el artefacto, etc., pero, Mayo piensa correctamente, me parece, que no hay un estudio sistemático que permita entender cómo la teoría y el experimento se relacionan en diferentes estadios de desarrollo, y por lo tanto no hay un estudio sistemático que formule de manera clara qué quiere decir la frase de Hacking de que la experimentación tiene vida propia. La vía apropiada para desarrollar esa requerida fundamentación del nuevo experimentalismo es, según Mayo, el estudio a fondo de la metodología estadística, tal y como se utiliza en la ciencia, este es un elemento básico para todo estudio sistemático de la experimentación que los nuevos experimentalistas simplemente han ignorado. Mayo afirma:

Their experimental narratives offer a rich source from which to extricate how reliable data are obtained and used to learn about experimental processes. Still, nothing like a systematic program has been laid out by which to accomplish this. This task requires getting at the structure of experimental activities and at the epistemological rationale for inferences based on such activities. To my mind, the reason the New Experimentalists have come up short is that the aspects of experiment that have the most to offer in developing such tools are still largely untapped. These aspects cover the designing, modeling, and analyzing of experiments, activities that receive structure by means of statistical methods and arguments. (Mayo,p,58)

Según Mayo, la estadística del error puede basarse en la teoría de Neyman-Pearson que tal y como Mayo la reformula permite que estos tres propósitos se apoyen mutuamente.

Vamos ahora a ver en detalle cuál es la estructura del modelo sistemático que Mayo propone para estudiar la experimentación científica.

1.2 Los modelos de las investigaciones experimentales

Como ya lo mencionamos, una idea central del nuevo experimentalismo consiste en articular y defender la idea de que la experimentación tiene vida propia, una idea que se opone al modelo clásico de la filosofía de la ciencia que sostiene que el experimento está siempre subordinado a la teoría. Para los experimentalistas, los experimentos pueden tener otras metas además de probar, confirmar o falsificar las teorías; los datos o las evidencias experimentales pueden justificarse sin recurrir de manera esencial a la noción de “carga teórica”, y el conocimiento experimental puede mantenerse frente al cambio radical de las teorías. (Mayo, 1994, pp,270-271) El nuevo modelo de la experimentación debe ser capaz de representar estas relaciones complejas entre las teoría y los experimentos. Mayo trata de construir los modelos de las investigaciones experimentales inspirándose en los modelos de los datos de Patrick Suppes. Por ello principio con una breve introducción a esta idea de Suppes.

1.2.1 Los modelos de los datos de Suppes

Patrick Suppes critica sin tapujos en su artículo “Model of Data” (1969) la simplificación de la estructura de la ciencia por parte de los positivistas lógicos, y plantea una jerarquía de diferentes modelos que relacionan las teorías con los datos. Dice Suppes:

Philosophers who write about the representation of scientific theories as logical calculi then go on to say that a theory is given empirical meaning by providing interpretations or coordinating definitions for some of the primitive or defined terms of the calculus. What I have attempted to argue is that a whole hierarchy of models stands between the model of the basic theory and the complete experimental experience. Moreover, for each level of hierarchy, there is a theory in its own right. Theory at one level is given empirical meaning by making formal connections with theory at a lower level. (Suppes, 1969, p.34)

Aunque Suppes comparte un verificacionismo fuerte con el positivismo lógico (la idea de que la fuente empírica de la evidencia es la base para la justificación del significado de los enunciados teóricos), ya tiende a formular este problema en una dirección más cercana a las prácticas científicas. Tomando el ejemplo de la teoría de aprendizaje (*learning theory*), Suppes clasifica jerárquicamente cinco teorías de modelos, cada uno de las cuales requiere de una teoría aparte para explicar la relación entre las teorías y los datos.

La primera es la teoría de modelos de respuesta lineal (*linear response models*), cuya tarea principal es decidir el nivel de significación para el análisis de los datos en los pasos subsecuentes y

definir las condiciones bajo las cuales la hipótesis está confirmada o desconfirmada. La segunda es la teoría de los modelos de los experimentos. La tarea en esta etapa consiste en decidir cuántas veces se requiere repetir el experimento, y en escoger los parámetros que pueden someterse a prueba. La tercera es la teoría los modelos de los datos, en los cuales los científicos comparan el resultado obtenido por medio del cálculo estadístico sobre los datos, conseguidos por medio del experimento, con el nivel de significación a fin de rechazar o aceptar la hipótesis nula. La cuarta es la teoría del diseño experimental, y la quinta la teoría de las condiciones *ceteris paribus*. Las tres primeras teorías, como Suppes muestra, pueden sumergirse formalmente en la teoría de conjuntos formulada por Suppes, y según él, el modelo de los datos es el núcleo de la jerarquía ya que es donde los científicos deciden aceptar o rechazar una hipótesis. Así afirma Suppes:

A model of the data is designed to incorporate all the information about the experiment which can be used in statistical tests of the adequacy of the theory.”
(Suppes, 1969, p.31)²

Según Mayo, el hecho que la jerarquía de los modelos deja ver que hay diferentes problemas que tienen que resolverse en cada uno de los modelos sugiere el tipo de estudio por partes de los problemas de la inferencia estadística. Mayo no intenta formular una teoría formal para reducir estas relaciones entre los modelos a una teoría única como pretende Suppes, sino que acepta *prima facie* esta jerarquía como punto de partida para modelar la estructura de los experimentos. Ahora continuamos con la presentación de la propuesta de Mayo.

1.2.2 El modelo primario

El primer modelo de la jerarquía es a lo que Mayo llama modelo primario. Mayo describe el funcionamiento de este modelo en los siguientes términos:

A substantive scientific inquiry is to be broken down into one or more local or “topical” hypotheses, each of which corresponds to a distinct primary question o primary problem. In a comprehensive research situation, several different piecemeal tests of hypotheses are likely to correspond to a single primary question. Typically, primary problems take the form of estimating quantities of a model or theory, or of testing hypothesized values of these quantities. They may also concern the form or equation linking theoretical variables and quantities. (Mayo, p.129)

Así, este modelo busca localizar una hipótesis local que pueda someterse a prueba. Esto es una idea compartida por la mayoría de los filósofos de la ciencia: una teoría global no puede ser

probada directamente, sino por medio de las hipótesis deducidas de esta teoría con condiciones iniciales y enunciados *ceteris paribus*. Estas hipótesis locales permiten elaborar un juicio acerca de las hipótesis que pueden someterse a prueba. En el modelo de Mayo, la decisión que decide el límite para mantener o rechazar una hipótesis, se lleva a cabo por medio de un análisis de la distribución estadística implicada por las hipótesis primarias. Además, el modelo primario permite decidir el alcance de una hipótesis que según Mayo depende del siguiente paso.

A manera de ilustración consideramos el experimento de J. Perrin sobre el movimiento browniano. El propósito de Perrin es el de probar la teoría de Einstein-Smoluchowski según la cual el movimiento browniano es causado por las moléculas en movimiento, teoría que se identifica como la cuestión primaria o el problema primario (*primary question* o *primary problem*). Un modelo primario tiene que llevarnos a determinar una hipótesis local que pueda sostenerse por medio de una prueba. La hipótesis local en este caso es la siguiente: si el movimiento browniano es causado por los movimientos de las moléculas en el medio, entonces los desplazamientos totales de las partículas brownianas tendrán una media que concuerda con la distribución de Gauss, o sea, la distribución normal para dos parámetros: la media que en este caso es 0, y la variación que en este caso es $2Dt$. Esta hipótesis en el modelo primario según Mayo se expresa de este modo:

La hipótesis primaria H : S_t , el desplazamiento de una partícula browniana en un tiempo t , obedece la distribución normal (Gaussiana) con $\mu=0$, y la variación= $2Dt$.³

1.2.3 El modelo experimental

El modelo experimental tiene dos funciones principales. La primera consiste en establecer una relación con el modelo primario. En el modelo experimental, las hipótesis o los problemas del modelo primario se reducen a problemas concretos, tales como ¿qué tipo de experimentación puede ser efectivo? ¿cuántas repeticiones se necesitan para que haya una prueba confiable? ¿qué tipo de resultado, generalmente estadístico, puede ser esperado por los experimentadores? Si un problema primario tiene como propósito la prueba de una hipótesis, la tarea del modelo experimental consiste en caracterizar, por medio del planteamiento de las hipótesis auxiliares requeridas por un experimento, los posibles resultados, en caso de que la hipótesis primaria sea correcta. Durante este proceso, las características principales de un experimento se especifican para los siguientes pasos, aunque muchas consideraciones todavía se manejan de manera muy idealizada, por ejemplo, el gas que se asume es un gas ideal, el movimiento sin fricción, o los controles perfectos, etc.

La segunda función del modelo experimental consiste en establecer una relación con el modelo de los datos por medio de la cual se especifican las técnicas de análisis de los datos. Aquí es donde entran los conceptos estadísticos. Aunque los objetos de las investigaciones científicas son los hechos en el mundo, los científicos generalmente no los tratan directamente sino sólo a través de un muestreo de los datos, y de la estadística que nos permite analizar esas muestras. Según Mayo, la relación entre el modelo experimental y el modelo de los datos generalmente se formula estadísticamente, si bien la hipótesis primaria no es necesariamente una hipótesis estadística. Hay dos maneras de formular esta relación estadística. Primera, la misma predicción experimental puede formularse como una hipótesis estadística que permite un margen de error. Segunda, la predicción experimental se considera fija y se clasifican los datos de acuerdo con la predicción, siempre y cuando la distribución de los datos no esté muy lejos del margen de error.

El interés de Mayo sobre la prueba estadística es en particular en la prueba de Neyman-Pearson o la prueba de significación (*the significance test*) como medio para realizar esta reducción de las hipótesis al modelo experimental. Esto lo voy a abordar en detalle en una sección posterior, puesto que se trata de un tema muy importante del proyecto de Mayo. Aquí me limitaré a mencionar que esta prueba permite a los experimentadores tomar dos tipos de decisiones. Una es respecto al nivel de significación, que es una estimación previa, acerca del valor de los parámetros estadísticos calculados por un experimento; y el otro tipo de decisión, que depende de la primera decisión, permite estimar el número de los errores posibles. Mayo cree que los dos tipos de decisiones anteriores son suficientes para establecer la confiabilidad o la severidad de una prueba.

En el caso del experimento de Perrin, la hipótesis primaria arriba mencionada tiene que concretarse en una hipótesis estadística realizable y manejable en un experimento. Esta concretización puede expresarse como sigue:

La hipótesis del modelo experimental H: La distribución de desplazamiento es de una población distribuida de acuerdo a una distribución Gaussiana M en donde el valor del parámetro a es una función de N^* .

Aquí, N^* es el número de Avogadro asumido por la teoría cinética. Puesto que D en la hipótesis primaria es una función del número de Avogadro N^* , éste también puede ser calculado cuando se estima D .

Esta hipótesis del modelo experimental implica dos cosas, una sobre la distribución M y otra sobre el parámetro N^* , de manera que se requieren dos pasos del experimento y por lo tanto dos pruebas. Aquí sólo voy a mencionar el primero, el presupuesto de que el proceso obedece la distribución de Gauss.

Para especificar este paso, transcribo la primera parte de la hipótesis H:

j: Los datos del desplazamiento de las partículas de polen en el experimento E aproximan una muestra azarosa de el proceso normalizado M.

Como requiere la prueba de Neyman-Pearson, la hipótesis se considera como la hipótesis nula, y la hipótesis contraria se formula de la siguiente manera:

j': Los datos de desplazamiento de las partículas de polen en el experimento E tienen efectos sistemáticos (i.e con una tendencia a desplazarse en una determinada dirección) y no azarosos.

La teoría de Einstein y Smoluchowski requiere que j sea correcta o j' sea refutada. Si son factores aparte del medio líquido la causa del movimiento, en cuyo caso j' es correcta, las partículas de polen van a tender a moverse aproximadamente en una dirección determinada. Si una partícula de polen se mueve al azar con respecto al movimiento de las otras partículas, la hipótesis j se sostiene. En el experimento, Perrin calculó la desviación de la distribución de 500 desplazamientos de las partículas de polen con la distribución normal estadística de ji-cuadrado (χ^2), y llegó a la conclusión que, si esta desviación era menor que el nivel significativo entonces j queda aceptada, y que, si es mayor que el nivel significativo, entonces j' queda aceptada. En términos del modelo experimental, lo que Perrin hizo fue especificar las veces que el experimento debía repetirse, escoger la estadística experimental apropiada (el tipo de la distribución normal) y decidir el nivel de significación.

1.2.4 El modelo de los datos

En un experimento, los científicos manejan y manipulan los datos seleccionados a partir de materiales no procesados previamente. Precisamente la tarea del modelo de los datos consiste en la generación de los datos, o sea, en procesar la materia prima en una forma canónica requerida para la aplicación de las herramientas analíticas o estadísticas del modelo experimental. La hipótesis nula en el nivel del modelo experimental conlleva presupuestos acerca de lo que deben ser los datos para el experimento y requiere que se satisfagan estos presupuestos experimentales en el nivel del modelo de los datos. Generalmente hay dos maneras de satisfacerlos. Una es diseñar un experimento plenamente de acuerdo con los presupuestos experimentales. La otra es encontrar los datos de manera que puedan eliminarse las violaciones de los requerimientos que se le han impuesto al modelo experimental. La primera utiliza teorías del diseño experimental y la segunda teorías de los datos. Durante ambos procesos, los experimentadores tienen que averiguar si los datos concuerdan o no con los presupuestos experimentales antes y después de los experimentos. Mayo considera que esta es

una cuestión secundaria en un modelo de la experimentación, y piensa que generalmente la prueba sobre este asunto se hace con los mismos tipos de metodología estadística, de manera que toda la jeraquía del modelo de la experimentación se reduce a una cuestión estadística. Describe ella:

A statistical significance test may also be used to check that the data satisfy the assumptions of the experimental model. The null hypothesis would now assert that a given assumption is met by data, and the data would now be the actual, not the condensed, data set...[T]he statistical significance tests for checking data-model assumptions are part of the battery of standard statistical methods...[T]he ability and need to test the experimental-model assumptions are an important part of why standard statistical methods are often referred to as "objective". (Mayo,pp,160-1)⁴

What should be included in data model? The overriding constraint is the need for data models that permit the statistical assessment of fit(between prediction and actual data)... (Mayo,p,135)

Veamos cuál es la tarea de Perrin con respecto al modelo de los datos. Habiendo establecido los presupuestos experimentales al nivel del modelo experimental, o sea, aquellos presupuesto que se requieren para poner a prueba la hipótesis nula y la contraria, y para estimar el nivel de significación, le queda a Perrin la cuestión de cómo encontrar los datos que satisfagan estos presupuestos, y esto es el trabajo del modelo de los datos y de la teoría del diseño experimental. Su idea consiste en manejar los 500 desplazamientos señalados como una muestra azarosa (*random sample*) de la población con distribución normal. Perrin puso las partículas de polen en nueve círculos concéntricos dibujados como un blanco, y registró los desplazamientos de las partículas de polen en las nueve zonas. La ley de los grandes números le permite predecir una distribución normal e hipotética calculable en estas nueve zonas. Lo que Perrin hizo fue averiguar la desviación del número de los desplazamientos observados(n_o) con los calculados(n_c) por la fórmula estadística: $(n_o - n_c)^2 / n_c$. Comparando el resultado con el nivel de significación de la distribución de ji-cuadrado, Perrin encontró que la desviación no es suficiente significativa, es decir, la hipótesis j pasó la prueba. Mayo describe así el razonamiento de esta decisión estadística:

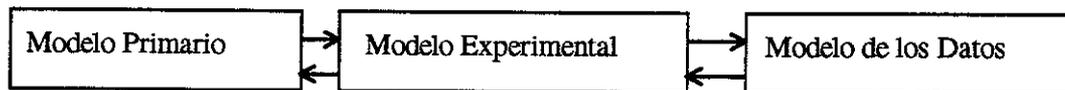
What matters is being able to sustain the argument of the significance test (i.e., answer the "significance question"). This we can do because we can determine the probability that a purely random sample of measurements taken from the hypothesized model distribution j would show worse agreement with the model j than is shown by the actual set. If the differences observed are of the sort frequently "caused by" chance(i.e., if they are typical under j), then the sample data are in accord with the hypothesized

experimental model. In Perrin's experiment E above, it turned out that the observed differences, or one even larger, are not infrequent but rather are typical, assuming j. That is, the observed difference is not statistically significant from what is typical under j. Hypothesis j passes. (Mayo,p,228)

Para realizar este experimento, Perrin tuvo que hacer mucho trabajo para diseñar adecuadamente los materiales e instrumentos experimentales a fin de satisfacer los presupuestos experimentales. Le llevó varios meses encontrar la manera de preparar las partículas de polen requeridas para que fueran del mismo tamaño y peso. Además, también tuvo que hacer esto de manera que pudieran excluirse influencias esenciales por factores extraños (no controlados en el experimento) sobre las condiciones *ceteris paribus*, tales como el cambio de temperatura, el color y la intensidad de luz, la vibración y la impureza de las partículas de polen durante el experimento. Estos trabajos requerían tanto de la habilidad de los experimentadores como de conocimientos de una serie de experimentos anteriores que habían sido llevado a cabo durante medio siglo.⁵

1.2.5 La jerarquía de los modelos experimentales

Ahora podemos apreciar más claramente por medio de un esquema de la jerarquía de los modelos experimentales según Mayo:



Existen cuatro relaciones en el esquema arriba dibujado, y cada relación tiene su propia cuestión que se requiere que se resuelva localmente:

- (1) La relación entre el modelo primario y el modelo experimental captura la cuestión de cómo se someten las cuestiones primarias a una investigación realizable, o sea, cómo reducirlas a preguntas más locales que puedan concretarse en pruebas;
- (2) La relación entre el modelo experimental y el modelo primario se refiere a cómo se relacionan las cuestiones primarias locales con las experimentales y canónicas de los experimentos concretos;
- (3) La relación entre el modelo experimental y el modelo de los datos se centra en la cuestión de cómo conectar los datos con los presupuestos experimentales, y finalmente
- (4) La relación entre el modelo de los datos y el modelo experimental, por medio de la cual se manejan las materias primas a fin de acomodarlas de una forma canónica preestablecida. Esto es lo que nos permite decidir si los datos generados satisfacen o no los presupuestos experimentales.

Así, las cuestiones primarias o teóricas no se relacionan directamente con las de los datos, sino que se conectan indirectamente a diferentes niveles. Y cada nivel tiene sus propios problemas por resolver, y sus propias teorías para enfrentarlos. A diferencia de la idea clásica acerca de la relación entre las teorías científicas y los experimentos, según la cual que ambos se conectan directamente, este esquema de Mayo tiene la ventaja de ser más realista respecto a las prácticas científicas.

Mayo introduce la metodología estadística en este esquema e insiste que esta metodología estadística es la clave para entender la inferencia científica, y sobre todo, para la construcción de una filosofía de la experimentación adecuada.

1.3 El uso de la estadística en los experimentos

En esta sección voy a estudiar brevemente la metodología estadística que Mayo introduce para analizar los experimentos y los dos conceptos que ella considera claves para entender su filosofía de la experimentación: la argumentación a partir de los errores y la severidad de los experimentos. Veremos que la metodología estadística basada en estos conceptos es el núcleo de la jerarquía estudiada en la sección anterior, o sea, que la construcción de la jerarquía se base sobre esta metodología estadística.

La idea principal de la prueba de Neyman-Pearson es que pueden compararse la distribución normal o teórica con las distribuciones que se obtienen de los experimentos, y que esta comparación puede servir de base a las inferencias que nos permiten conectar los resultados obtenidos en las muestras con propiedades de poblaciones. Usando el simbolismo de Mayo, primero voy a analizar el proceso de la prueba de Neyman-Pearson en una forma no formal pero sencilla, para luego ejemplificarla con un caso concreto.

Supongamos que tenemos una hipótesis nula H y una hipótesis contraria J que son inconsistentes:

$$H: \mu = \mu_0$$

$$J: \mu \neq \mu_0$$

μ_0 es la media de la población hipotética y μ es la media de las muestras. En los experimentos concretos, μ se expresa como \bar{X} , que se refiere a la media de los datos que se obtiene en los experimentos. Lo que nos interesa es la diferencia entre \bar{X} y μ_0 , si lo que queremos es formarnos un juicio sobre la aceptación o rechazo de H y J , ya que esta diferencia calculada nos señala la diferencia entre las propiedades de las muestras que los científicos han manejado y las propiedades de las poblaciones hipotéticas en cuestión. La diferencia estadística D se expresa así:

$$D = \bar{X} - \mu_0$$

La prueba de Neyman-Pearson T^* es un cálculo estadístico que señala el rechazo de H y la aceptación de J dependiendo de si la variable \bar{X} se aleja de H más de \bar{X}^* . El punto \bar{X}^* se especifica por medio de la probabilidad de rechazar H cuando H es correcta, o sea cuando la decisión de rechazar H es un error. Llamamos a este error el error de tipo I. Esta probabilidad se expresa por α , la que se le conoce en estadística con el nombre del nivel de significación o el tamaño (*size*) de la prueba. La decisión de qué α se escoge depende del tipo de experimentos y de contextos. Supongamos que en un experimento, el nivel de significación α es 0.03, o sea, la posibilidad de cometer el error de rechazar H cuando H es correcta no es más que 3 por ciento, en este caso la prueba de Neyman-Pearson se expresa así:

la prueba T^* : rechazar H en el nivel 0.03 si y sólo si $\bar{X} > \mu_0$ por 2 d.e. (x_n)

donde d.e. (x_n) es la desviación estándar. Si ya tenemos el nivel de significación, podemos encontrar d.e. (x_n) correspondiente en la curva de la distribución normal.

Mientras tanto, los experimentadores también se interesan por estimar la probabilidad de cometer el error de aceptar H cuando H es falsa o J es correcta. A este se le conoce con el nombre del error de tipo II, y se representa con el símbolo β :

$\beta(\mu')$: la posibilidad de que la prueba T^* acepte H cuando H es falsa.

donde μ' es la media límite cuando J es correcto. Dado que J dice que $\mu > \mu_0$, tenemos la hipótesis contraria J' respecto a J :

$$J': \mu = \mu'$$

Entonces, podemos expresar $\beta(\mu')$ como así:

$$\beta(\mu'): \Pr(T^* \text{ acepta } H / J' \text{ es correcta})$$

Recordamos que T^* acepta H cuando \bar{X} es menos que \bar{X}^* , así:

$$\beta(\mu'): \Pr(\bar{X} < \bar{X}^* / J' \text{ es correcta})$$

En los experimentos reales, los investigadores suelen usar el concepto de poder (*power*) para referirse a β , esto es para expresar la capacidad de un experimento de no cometer el error de tipo II:

$$\text{el poder de } T^* \text{ contra } J' = \Pr(\bar{X} \geq \bar{X}^* / J' \text{ es correcto})$$

Así, el poder de no rechazar H cuando J' es correcta es $1 - \beta(\mu')$.

Ahora vamos a ver cómo los experimentadores usan las ideas estadísticas expresadas arriba por un caso concreto, modelado por medio de la jerarquía de los modelos experimentales de Mayo.

Un doctor quiere saber si fumar afecta al temblor de las manos y hace un experimento. La idea en el modelo primario es medir los grados del temblor en dos grupos de gente, un grupo que fuma y el otro que no fuma. Llamamos al primer grupo el grupo en prueba y al segundo grupo el grupo de

control. Si no hay diferencia significativa entre los dos grupos, deberíamos concluir que el fumar no influye en el temblor de las manos. Por el contrario, si el resultado del grupo en prueba es significativamente mayor del grupo de control, la conclusión debería ser el que fumar afecta al temblor de las manos. Utilizando el simbolismo anterior, μ_f es la media de la población de los fumadores, y μ_n es la media de la población de la gente que no fuma. La hipótesis nula en el modelo primario puede expresarse así:

$$H_0: \mu_f - \mu_n = 0$$

En el modelo experimental, lo que los experimentadores manejan no son las poblaciones hipotéticas sino los resultados concretos del experimento. Si definimos $D = \bar{X} - \mu_0$ donde \bar{X} es la media del resultados del grado de temblor de las manos en el grupo en prueba, y μ_0 es la media en el grupo de control, la hipótesis nula que niega la correlación entre el fumar y el temblor de las manos se escribe así:

$$H: D = 0$$

y la hipótesis contraria que afirma la existencia de esta correlación:

$$J: D > 0$$

Una tarea del modelo experimental es la de decidir el nivel de significación α , que dice que cuando \bar{X} llega la cantidad de \bar{X}^* , la probabilidad de rechazar H cuando H es correcta es α . Supongamos que los científicos estén de acuerdo con que α sería 0.03, la prueba de Neyman-Pearson es

$$T^+: \text{rechaza H en el nivel de 0.03 si y sólo si } D \text{ tiene 2 d.e.}(x_n)$$

Para realizar este experimento, hay que hacer mucho trabajo en los modelos de los datos. Los científicos deben tener en cuenta la edad, el sexo, la experiencia antes de fumar y del temblor las manos en los integrantes de ambos grupos, y toman en cuenta otras posibles variables extrañas que puedan afectar al resultado, y sobre todo, el elemento psicológico. Mucha gente tiene la impresión de que el fumar tabaco afecta al temblor las manos, de manera que esta gente que fuma en el experimento muestra un alto grado de temblor, no por el efecto del tabaco sino por la sugerencia psicológica. En el experimento real, el experimentador C.L.Hull cuidadosamente preparó dos tipos de pipa. Uno que contenía el tabaco real adentro, y el otro sólo el aire húmedo y tibio. La gente que tomó ambos tipos de pipa creía que estaba fumando el tabaco.(Hull, p.161)

Ahora tenemos el resultado de los dos grupos como así:

sujeto	fuma(a)	no fuma(b)
A	50	44
B	37	28
C	63	54
D	52	37
E	83	62

F	48	40
G	43	44
H	55	49
I	47	53
J	25	23
K	71	69
L	58	51
M	47	42
N	51	51
O	66	60
P	52	45

Frente a estos datos, se requieren varios tratamientos estadísticos para calcular \bar{X} y para escoger el nivel de significación α apropiado. Los cálculos son los siguientes:

sujeto	la diferencia $X=a-b$	la desviación $x=X-\bar{X}$	la desviación cuadrada x^2
A	6	0	0
B	9	3	9
C	9	3	9
D	15	9	81
E	21	15	225
F	8	2	4
G	-1	-7	49
H	6	0	0
I	-6	-12	144
J	2	-4	16
K	2	-4	16
L	7	1	1
M	5	-1	1
N	0	-6	36
O	6	0	0
P	7	1	1
total	96	0	592

$$\bar{X}=96/n=6$$

$$S_x=\sqrt{\sum x^2/(n-1)}=\sqrt{592/15}=6.28$$

$$S_{x^*}=S_x/(\sqrt{n})=6.28/(\sqrt{16})=1.57$$

$$d.e.(x_n)=(\bar{X}-\mu_0)/S_{x^*}=(6-0)/1.57=3.82$$

El resultado es que la desviación estándar llega a ser 3.82 a lo que corresponde un nivel de significación de 0.01 que es menor que el α establecido:

$$\Pr(\bar{X} \geq 6/H) < 0.01$$

Esta fórmula significa que es menor a 0.01 el riesgo de cometer el error de que se rechaza H y se acepte J cuando en realidad H es correcto. Dada esta probabilidad, los científicos tendrán buena

razón para creer que el fumar afecta el temblor de las manos. De la misma manera también se puede calcular el poder β .

Aunque quizá parezca este un caso demasiado sencillo como para hacer comparación con los experimentos en las investigaciones científicas reales, este ejemplo muestra los principios básicos de la prueba estadística de Neyman-Pearson y su relación con la jerarquía de los modelos experimentales descritos por Mayo. La jerarquía de los modelos experimentales de Mayo, en contraposición a la idea clásica de una relación sencilla y directa entre las teorías y la evidencia, requiere de un estudio por medio de una metodología por partes de esta relación. Sin embargo, la jerarquía misma no tiene alcance sobre los temas más amplios de la inferencia estadística. Mayo cree que la argumentación a partir del error sirve de puente entre inferencia estadística y inferencia científica:

In contrast to the global inductive approaches--a rule for any given data and hypothesis--so attractive to philosophers, I favor a model of experimental learning that is more of a piecemeal approach, whereby one question may be asked at a time if carrying out, modeling, and interpreting experiments--even to determine what "the data" are...By insisting on a global measure of evidential-relationship, philosophers have overlooked the value of piecemeal error-statistical tools for filling out a series of models that link data to experiments. But how are the pieces intertwined so that the result is genuinely ampliative? Unlocking this puzzle occupied me for some time. One way to describe how statistical methods function, I came to see, is that they enable us, quite literally, to learn from error. A main task of this book is to develop this view.(pp.xi-xii)

Eso significa que para formular una filosofía de experimentación, Mayo necesita plantear algo más: cómo conectan las inferencias científicas con las inferencias estadísticas que sólo se estudian por medio de una metodología por partes. Para puntualizar este punto, Mayo formula dos conceptos importantes: la argumentación a partir del error y la severidad de los experimentos.

Durante los experimentos reales, existen muchas oportunidades para que se cuelen muchos tipos de errores. Mayo sintetiza cuatro tipos principales:

- a) cuando se toman efectos espurios por algo real, y por lo tanto se consideran efectos falsos como correlaciones y regularidades genuinas;
- b) errores sobre la cantidad o el valor de un parámetro;
- c) errores sobre los factores causales;
- d) errores en los presupuestos experimentales.

La argumentación a partir del error según Mayo consiste en el conocimiento que derivamos de la ausencia y la presencia de estos cuatro tipos de errores:

It is learned that an error is absent when (and only to the extent that) a procedure of inquiry (which may include several tests) having a high probability of detecting the error if (and only if) it exists nevertheless fails to do so, but instead produces results that accord well with the absence of the error.(p,64)⁶

Supongamos que con base en experiencias anteriores los científicos piensan que podría existir un error. Pueden diseñar un experimento que tenga una alta probabilidad de detectar este error. Cuando aún así no lo encuentren, tendrán una buena razón para descartar su existencia. Mayo llama este proceso de razonamiento inferencia experimental, que es más amplia que la inferencia estadística:

Although a single inquiry involves a network of models, an overall logic of experimental inference emerges: data e indicate the correctness of hypothesis H , to the extent that H passes a severe test with e . All the tasks of the interconnected models are directed towards substantiating this piece of reasoning. To remind us, hypothesis H passes a severe test with e if e fits H , and the test procedure had a high probability of producing a result that accords less well with H than e does, if H were false or incorrect.(p,445)

Aquí, podemos considerar que H es la hipótesis sobre la existencia o no de un error. Podemos decir entonces que la lógica de la inferencia experimental es el núcleo de la argumentación a partir del error.⁷

Mayo insiste en que la argumentación a partir del error es muy importante para elaborar una filosofía de la experimentación, ya que es sobre esta base que radica precisamente el conocimiento experimental no formalizado:

Experimental knowledge, as I understand it, may be construed in a formal or informal mode. In its formal mode, experimental knowledge is knowledge of the probabilities of specified outcomes in some actual or hypothetical series of experiments. Its formal statement may be given by an experimental distribution(a list of outcomes and their associated probabilities), or by a standard "random" process such as a coin-tossing mechanism. Typically, the interest is only in some key characteristic of this distribution--a parameter--such as its arithmetic mean. In its informal mode, the one the practitioner is generally engaged in, experimental knowledge is knowledge of the presence or absence of errors.(pp.12-13)

El otro concepto importante es el de la severidad de un experimento. Esta idea es muy importante para la inferencia científica porque si un experimento se hace con un mayor grado de severidad, los experimentadores tendrán más razones para confiar en sus juicios basados sobre los cálculos estadísticos. Una definición de la severidad de un experimento puede expresarse así:

El criterio de severidad para una prueba de aceptación o rechazo (*pass-fail*): Es muy alta la probabilidad de que H no pase la prueba E, dado que H es falsa.

De esta manera, se puede entender la severidad como grado de posibilidad de incorporar un error: cuando la severidad es alta, este grado es bajo. En el caso del experimento sobre la relación entre el fumar y el temblor de las manos, el nivel de significación α es 0.03, ello quiere decir que si el resultado es menor que 0.03, los experimentadores van a rechazar la hipótesis H y aceptar J porque la probabilidad de que este juicio sea erróneo es sólo 0.03. La severidad de este juicio se define como $1-\alpha$, que es 0.97. El resultado de un experimento es:

$$\Pr(\bar{X} \geq 6/H) < 0.01$$

Esto significa que si la hipótesis H, que dice que el fumar no afecta al temblor de las manos, es correcta, que la posibilidad teórica calculada por medio de la desviación estándar \bar{X} sea igual o más de 6 es menor que 0.01, que es menor que el valor de 0.03 planteado con anterioridad del experimento. En este caso, los experimentadores rechazan H y aceptan J, y sus juicios son confirmados por un experimento con grado de severidad de 0.97.

Mayo cree que con este concepto de la severidad podemos calcular la probabilidad del error de rechazar una hipótesis H cuando H es correcta, al que hemos llamado error de tipo I, y también podemos calcular la probabilidad del error del tipo II, que consiste en aceptar H cuando H es falsa. Se trata de errores por los cuales están más preocupados los experimentadores, que quieren hacer una inferencia sobre las hipótesis primarias. Puesto que esta probabilidad se obtiene por medio de las inferencias estadísticas, el concepto de la severidad tendrá el alcance sobre la inferencia científica. Mayo explica la relación entre el concepto la argumentación a partir del error y el de la severidad, y define el alcance de la severidad sobre los problemas primarios de la siguiente manera:

In the canonical arguments from error, the probabilistic severity requirement may capture the argument from error so well that no distinction between so-called formal and informal modes is needed. In general, however, asserting that a hypothesis H passed a highly severe test in this formal sense is but a pale reflection of the actual experimental argument that sustains inferring H. The purpose of the formal characterization is to provide a shorthand for the actual argument from error, which necessarily takes on different forms. The formal severity criterion may be seen to

represent a systematic way of scrutinizing the appropriateness of a given experimental analysis of a primary question.(p,185)

En este capítulo, pasando por alto muchas cuestiones de detalle del sistema de Mayo, hemos repasado muy brevemente sus ideas principales sobre la filosofía de la experimentación. Las ventajas de este proyecto son muy claras. Primero, la jerarquía de los modelos experimentales nos provee de una descripción de la investigación en que las teorías no están directamente conectadas con la evidencia, como se tiende a aceptar por los filósofos de la ciencia. Esta metodología que plantea estudiar los problemas “por partes” es más cercana a las prácticas científicas y puede evitar muchas distorsiones típicamente asociadas con una filosofía de la ciencia dedicada exclusivamente al estudio de teorías. En segunda lugar, Mayo señala correctamente que la metodología estadística ha sido usada muy frecuentemente en las investigaciones científicas, pero, que no ha sido estudiada con suficiente atención por los filósofos, ni mucho menos por los nuevos experimentalistas. Una filosofía de la experimentación no debe olvidar el papel decisivo que juega la metodología estadística.

Sin embargo, me parece muy problemática la intención de Mayo de extender las inferencias estadísticas a todas las áreas de la experimentación y la inferencia científica. El propósito de mi tesis radica precisamente en estudiar los problemas que surgen tanto en los modelos experimentales como en las inferencias científicas en general, cuando se pretende tomar este enfoque como el único y definitivo apropiado para el planteamiento de los problemas filosóficos que surgen de las prácticas experimentales.

Notas

1. Los bayesianos, han habido varios intentos para evitar que la subjetividad en la asignación de probabilidad anterior tenga como consecuencia la acusación de la subjetividad a la asignación de la probabilidad anterior. El más popular es el proyecto de convergencia (*wash out*), que sostiene que cuando las evidencias se acumulan, las diferencias entre diferentes creencias disminuyen, de manera que a largo plazo (*in the long run*), por fin se pueden llegar a una convergencia intersubjetiva. Earman señala que esto proyecto es efectivo sólo en algunos casos, pero no funciona en otros por las siguientes razones:

(1) Los científicos pueden tener diferentes estimaciones en cuándo la convergencia llega (*the rate of convergence*), pero las diferentes opiniones sobre eso impiden que el proyecto de la convergencia sirva para resolver el problema de la subjetividad en la asignación de la probabilidad anterior. Dice Earman: “...[In] the long run we are all dead has no sting in the present context if we can know in advance how long the long run is...Nor does it seem possible to derive informative estimates in the present general setting.” (Earman, 1992,p,148)

(2) No hay una manera de representar las diferentes formas de la valuación sobre las evidencias relevantes por siguientes dos razones: (a) “...[In] the general case, the relevant evidence can come in myriad forms, and within a form the order can matter crucially. Some sort of estimate of rate of convergence could be produced by averaging over the rates for different sequences of evidence strings”; y (b) “Scientists often agree that a particular bit of evidence supports

one theory better than another or that a particular theory is better supported by one experimental finding than another...What happens in the long or the short run when additional pieces of evidence are added is irrelevant to the explanation of shared judgements about the evidential value of present evidence.” (Earman,1992,p,149) Vale la pena mencionar que son precisamente estas dos razones las que Mayo usa en su libro para en contra el proyecto de convergencia. (Mayo, pp.83-85)

(3) En caso de que dos teorías son empíricamente indistinguibles, o sea, son infradeterminadas, el teorema bayesiano no funciona. (Earman, 1992,pp,149-153)

Otro intento importante es el proyecto de restricción de la probabilidad anterior (*constraining priors*), el cual sostiene que podemos restringir el subjetivismo por medio de recurrir a reglas artificiales. Por ejemplo, podemos restringir que las teorías más simples tienen la probabilidad anterior más alta. Sin embargo, Earman señala que el problema de este proyecto es que los científicos pueden tener diferentes forma de aplicar estas reglas, y esta diferencia puede conducir resultados que estén en conflictos. (Earman,1992,p,139) Además, Earman han analizado otros intentos menos populares pero igualmente importantes ((1)*definitional solution*, (2)*socialism*, (3)*evolutionary solution*, (4)*modest but realistic solutions*, (5)*non-Bayesian solutions* y (6)*retrenchment*), y concluye que todos estos intentos son efectivos sólo en ciertos casos, pero no pueden resolver el problema del subjetivismo.

Otra dirección de estos intentos es la tesis de “la impecable objetividad” defendida por Howson y Urbach, quien insiste en que la intención de la inferencia bayesiana es para proveernos una lógica inductiva. (Howson, 1996) Como una lógica deductiva, la lógica inductiva sólo interesa a la inferencia válida entre premisa y conclusión, y no se preocupa si la premisa es correcta o no. Por eso, “[Bayesian theory] is unimpeachably objective: though its subject matter, degrees of belief, is subjective, the rules of consistency imposed on them are not at all.” (Howson y Urbach,1989, primer edi,p,290) Mayo cree que esta dirección bayesiana no es aceptable en modelar la inferencia científica, porque no puede captar lo que los científicos realmente quieren hacer. Por ejemplo, la meta de un experimento no es para demostrar la coherencia de las creencias de los científicos, sino que es para descubrir o controlar los posibles errores en los procedimientos de las investigaciones científicas. (Mayo, pp.85-86) Dice ella: “What the Bayesian has given us comes down to a tautology: if the agent assigns a high posterior degree of belief to the falsity of A', then the subjective probability that A' is false is high. For the error statistician, a hypothesis is warranted only if its errors have been ruled out; that is not required in order for hypotheses to be strongly believed.” (Mayo, 1997, pp.325-326)

2. Según Suppes (1969), la expresión formal de estas tres modelos puede expresarse así:

(1) Definimos la teoría de los modelos de la respuesta lineal como un triple ordenado $\lambda = \langle X, P, \theta \rangle$; donde

(a) X es el conjunto de todas las sucesiones de pares ordenados, cuyo primer miembro es un elemento del conjunto A, y el segundo miembro es un elemento del conjunto B. En el caso más simple posible, el conjunto A incluye dos posible respuestas, y el conjunto B incluye dos posible refuerzos asociados con cada una de las respuestas en A. Las respuestas en A pueden interpretarse como dos hipótesis y los refuerzos respectivos pueden interpretarse como rechazo o aceptación.

(b) La función P es el cálculo estadístico sobre X, y

(c) θ es el número real que está adentro el intervalo $0 < \theta \leq 1$, representando el nivel de significación.

Para incluir los números finitos al representar la repetición de las pruebas, y para representar θ que no es directamente observable ni constituye una parte de los datos, Suppes introduce los siguientes axiomas:

A1: si $\Pr(B_{i,n} A_{r,n} x_{n-1}) > 0$, entonces $\Pr(A_{i,n+1} / B_{i,n} A_{r,n} x_{n-1}) = (1-\theta)\Pr(A_{i,n} / x_{n-1}) + \theta$.

A2: si $\Pr(B_{j,n} | A_{i,n-1}) > 0$, y $i \neq j$, entonces $\Pr(A_{i,n+1} | B_{j,n} | A_{i,n-1}) = (1-\theta)\Pr(A_{i,n} | x_{n-1})$.

donde $i, j=1, 2$; $A_{i,n}$ es la respuesta A_i en la prueba n ; $B_{j,n}$ es el refuerzo B_j en la prueba n ; y x_n es la clase equivalente de todas sucesiones de X en la prueba n . El A1 afirma que cuando una respuesta está reforzada (o confirmada), la probabilidad de la aparición de la misma respuesta en la siguiente prueba aumenta. El A2 afirma que cuando una respuesta no está reforzada (o otra respuesta está reforzada), la aparición de la misma respuesta en la siguiente prueba disminuye.

(2) Definimos la teoría de los modelos experimentales así: $\forall \langle Y, P \rangle$, donde

(a) Y es un conjunto finito de par ordenado cuyo primer elemento será un miembro de A . Y representa todas las sucesiones de todas las repeticiones de una prueba, y

(b) P es la probabilidad calculada sobre todos los subconjuntos de Y .

(3) Para definir la teoría de los modelos de los datos, introducimos la N -tuplas Z que son los elementos de Y , donde N es el número de los sujetos en una prueba. Definimos que Z es el modelo de los datos cuando las frecuencias relativas condicionales de refuerzos $B1$ y $B2$ encajan, dentro el nivel de significación, con la probabilidad P calculada en el modelo experimental.

3, D es el coeficiente de difusión, un concepto estadístico, y t es el tiempo. S_t es la media de los datos y μ es la media de la población. La deducción de expresión matemática y formal se ve en la página 221 de Mayo.

4, El razonamiento de la utilización de la prueba de significación la expresa Suppes por medio de usar la teoría de comprensión: "For each possible realization Z of the data, we define a statistic $T(Z)$ for each question [about assumptions]. This statistic is a random variable with a probability distribution... We "accept" the null hypothesis if the obtained value of the statistic $T(Z)$ has a probability equal to or greater than some significance level α on the assumption that indeed the null hypothesis is true." (Suppes, 1969, p.29) (sobre Z , véase la nota 2)

5. Sin embargo, hay un punto inconsistente en este modelo de los datos que Mayo plantea. Es sabido que los problemas del diseño experimental, de la generación de los datos y de la especificación de las condiciones *ceteris paribus* son elementos importantes para una filosofía de la experimentación. Mayo separa estos tres problemas del modelo de los datos, porque considera que estos tres problemas no pueden reducirse a cuestiones estadísticas, que se requiere del modelo experimental para relacionar la hipótesis primaria con el modelo de los datos. Dice ella: "there are also numerous and sundry factors that are not incorporated in any of the models of data for assessing the primary hypothesis. There are the leftovers that Suppes places under the heading of the *ceteris paribus* conditions of an experiment. 'Here is placed every intuitive consideration of experimental design that involves no formal statistics.'... For two reasons I think it is better to include them [*ceteris paribus*] with experimental design: First, even those features of data generation that do relate explicitly to data models need not always be checked by formal statistics. Second, features assumed to be irrelevant or controlled may at a latter stage turn out to require explicit scrutiny. For example, the sex of an experimenter in a study on human subjects may not be explicitly incorporated in a data model, and thereby placed under *ceteris paribus*, yet a suspected influence of gender on a subject's response may later be studied as a possible obstacle to satisfying the experimental model assumptions." (Mayo, p.139) Es muy claro que la separación artificial hecho por Mayo se basa en su manera de enfocar el estudio de la metodología estadística. Pero es muy difícil hablar de la función del modelo de los datos sin mencionar problemas relacionados con la metodología estadística. De hecho en muchos

lugares de su libro la autora estudia los casos concretos de los experimentos sin hacer distinción entre el modelo de los datos y los tres problemas mencionados. Sin embargo, la jerarquía de los modelos experimentales esbozado por Mayo sólo permite relaciones entre el modelo primario y el modelo experimental, y entre el modelo experimental y el modelo de los datos, y no permite un lugar para estos tres problemas que son elementos muy importantes en la experimentación. Esto muestra un defecto de la jerarquía de los modelos experimentales de Mayo, puesto que se enfoca sólo y exclusivamente a los elementos estadísticos y no es capaz de incorporar los otros elementos que ella misma reconoce como importantes en una filosofía de la experimentación. Los nuevos experimentalistas han hecho bastantes estudios sobre los problemas del diseño de los instrumentos experimentales y el proceso de la generación de los datos, en los cuales experiencias personales, tanto como las manipulaciones y prácticas han sido considerados elementos importantes de los experimentos. Mayo insiste en que la diferencia de la actitud sobre estos tres problemas mencionados entre ella y los nuevos experimentalistas es por diferente énfasis sobre el mismo tema: "To some extent, the difference between the typical experimental narrative and what I am calling for is a matter of emphasis. For example, an important consideration under experimental design would be various protocols for instrumentation--something already found in many narratives [of New Experimentalists]--but the emphasis should be on how the protocols bear on experimental assumptions, for example, assessing experimental errors, making any needed adjustments, circumventing pitfalls... (Mayo,p,139)" Pero, me parece que lo que Mayo hace aquí no es un cambio de énfasis sobre el mismo tema, sino realmente está proponiendo un nuevo criterio. Bajo este criterio, ella tiene la dificultad de incorporar estos tres problemas en su modelo de la experimentación, debido a su intención de reducir todos los modelos en la jerarquía a cuestiones estadísticas.

6, La argumentación a partir del error se puede expresar como así: "It is learned that an error is present when a procedure of inquiry that has a very high probability of not detecting an error if (and only if) none exists nevertheless detects an error." (Mayo,p,185)

7, Se puede considerar este párrafo como aquí que dice Mayo: "If a hypothesized error is not detected by a test that has an overwhelmingly high chance of detecting it, if instead the test yields a result that accords well with no error, then there are grounds for the claim that the error is absent, We can infer something positive, that the particular error is absent (o is no greater than a certain amount). Equivalently, we have grounds for rejecting the hypothesis, H', that the error is present, and affirming H, that it is absent. When we have such information, we say that H has passed a severe test. Alternatively, we can say that the test result is a good indication that H is correct." (Mayo,p,7)

Capítulo 2 La Limitación del Modelo Primario

Como se dijo en el capítulo anterior, para Deborah Mayo, el modelo primario tiene la función de permitir que las teorías globales, que carecen de consecuencias empíricas, generen hipótesis o problemas primarios y locales, que a su vez son capaces de sugerir experimentos. Generalmente, un problema primario se avoca a la obtención de las pruebas para las hipótesis locales, o de las estimaciones de los valores de los parámetros de ciertas teorías, dados la forma de las ecuaciones. También tiene que especificar los procedimientos para la obtención de estas pruebas, o de las estimaciones para el modelo de los datos. Esta idea de modelo primario tiene dos consecuencias graves que voy a criticar en este capítulo:

- (1) Todos los experimentos se origina en el modelo primario, porque éste decide la dirección y la forma de la ejecución del experimento. Puesto que los problemas primarios son problemas locales derivados de las teorías globales, los experimentos siempre dependen de los intereses o propósitos de las investigaciones de las teorías globales, por medio de la decisión o la especificación de la distribución estadística.
- (2) Puesto que los problemas primarios se limitan exclusivamente a las consideraciones estadísticas, o son reducibles a hipótesis estadísticas, a la dicotomía entre hipótesis nulas y las contrarias, todos los experimentos son esencialmente estadísticos.

En este capítulo, examinaré críticamente dos consecuencias del modelo primario y de la jerarquía de los modelos experimentales de Mayo. Mi tesis principal es que el enfoque estadístico de Mayo sirve como descripción sólo de una parte de lo que son los experimentos científicos, aunque sea una parte muy importante. Dista mucho sin embargo de proveernos de una filosofía de la experimentación completa como Mayo pretende. Primero, quiero hacer ver que Mayo entiende la relación entre las teorías y los experimentos de modo diferente a como los piensan los nuevos experimentalistas. Esta discrepancia se refleja en la consecuencia (1). Lo que haré es mostrar diferentes tipos de experimentos que no están relacionados con la metodología estadística en el sentido de Mayo. Finalmente, explicaré la manera en que Peter Galison clasifica los experimentos y una tradición experimental que no es estadística, a fin de sostener, que contra la consecuencia (2), la jerarquía de los modelos experimentales de Mayo no puede abarcar todas las formas que asumen las investigaciones experimentales.

2.1 “La experimentación tiene vida propia”

Todos los nuevos experimentalistas sostienen el lema formulado por Hacking: “la experimentación tiene vida propia”. Hemos visto en 1.1 los tres sentidos entendidos por Mayo de este lema. Estos tres sentidos muestran que el estudio de la experimentación debe enfocarse hacia las situaciones locales. Estoy de acuerdo con esto. Sin embargo, la interpretación de Mayo del primer sentido del lema de Hacking no es correcta. Lo que Mayo entiende por diferencia entre las metas de los experimentos y a las de las teorías, consiste tan sólo en que los experimentos asumen las tareas más locales y inmediatas, como calibrar los instrumentos, eliminar las variables extrañas, estimar sobre el nivel de significación, distinguir los efectos reales de los artefactos y estimar las influencias de los factores contextuales. Pero la jerarquía de los modelos experimentales de Mayo, en particular el modelo primario, requiere que todas estas tareas locales se realizan bajo el control de los problemas primarios. Así, Perrin habría hecho este tipo de tareas para probar la teoría de Einstein y Smoluchowshi. Bajo la jerarquía de los modelos experimentales de Mayo, los experimentos se relacionan y están bajo el control, aunque no directamente, de las teorías globales. Es a través del medio del modelo primario que se decide la distribución estadística para los experimentos; distribución con la que se ponen a prueba ciertas hipótesis locales y se estiman los parámetros de ciertas teorías. De ahí la inevitabilidad en el esquema de Mayo de que un experimento siempre tenga que ser precedido por una teoría. Eso es precisamente lo que Hacking critica. Las metas de los experimentos según Hacking y otros nuevos experimentalistas son muy diferentes de las que Mayo propone.

La idea de que las teorías deciden el curso de los experimentos, provino de las críticas al positivismo lógico. Después de que aparecieron los conceptos de “armazón” (*framework*) del segundo Carnap, “esquema conceptual” y “medio ambiente conceptual” de Frank y Quine, así como el de “paradigma” de Kuhn, los anti-positivistas se convencieron de que las observaciones están cargadas de teoría, y es imposible la neutralidad en los enunciados observables básicos. Esto hace insostenible la tesis reduccionista sobre la relación entre las teorías sucesivas en un campo científico, como la defendía el positivismo lógico. Los anti-positivistas siempre han compartido, sin embargo, una tesis con los positivistas lógicos, la tesis de que los experimentos están (desde un punto de vista epistemológico) subordinado a teorías.

Cuando Hacking critica la tesis de que los experimentos siempre están subordinados a teorías, él es cuidadoso en distinguir entre una versión débil y una fuerte de ésta. En la versión débil la realización de todo experimento presupone ciertos conocimientos sobre la naturaleza y sobre los instrumentos usados. Como ya lo mencionamos, entre estos conocimientos presupuestos están lo que Mayo llama las tareas locales que determinan las metas de los experimentos. Hacking acepta esta

versión débil pero se opone a la versión fuerte según la cual un experimento es significativo sólo si su objetivo es poner a prueba una teoría sobre los fenómenos bajo escrutinio. (Hacking, 1983, p. 182) La descripción de Mayo asume la versión fuerte, al afirmar que la relación entre una teoría y un experimento se da por medio del modelo primario; aunque sea de una manera más sofisticada que los popperianos o los positivistas lógicos.¹

Los nuevos experimentalistas se oponen a un supuesto como el del modelo primario de Mayo. No sólo se oponen a que una teoría tiene que estar establecida antes de que un experimento tenga sentido, se oponen también a la idea que los experimentos se realizan en última instancia para probar teorías o estimar sus parámetros. Hacking insiste en que la teoría no necesariamente va primero, y sostiene que existen cuatro posibilidades para la relación entre un experimento y una teoría:

A veces hay trabajos experimentales profundos que la teoría genera en su totalidad. Algunas teorías importantes salen del experimento preteórico. Algunas teorías languidecen por falta de conexiones con el mundo real, mientras que algunos experimentos no tienen nada que hacer por falta de teoría. (Hacking, 1983, p. 187)

La jerarquía de los modelos experimentales de Mayo, al igual que los anti-positivistas y los popperianos, sólo considera la primera posibilidad, es decir, que las teorías preceden a los experimentos.

Para combatir esta idea, Hacking y Everitt describen numerosos experimentos cuyos propósitos no fueron poner a prueba una teoría o estimar sus parámetros.

Algunos experimentos pueden estar guiados por la pura curiosidad de los científicos. Por ejemplo, entre 1600 y 1800, muchos científicos investigaban los dos rayos “extraordinarios” que producía el espato o calcita de Islandia. Estos estudios no estuvieron guiados por ninguna teoría; la teoría ondulatoria de la luz en la que luego hallaron acomodo fue formulada sólo después de 1800 por Fresnel y Thomas Young. Sin embargo, como afirma Hacking:

Por supuesto que Bartholin, Grimaldi, Hooke y Newton [quienes estudiaban el espato] no eran empiristas insensatos sin ninguna idea en la cabeza. Ellos vieron lo que vieron porque eran curiosos, inquisitivos, gente reflexiva. Estaban tratando de formar teorías. Pero en todos estos casos está claro que las observaciones precedieron a cualquier formulación teórica. (Hacking, 1983, p. 184)

Por si se arguyera que estos casos se limitan sólo a las observaciones que no son experimentos en el sentido estricto, podemos mencionar otros ejemplos más persuasivos. Por ejemplo, los experimentos de óptica que E.L. Malus y Brewster realizaron antes de que se formulara la teoría de la polarización de la luz por reflexión. Los experimentos de R.W. Wood sobre efectos como la

radiación de resonancia, la fluorescencia, los espectros de absorción y los espectros Raman fueron hechos en condiciones en que no había surgido la teoría cuántica que es la teoría que permite explicarla. Mientras tanto, el movimiento browniano y el efecto fotoeléctrico de A.C. Becquerel fueron observados muchos años antes de que una teoría los pudiera explicar. Hacking señala que, a pesar de que muchos historiadores describen la observación de Davy sobre las algas en agua como consecuencia de sus consideraciones teóricas, este caso puede entenderse como el resultado de la curiosidad pura de Davy. De hecho, su curiosidad era el gran motor de sus investigaciones. Cuando estaba haciendo los experimentos de la descomposición de los álcalis, notó que “aunque había una elevada intensidad de acción, solamente quedaba afectada el agua de las disoluciones, separándose el hidrógeno y el oxígeno”. Sin ninguna explicación teórica en su cabeza, se le ocurrió la idea curiosa: ¿qué sucedería si se prescindiese del agua? Calentando “una cuchara de platino que contenía patasa, este álcali fue mantenido durante algunos minutos en estado de perfecta fluidez.” Cuando lo conectó con el polo positivo y negativo de la batería, descubrió los óxidos metálicos bautizados por Davy como “Potasium” y “Sodium”.

Muchas veces, los experimentadores intentan cosas que la teoría no contempla. Se refiere Derel de Solla Price a este tipo de experimentación en los siguientes términos:

Any effect or phenomenon like the Edison effect, Cerenkov radiation, and the creep of liquid helium, might be just the thing to measure or reveal that which we did not know before. Such experimentation is a sort of fishing expedition because you never quite know what you will catch, and always hope for the unexpected. It cannot be planned with an eye towards any particular objective, though of course it is common to say as a necessary condition for getting funding. (citado por Buchwald, p.169)

Hay otro tipo de experimentación que tiene con fin la invención tecnológica. Everrit dice que hay dos maneras de lograr esto. La primera es elaborar simultáneamente una teoría y su experimento, y después aplicarlos a problemas prácticos; en la segunda las invenciones siguen su propio curso práctico y la teoría se deriva de este de manera indirecta. En el segundo caso, los experimentos se efectúan antes de que haya una teoría (Hacking, 1983, p.191). Un ejemplo de ello es la secuencia de invención de la máquina de vapor. Primero, Newcomen inventó la máquina atmosférica. Watt inventó después la máquina de condensación que resultaba ser más efectiva y económica; y al final Trevithick inventó la máquina de alta presión. Todos estos experimentos se hicieron con los propósitos puramente prácticos. Varios años después, Sadi Carnot observó que la ventaja de la máquina de alta presión estaba en la velocidad de elevación del punto de ebullición del agua mediante la presión, y

formuló la teoría termodinámica; teoría que surgió para entender las máquinas, y que se desarrolló después como una ciencia general. Concluyen Hacking y Everitt:

Así, la palabra “termodinámica” nos recuerda que esta ciencia surgió del análisis profundo de una sucesión notable de invenciones. El desarrollo de esa tecnología requirió una gran cantidad de “experimento”, pero no en el sentido de contrastación popperiana de teorías, ni en el sentido de la inducción de Davy. Los experimentos eran intentos imaginativos requeridos para la perfección de la tecnología que se encuentra en el centro de la revolución industrial. (Hacking, 1983, p. 192)

Otro tipo de experimentación muy importante se ocupa de la medición. La medición desempeña diferentes funciones en la ciencia. Muchas mediciones se hacen para contrastar las teorías, como la medición del número de Avogadro en el experimento de Perrin. Pero las mediciones se relacionan con las teorías de muy diferentes maneras. Popper por ejemplo considera que la precisión misma es un criterio para probar las teorías, pues mientras más precisas las mediciones estas implican mayor riesgo para la teoría. Kuhn a su vez sostiene que la medición desempeña una tarea especial en la ciencia normal: encontrar “diferencias esotéricas” como las anomalías. Hacking plantea otra función de la medición independiente de las teorías: la precisión misma y el mejoramiento de las técnicas de medición. Entre los ejemplos de este tipo de mediciones están la del diámetro de la Tierra hecha por Aristarco, la del “peso” de la Tierra por Cavendish, la de la velocidad de la luz por Fizeau; y la de la carga del electrón por Millikan. Quizá el caso más típico es el de J.J. Berzelius, quien se ocupó toda su vida de perfeccionar las mediciones químicas. La teoría atómica de Dalton implicaba que cuando un determinado peso de un elemento A se combina con un cierto peso de un elemento B, si el mismo peso de A se combina con un determinado peso de un tercer elemento C, entonces cuando los elementos B y C se combinen estos se hará con una relación definida. Berzelius se obsesionaba por las mediciones de estas relaciones relativas entre los elementos. Su insistencia en medir los compuestos oxigenados, a saber, “(peso de óxido–peso de metal)/peso de metal” fue célebre en su época. El valor de estos experimentos radicó en la precisión alcanzada y en las técnicas diseñadas para hacer las mediciones. Según Harré:

El secreto de su éxito fue una especie de perfeccionismo, una auténtica obsesión por la precisión. “No tuve éxito en mis primeros intentos--dice Berzelius--. Aún no tenía yo experiencia en lo tocante a la gran exactitud que era necesaria, ni de cómo podría obtenerse precisión mayor en los resultados finales.” La solución de estas dificultades residía en prestar la máxima atención a los detalles. Era preciso diseñar el instrumental de modo que las pérdidas de material fueran lo más reducidas posible. En

las reacciones donde era preciso verter líquidos, las vasijas habían de tener labios que descargasen hasta última gota. Los filtros de papel no solo habían de tener unos residuos de cenizas perfectamente normalizados; sino que era aconsejable humedecerlos antes de usarlos, para evitar en lo posible que algunas de las sustancias en disolución pudieran ser absorbidas por las fibras de papel. Pero, por encima de todo, la técnica de manipulación tenía que ser precisa. Consistía en “observar un gran número de pequeños detalles, que, de no ser tenidos en cuenta, muchas veces podían echar a perder varias semanas de cuidadoso trabajo” (Harré,1981,pp.184-185)

En el año 1818, Berzelius había medido los pesos atómicos de 45 de los 49 elementos conocidos. El momento de su muerte en 1848, seguía, como lo había hecho esos 30 años mejorando y perfeccionando los resultados de sus mediciones. Pero ¿cómo entender el papel de una medición precisa? ¿Porqué resulta tan importante para las investigaciones científicas? Hacking lo explica así:

¿Por qué son importante estos experimentos? Hay dos razones para admirarlos. En primer lugar, fueron extraordinariamente exactos. No vamos a corregir los resultados de estos pioneros de ninguna manera significativa. En segundo lugar, cada individuo produjo una nueva técnica. Cada uno de ellos tuvo la genialidad de concebir no sólo una idea experimental brillante, sino la capacidad de ponerla en práctica por medio de la invención de una serie de concepciones experimentales adicionales y de innovaciones tecnológicas. (Hacking,1983,p,264)

Además de las cuatro posibilidades arriba mencionadas de relación entre teorías y experimentos, existe otra posibilidad relativamente rara, llamada por Hacking el “encuentro feliz”. En 1965 los radioastrónomos Arno Penzias y R.W. Wilson instalaron un radiotelescopio para estudiar las fuentes del ruido en el radio, lo que tenían propósito práctico para el desarrollo tecnológico. Pero descubrir una fuente de radiación de una temperatura de 3^0K , que parecía estar en todas partes del espacio, distribuida uniformemente. Penzias y Wilson no tenían ninguna idea para explicar este fenómeno y no se atrevían a publicar los resultados. Mientras tanto, la teoría cosmológica del “big-bang” se formuló sin ninguna relación con los experimentos de Penzias y Wilson. Esta teoría sostiene que el universo nació por una gran explosión y que uno de sus efectos consiste tendría que ser la existencia de una radiación con temperatura de 3^0K como energía residual en todo el espacio. Cuando se unieron finalmente, el experimento de hecho “confirmó” la teoría, pero cuando se obtuvo el resultado, este no tuvo nada que ver con la teoría.

También se puede llamar encuentro feliz a los descubrimientos accidentales, no anticipados por las teorías. Un ejemplo sería la famosa invención de la vacuna por Pasteur. Por mucho tiempo,

Pasteur trataba de buscar alguna forma de atenuar la virulencia, pero no la encontró. En 1679, se fue de vacaciones por varios meses. Cuando regresó su laboratorio, se encontró con que todavía quedaban ciertos cultivos de microbio hechos antes de sus vacaciones. Usó estos cultivos para seguir con sus experimentos, pero las gallinas no mostraban síntomas de la enfermedad de cólera aviar. Pasteur pensó que esto se debía a la conservación de los cultivos durante demasiado tiempo. Entonces, procedió a inyectar nuevos cultivos a estas gallinas, pero tampoco ocurrió nada con ellas. De esta manera, Pasteur llegó a la conclusión de que había sido descubierto un método para atenuar artificialmente el “virus”. Algo similar ocurrió con el descubrimiento de la transmutación artificial de un elemento. Rutherford hizo el experimento no con el propósito de probar ninguna hipótesis, sino para explorar la intensidad del momento cinético impartido a los productos de la desintegración de los átomos. Pudo observar que la fuente de las partículas α produjo los centelleos mucho más largas de lo debido. Después de varios experimentos, a fin de eliminar otras explicaciones, Rutherford sacó la conclusión de que el fenómeno se debía a la transmutación de los elementos, e identificó que estas partículas de largo alcance eran los átomos de hidrógeno. ¿Cómo se explica la relación entre una teoría y un experimento en estos casos? Me parece claro que no puede ser explicada por los modelos de Mayo. Quizá la única explicación es la de George Darwin, citada por Hacking, en el sentido de que “uno debería hacer de vez en cuando un experimento totalmente loco, como sonarles una trompeta a los tulípanes cada mañana durante un mes. Tal vez nada va a pasar, pero si algo pasara, sería un descubrimiento estupendo.” (Hacking, 1983, p, 182)

A través de estos casos, podemos concluir que no hay una relación simple y unívoca entre teorías y experimentos. Las diferentes situaciones mencionadas no pueden conciliarse con el modelo primario de Mayo, porque en ellas los experimentos no tienen la tarea de conectar las teorías globales a las hipótesis locales. La explicación que dan los nuevos experimentalistas, encabezados por Hacking, a la tesis de que “la experimentación tiene muchas vidas propias” es mucho más amplia que lo que Mayo plantea. Resulta así muy dudoso que un modelo pueda ser suficiente para una filosofía de la experimentación, ya que no tiene la capacidad de dar cuenta una gran cantidad de experimentos científicos importantes en la práctica científica real.

Ante objeciones análogas, Mayo se defiende afirmando que su metodología estadística en su filosofía de experimentación no está dominada por la teoría (*theory-dominated*). (pp.67-69; 1994, p, 272) Según ella, la filosofía de la confirmación o de la prueba se subyuga a la teoría cuando:

(1) the philosophical work begins with data or evidence statements already in hand; (2) the account seeks to provide uniform rules for relating evidence to any theory or conclusion (or decision) of interest; and (3) as a consequence of (1) and (2), the account

functions largely as a way to reconstruct a scientific inference or decision, rather than giving us tools scientists actually use or even a way to model the tools actually used.
(Mayo,p,68)

Los ejemplos de este tipo de dominación teórica son la lógica inductiva de Carnap y la inferencia bayesiana. Según Mayo su filosofía difiere de ellos por su método “por partes” para entender la relación entre teorías y experimentos. Dicho en otras palabras, la relación entre una teoría y un experimento según Mayo no es una relación directa, como es para Carnap y los bayesianos. Esto es cierto, pero Mayo comparte con ellos una tesis: la de la existencia de una relación fija y unívoca (no importa cuán indirecta y sofisticada) entre teorías y experimentos. Mayo fila esta relación por medio de los estrategias centrales a su propuesta. Una es que, como lo hemos visto, por medio de la distribución estadística en el modelo primario, las teorías globales controlan la dirección y el curso de los experimentos. La otra consiste en que los datos experimentales deben conformarse a los requisitos estadísticos del modelo de los datos, esto lo analizaremos en el siguiente capítulo. Pero a lo que se oponen los nuevos experimentalistas es precisamente a esa relación fija que Mayo y aún los antipositivistas comparten con Carnap y los bayesianos. A través de los ejemplos que hemos visto en esta sección, queda clara que el modelo primero que Mayo propone no es plausible porque todavía sigue la idea de que los experimentos siempre están subordinados a teorías, idea que simplifica los procedimientos complejos de la experimentación. Además, el modelo primario tiene otra idea implausible: todos los experimentos son reducibles a los experimentos estadístico. Esto voy a examinar en las siguientes dos secciones.

2.2 Los experimentos no estadísticos

Una tesis importante de la filosofía de la experimentación de Deborah Mayo es la utilización de la inferencia estadística para modelar los procedimientos experimentales. En la sección llamada “Put Your Epistemology of Experiment at the Level of Experiment”, Mayo critica la omisión que hacen los nuevos experimentalistas de las teorías estadísticas en sus estudios sobre los razonamientos experimentales:

In rejecting old-style accounts of confirmation as the wrong way to go, the New Experimentalists seem dubious about the value of utilizing statistical ideas to construct a general account of experimental inference. Theories of confirmation, inductive inference, and testing were born in a theory-dominated philosophy of science, and this is what they wish to move away from. (Mayo,p,67)²

Para Mayo la estadística es el núcleo de la epistemología de experimentación. Al ser ésta la función principal del modelo experimental, es incorporada al experimento por el modelo primario:

Suppose primary question is a test of a hypothesis. Because of the many sources of approximation and error that enter into arriving at the data, the data would rarely if ever be expected to agree exactly with theoretical predictions. As such, the link between data model and experimental hypothesis or question may often be modeled statistically, whether or not the primary theory or hypothesis is statistical. This statistical link can be modeled in two ways: the experimental prediction can itself be framed as a statistical hypothesis, or the statistical considerations may be seen to be introduced by the test (in its using a statistical test rule)...Either way, the primary hypothesis may be seen to entail something about the probability with which certain experimental results would occur. This is an experimental distribution--and it is in terms of the experimental distribution (in the experimental model) that the primary question is asked and answered. More exactly, when the primary question is tackled it is by means of a hypothesis about an experimental distribution (i.e., a statistical hypothesis). (Mayo,p, 134)

La idea principal es que una hipótesis primaria H puede ser no estadística, (por ejemplo, la hipótesis sobre la relación entre fumar y el temblor de las manos) pero implica siempre una distribución experimental o hipótesis estadística cuando se relaciona con el modelo experimental; es decir, cuando se baja la hipótesis primaria H a la hipótesis nula H_0 o a la hipótesis contraria J . Así, todos los experimentos implican a fin de cuentas una distribución estadística.

Creo que esta manera de reducir todos los experimentos a lo estadístico representa un enfoque demasiado estrecho. En las investigaciones científicas reales hay muchos experimentos que no son estadísticos, y los experimentos estadísticos son sólo una parte de la experimentación, tal como voy a mostrar en esta sección y la siguiente.

Hay muchos experimentos que persiguen el objetivo de probar una hipótesis pero que no implican ninguna distribución estadística. Veamos algunos ejemplos.

En el siglo XVI, la mayoría de la gente que estudiaba los fenómenos magnéticos creía que la tendencia de la aguja magnética a girar hacia los polos se debía a un “punto atractivo” que arrastraba hacia sí el de la aguja señalando el norte. Pero Robert Norman descubrió en su brújula una inclinación magnética. Es decir, observó una tendencia de la aguja imantada a apuntar, no solo hacia el norte, sino también hacia arriba o hacia abajo según la latitud. Esta desviación de la horizontal era

regular. El fenómeno provocó dudas en Norman sobre la idea del punto atractivo. Por ello, planteó un experimento para poner a prueba la hipótesis del punto atractivo.

“Ahora,” en palabras del mismo Norman. “para demostrar que no hay punto atractivo...se tomará un troza de alambre de hierro o acero, de dos o más pulgadas de longitud, y se atravesará con él una pieza de corcho compacto, lo bastante grande como para que...pueda mantener sobradamente el alambre sobre el agua, al flotar el dicho corcho en medio de ella...Después se tomará un vaso hondo, bol, copa grande, u otro recipiente, y se llenará de agua clara, dejándolo en algún lugar donde pueda permanecer quieto, y al abrigo del viento. Hecho esto, se corta el corcho con circunspección y poco a poco, hasta que el corcho con el alambre de tal modo sea adecuado, que pueda permanecer a dos o tres pulgadas bajo la superficie del agua, sin ascender ni descender...Sáquese después del agua este mismo alambre, sin desplazar el corcho, y tóquese con la piedra, un extremo con el sur de la piedra, y el otro extremo con el norte, y vuélvase a sumergir en el agua, y se verá que ahora gira por sí solo sobre su centro, exhibiendo la precipitada propiedad de declinación, sin descender hasta el fondo, como por razón debería hacer, si hubiera alguna atracción hacia abajo, por estar el fondo del recipiente más cercano a ese punto, que la superficie del agua.” (Norman,R “The Newe Attractive”, 1581, citado por Harré,1981, pp,46-47)

Este experimento fue repetido por Gilbert en la investigación de la idea del campo magnético terrestre. El experimento para refutar la hipótesis del punto atractivo no tiene jamás una consideración estadística. Podemos pensar en la siguiente la hipótesis nula: la inclinación de la aguja no desciende hasta el fondo, y la hipótesis contraria que afirma lo contrario. El resultado del experimento de Norman pasó la hipótesis nula. Pero esta decisión no tuvo nada que ver con conceptos estadísticos como el nivel de significación y la severidad del experimento. Semejante experimento no requiere el cálculo de las muestras o los datos, y sólo basta la demostración de un fenómeno.³ Hay muchos casos de esta naturaleza en las investigaciones científicas. Veamos brevemente otros dos casos.

En 1670, Malpighi dedujo a partir de una gran cantidad de observaciones microscópicas de plantas, que no solamente debía existir un movimiento ascendente de savia, desde las raíces hacia las hojas, sino igualmente un movimiento descendente que transportaba los materiales de construcción del cuerpo vegetal hasta donde hubieran de ser utilizados. En el año 1716, Stephen Hales hizo una series de experimentos para probar esta idea. Primero tuvo que determinar ¿cuál es la fuerza impulsora del flujo de savia? Había dos hipótesis en aquella época. La primera, que podemos formar como la hipótesis nula H, proponía que la ascensión de agua se debía a la presión en las raíces; la segunda, la hipótesis contraria J, afirmaba que era un proceso de extracción o succión desde las

hojas. Hales enchufó una rama de manzano a un tubo y llenó el tubo de agua. Después sumergió toda la rama en un cubo lleno de agua. Observó que el agua en el tubo descendió 6 pulgadas en las dos primeras horas y 6 pulgadas en la noche siguiente. Al tercer día, Hales sacó la rama del agua y la colgó con el tubo atado a ella al aire libre. Observó que la rama embebió 27.5 pulgadas en 12 horas. Hales concluyó que el experimento mostraba la gran fuerza de la perspiración. Es decir, fue corroborada la hipótesis contraria J.

Otra cuestión permaneció: ¿cuál es el mecanismo de circulación de la savia en una planta? También habían dos puntos de vista al respecto. Una hipótesis sostenía que la savia ascendía por la parte interior del tallo y descendía por la externa; mientras tanto la hipótesis rival sostenía que era un proceso de flujo y reflujo, como el de las mareas. Hales cogió una rama de manzano y la unió fuertemente a un tubo mediante un sifón de plomo, pero antes había recortado la cabeza y el anillo leñoso del último año en una longitud de 3 pulgadas. Después llenó de agua el tubo pero recortando antes una abertura y a través de la corteza y de la madera del último año a 12 pulgadas del extremo inferior del vástago. El agua fue libremente embebida en un minuto. Al cabo de media hora Hales pudo percibir perfectamente que la parte inferior de la abertura estaba más húmeda que antes, mientras que la parte superior de la herida tenía aspecto seco y blanquecido. Con este fenómeno, Hales rechazó la hipótesis rival con el siguiente razonamiento: el agua debía necesariamente ascender desde el tubo, a través de la más interior, porque la madera del año pasado fue eliminada, por espacio de 3 pulgadas en torno a toda la vara; y consiguientemente, si la savia, al seguir su curso natural, descendiera por el último anillo anual, el agua hubiera descendido por la madera del último año, y humedecido primero la parte superior de la abertura; pero, por el contrario, fue la parte inferior la humedecida, y no la superior.

Durante todo el curso del experimento, Hales fue planteando cuidadosamente cada una de sus hipótesis locales para que cada paso fuera efectivamente realizable; sin embargo, ninguna hipótesis local conllevaba la distribución estadística, sino bastaba un fenómeno artificial para obtener un juicio.

Otro ejemplo es el experimento de William Beaumont sobre la digestión. En el siglo XIX, la gente conocía la existencia del jugo gástrico, merced a la observación anatómica, se pensaba, sin embargo, que se trataba tan sólo de una sustancia auxiliar, cuya función era humedecer los alimentos. Fue el cirujano Beaumont, quien aprovechó un accidente para demostrar que esta idea sobre el jugo gástrico no era correcta. En 1833, el soldado Alexis St. Martin fue herido en el abdomen por un tiro de mosquete y atendido por Beaumont. Se recuperó a los 18 meses, pero en la herida se formó un repliegue de doble capa, que cubría el orificio herido, evitando así la necesidad de

las compresas y vendajes para retener el contenido del estómago. Beaumont se valió de esta válvula artificial para hacer su estudio sobre la digestión. Por un lado estudió cómo eran digeridas diversas sustancias en las condiciones naturales del estómago; por el otro, extrajo los jugos estomacales y estudió fuera del cuerpo las condiciones en que ejercían su acción sobre las materias alimenticias. Con el jugo extraído del estómago de St. Martin, Beaumont observó sistemáticamente las velocidades de disolución de diferentes tipos de comida a diferentes temperaturas. Después de varios años de experimentación, Beaumont llegó a la conclusión de que el jugo gástrico no era una sustancia auxiliar, sino un disolvente químico de las materias alimenticias.

A diferencia del experimento de Norman y Hales, el de Beaumont no se basaba en un fenómeno sino en una serie de observaciones sistemáticas; sin embargo, tampoco en este caso se introdujo la metodología estadística. Beaumont jamás se interesó por la distribución estadística, sino por problemas como ¿para alimentos de hasta qué dureza el jugo gástrico puede ser efectivo? ¿a qué temperatura?, o ¿cuál es la duración de la capacidad disolvente del jugo gástrico?

Los tres experimentos que acabo de describir muestran que las hipótesis primarias no conllevan necesariamente la consideración de la distribución estadística como lo exige Mayo. Existen muchos experimentos que no son estadísticos.

Quizá la clasificación usada por Ronald Giere (1984) sobre los tipos de experimentación pueda ser ilustrativa. Según Giere, existen dos tipos de modelos teóricos: el modelo determinístico y el modelo estocástico. Para entender la diferencia entre uno y otro, Giere introduce el concepto de estado de un sistema teórico. El estado de un sistema teórico se define como un catálogo completo de todas las características del sistema en relación a un modelo teórico.⁴ Por ejemplo, el estado del sistema de partículas newtonianas está caracterizado por la masa, la posición y el momento de todas las partículas en un tiempo dado. El sistema de partículas newtonianas es un sistema determinístico porque un estado en un tiempo dado determina los estados de todos los otros tiempos. Para probar la hipótesis de que el sistema solar es un sistema de partículas newtonianas, una predicción del regreso de un cometa puede ser suficiente. Pero el modelo estocástico es diferente. En un sistema estocástico, un estado en un tiempo dado sólo determina la probabilidad de diferentes estados posibles en otros tiempos. Por ejemplo, según el modelo teórico mendeliano los factores hereditarios de los padres no son lo único que determina los caracteres de los hijos. Lo único determinado es la probabilidad de ciertas combinaciones de estos factores. Por eso, en la predicción de Halley sobre el cometa, como en los tres experimentos arriba descritos, no se requiere ninguna consideración estadística, pero en el experimento de Mendel sobre los chícharos sí. (Giere, 1984, pp,92-93; 191-193)

Es discutible que el modelo determinístico en los estudios sobre las investigaciones científicas refiere algo más que un esquema simplificado, conveniente para un texto pedagógico como el de Giere. Se podría plantear la objeción de que hay muchos modelos determinísticos que pueden trasladarse a modelos estocásticos, si incluimos las consideraciones sobre el problema del margen del error. Howson y Urbach explican esta situación:

...deterministic hypotheses are frequently confirmed by evidence that is only assigned some probability, for if it is a quantitative theory, its quantitative consequences may need to be checked with imperfect measuring devices, subject to what is known as experimental error. Take as an example a theory which predicts the position of a planet, this prediction being checked using an appropriate kind of telescope. Because of various unpredictable atmospheric conditions affecting the path of light to the telescope as well as other uncontrollable factors, some connected with the experimenter and some with physical vagaries, the actual reading is acknowledged in experimental work not to be completely reliable. For this reason, if the predicted value of an angle were being ascertained, the result of the measurement would normally be reported in the form of a range of values such as $a \pm b$. Here a is the reading recorded by the instrument, while the interval $a+b$ to $a-b$ signifies the range in which it is judged the true value probably lies. This calculation is usually based on a theory giving the probability that the instrument reading diverges by different amounts from the true value of the measured quantity. (Howson y Urbach, 1996, segunda ed. pp. 7-8)

Esto sin embargo no puede servir de argumento para afirmar que todos los experimentos son finalmente estadísticos como lo hace Mayo. Para Howson y Urbach, la distinción entre las hipótesis determinísticas y las hipótesis estocásticas no significa nada, porque ambos tipos pueden estudiarse con una misma metodología; la bayesiana. Sin embargo, la metodología estadística defendida por Mayo no tiene esta capacidad.⁵ Además, existen casos en los que una hipótesis estocástica puede resolverse de una manera determinística. Tomemos el ejemplo que Good usó para refutar la teoría de confirmación hempeliana. (Good, 1967, p. 322) En la isla X hay 100 cuervos negros y un millón de aves de otro tipo, y en la isla Y hay 1,000 cuervos negros, un cuervo blanco y un millón de aves de otro tipo. Un capitán tiene que distinguir entre cuál es X y cuál es Y por la evidencia de las aves observadas en ambas islas que su buque ha encontrado. El capitán pasa todo el día contando los cuervos negros de ambas islas para ver si la cantidad de una isla es 10 veces mayor que la de la otra. Es un procedimiento estadístico. Pero si de pronto aparece un cuervo blanco, el capitán puede

obtener la conclusión correcta sin ningún razonamiento estadístico; este cuervo blanco permite rechazar la hipótesis de que la isla en que éste vive es X.

Así, la división entre modelos determinísticos y estocásticos es adecuada y señala claramente que no todos los experimentos en la ciencia requieren las consideraciones estadísticas.

2.3 La tradición basada en “imágenes”

Hay un tipo de experimentación que se distancia aún más de la metodología estadística, y poco tiene que ver con la prueba de hipótesis.

Peter Galison en su libro “Image and Logic” (1997) hace un estudio de la historia del desarrollo de la física de las partículas elementales de este siglo. Este se concentran la relación entre los instrumentos experimentales y las teorías. Galison distingue dos tradiciones de invención de instrumentos. Llama a la primera tradición basada en imágenes, cuyo propósito es el de representar los procesos naturales de la manera más fiel y detallada posible; es decir, producir las imágenes lo suficientemente claras para servir de evidencias de una entidad o un efecto. Se trata de una representación mimética de la forma de la naturaleza. Esta tradición enfoca las características particulares y singulares de los objetos de estudio, y en ella basta una representación con buena calidad. Los ejemplos de este tipo de instrumentos en el libro de Galison son la cámara de niebla, la emulsión nuclear y la cámara de burbujas, cuya tarea consiste en señalar las líneas de la trayectoria o la emulsión de las partículas.

La segunda tradición mencionada por Galison es la de las máquinas destinadas a contar cierto tipo de los eventos que a su vez sirven de datos a fin de formular los argumentos estadísticos con los que se afirma la existencia de cierta propiedad en cierta partícula, o su efecto. En esta tradición lo que importan no son las imágenes, sino son las relaciones lógicas preservadas por los registros estadísticos, de manera que las características particulares de cada partícula inevitablemente se pasen por alto, y sólo se destaquen los rasgos comunes. Ejemplos de este tipo de instrumentos son el contador de chispas, la cámara de chispas y la cámara de alambre. (Galison,1997,p,19)

Las dos tradiciones se confrontan, como describe Galison, de esta manera:

Image machines and logic machines--each had their victories, their Nobel Prize-winning successes. At the same time, neither tradition was able to claim a privileged hold on truth for any length of time, and neither held unique sway (for long) over the physics community. Each found its own form of argumentation persuasive and judged the competition to be faulty in certain respects: image physicists argued that logic physicists could be fooled because they missed essential details of the physical

process. Only through the vividness of an image could the causal elements be traced all the way through a process. Ceding the force of such track-by-track arguments, logic physicists nonetheless remained dubious about demonstrations predicated on a handful events. By the 1960s, and logic experimenter could point to a myraid of examples of unrepeated(unrepeatable?) prints that had beckoned experimenters and theorists alike down blind alleys. In the words of one experimenter in the logic tradition, "Anything can happen once"...At every level the traditions clashed: golden event versus statistical demonstration, the objectivity of passive registration versus the persuasiveness of experimental control, vision versus numbers, and photography versus electronics. (Galison, 1997.pp.24-25)

El libro de Galison nos provee de una narrativa brillante sobre cómo, durante todo el siglo, estas dos tradiciones en la física de partículas elementales, competieron y se colaboraron sin perder sus propias características.

Sin embargo, la división de las dos tradiciones en el libro de Galison se limita a la invención, manipulación y utilización de instrumentos experimentales. Creo que sería muy plausible ampliar esta clasificación a toda la experimentación. Dicho en otras palabras, los experimentos se podrían clasificar en dos tipos: por un lado, el tipo lógico cuyo propósito es buscar los argumentos estadísticos para ciertos efectos o entidades, tal como Mayo plantea; y por otro lado, el tipo basado en una imagen, cuyo propósito es representar el proceso de la naturaleza, donde no tiene sitio la estadística. La existencia del tipo lógico no es polémica, pues todos los ejemplos de Mayo son de este tipo. Aquí trataré de encontrar los ejemplos históricos para demostrar la plausibilidad de la existencia general en la ciencia de experimentos del tipo basado en imágenes.

No es extraño que los primeros experimentos u observaciones deliberadamente planteadas en la historia de la ciencia sean precisamente del tipo basado en imagen. Conocer los rasgos determinantes de los procesos naturales es el primer motor de la ciencia. Un ejemplo es la observación de Aristóteles de desarrollo del polleulo descrita en su obra *Historia Animalium*. Fue un texto hipocrático, el primero en plantear la idea de que la comparación de las especies no humanas servía de vía para investigar la reproducción de los seres humanos. En *Sobre la Naturaleza del Niño* se encuentra la siguiente comparación entre el proceso embrionario de un ave y el de un hombre:

Tómanse veinte o más huevos, y pónganse a empollar con dos o más cluecas. Después, a partir del segundo día de incubación, hasta el último, el de la eclosión, se va retirando diariamente un huevo, que se abre, para examinarlo. Se encontrará que todo

concuera con lo que he dicho, hasta el extremo de que la naturaleza de un ave debería ser comparada con la de un hombre. (citado por Harré,1981, p,26)

Aristóteles adopta la técnica del experimento y la observación hipocrática, pero nos cuenta todos los pormenores:

...Cuando el huevo tiene ya diez días, el polluelo y todas sus partes son visiblemente distinguibles. La cabeza sigue siendo mayor que el resto de su cuerpo, y los ojos, mayores que la cabeza, pero todavía carentes de visión. Si hacia este tiempo se le extraen los ojos, se encuentra que son mayores que habichuelas, y negros; si se retira la capa de cutícula, se encuentra en su interior un líquido blanco y frío, que reluce intensamente a la luz del sol; pero sigue sin haber sustancia dura ninguna...En este momento son también visibles los órganos internos más grandes, así como el estómago y la disposición de las vísceras; y las venas que parecen proceder del corazón están ahora cercanas al ombligo. Desde el ombligo parten proceder un par de venas; una hacia la membrana que envuelve a la yema, y la otra, hacia esa membrana que envuelve globalmente la membrana en cuyo interior yace el polluelo, la membrana de la yema, y el líquido intermedio...

...Hacia el vigésimo día, si abrimos el huevo y tocamos el pollo, se remueve en el interior del cascarón y pía; y comienza a estar ya cubierto de plumón, cuando, tras haber pasado el vigésimo día, el pollito comienza a romper el cascarón... (libro 6, 561a3-562a20)

Veo en este caso un el ejemplo del origen del tipo de experimentación basado en imágenes. La intención de Aristóteles y del texto hipocrático era describir un proceso natural. Sin embargo, una idea muy popular en la filosofía de la ciencia es la que contrapone las puras observaciones con los experimentos. Según esta idea, un observador como Aristóteles permanece al margen, sin influir en el desarrollo de los acontecimientos por los que se interesa. Por otro lado, un experimentador interviene activamente en el curso de la naturaleza. Se trata de una distinción discutida por los aristotélicos y los baconianos en los siglos XVI y XVII. Claro que existe la diferencia entre las observaciones en la ciencia aristotélica y las que los partidarios de la revolución científica propusieron al inicio de la experimentación moderna. Pero en este caso de la embriología del polluelo, los observadores antiguos no fueron totalmente pasivos, (haciendo sólo el trabajo de registrar) sino que también intervienen. Como escribe Harré:

La ciencia griega fue en gran medida exploratoria y teórica. Pero en el uso controlado de la serie de huevos tenemos un ejemplo de técnica investigadora que

presupone cierta interferencia y planificación. Aristóteles no estuvo esperando pasivamente a que se le fueran presentando las etapas del desarrollo de polluelo, sino que intervino activamente en el proceso natural, según el ingenioso proceder sugerido por el autor hipocrático. (Harré,1981,p.31)

En la naturaleza existen muchos fenómenos cuyas causas son tan alejadas como para poderlas observar directamente. Una función del experimento basado en imagen es construir una imitación de estos fenómenos que permita indagar sus causas. Un ejemplo clásico es el experimento de Teodorico de Friburgo sobre las causas del arco iris. Aristóteles había planteado en su *Meteorologicos* que la aparición del arco iris se debía a la reflexión de la luz en gotas de lluvia recién formadas. La mayoría de los eruditos medievales creía que la forma circular del arco iris es meramente una reflexión del disco circular del Sol en el que las gotas de lluvia, al caer, actúan a modo de espejo. Fue Alberto Magno el primero en formular la teoría de que el arco iris se produce al interactuar la luz con cada una de las gotas. Teodorico hizo un experimento para imitar el mecanismo de este proceso. Aunque él no estaba de acuerdo con la idea de Alberto Magno de que los colores se producen en el seno de la cortina de gotas, por los efectos de algún tipo de estratificación. Primero, Teodorico usó un prisma hexagonal para decidir el orden de los colores. El color rojo se produjo más cerca de la superficie, en la parte más externa del prisma, mientras que el azul, el más oscuro, se vio en las profundidades, donde el medio es de máxima opacidad. Luego, para imitar el proceso de la formación de un rayo de arco iris, Teodorico utilizó un gran globo de cristal lleno de agua como modelo de una gota de lluvia, y observó que la luz se refractó en la superficie de la esfera y se reflejó en el interior. También se comprobó el orden de los colores en un arco iris. Durante el proceso de la refracción, el rojo se produce más cerca de la superficie, y el azul surge más hacia el interior. Pero cuando estos rayos rebasan la superficie interior, debido a la reflexión, el orden se invierte y se ve como aparece un arco iris. Con este modelo de globo del cristal, y bajo su suposición de que los rayos se refractan y se reflejan en cada gota individual, Teodorico pudo explicar todas las características, salvo el tamaño, de un arco iris. (Pedersen, 1993,p,177)

De hecho, la meteorología es una ciencia que conlleva muchos experimentos basados en imágenes, porque sus objetos de investigación (cosas como el viento, la lluvia, el trueno y el relámpago etc.,) son fenómenos cuyas causas no son directamente observables para la gente. En los siglos XVIII y XIX, la meteorología era un bastión de la tradición experimental basado en imágenes, y la invención de la cámara de niebla, según Galison, es una consecuencia directa de tal tradición.

En la época de la Ilustración, los logros de la matematización de la física estimularon el espíritu de las investigaciones cuantitativas en otras áreas. Los meteorólogos enfocaron su atención

hacia la invención de nuevos instrumentos, y hacia el perfeccionamiento de los viejos, para aumentar la precisión y la confiabilidad de las mediciones. (Feldman,1990,pp.143-146) Pero la meteorología nunca se convirtió totalmente en una ciencia abstracta demostrativa y reduccionista como los filósofos naturales pretendían. Conservó mucho del espíritu de la historia natural, o la ciencia “morfológica” como la llamó Goethe. Para los filósofos naturales, la meta de los experimentos era extraer las leyes universales a partir de los datos particulares, de modo que se obtuvieran las verdades más transcendentales que estaban tras los fenómenos. Pero, siguiendo la tradición de la historia natural, impulsado por Goethe y Humboldt, muchos meteorólogos y científicos:

realized that there were times when the abstract scientific method was inadequate. Difficulties arose because the natural philosopher who exploited analysis and abstraction exclusively was “forcefully reminded that he was in danger of dealing not with natural, but with artificial things. Instances are plentiful where through the elaboration of fanciful theories, the connection with the real world has been lost”...Opposing the “one-sided” working of abstract science lay another ideal of investigation, embodied in the morphological sciences. These sciences were motivated, as Merz somewhat rhapsodically put it, by “the genuine love of nature, the consciousness that we lose all power if, to any great extent, we sever or weaken that connection which ties us to the world as it is---to things real and natural: it finds its expression in the ancient legend of the mighty giant[Antaeus] who derived all his strength from his mother earth and collapsed if severed from her.”
(Galison,1997,p,79)

A través de esta predilección realista, los experimentos basados en imágenes encontraron su base filosófica: capturar la naturaleza no por medio de la búsqueda transcendental de causas sino por la imitación de los fenómenos naturales.

El geólogo M. Rudwick se resistió con tenacidad a los estudios de laboratorio por su incapacidad de reproducir las condiciones naturales. Para él, las investigaciones en el campo eran la única dirección correcta para la geología. Pero ya en la segunda parte del siglo XIX, los geólogos empezaron a subrayar la importancia de los experimentos de laboratorio, los cuales, sin embargo, lejos de ser estudios abstractos, heredaron el espíritu naturalista de imitación de los fenómenos. Como afirma Galison:

But instead of shunning experiment, late-nineteenth-century morphological scientists began to use the laboratory to reproduce these same natural occurrences. The morphologists strove to make laboratory versions of real phenomena with all the

richness the cylopes or glaciers had in nature, By re-creating nature in the controlled world of the laboratory, the scientists hoped to discover the physical processes underlying the natural world. It is this attempt to imitate nature in the small the I will refer to as the "mimetic tradition"; when the morphological sciences enter the laboratory, mimesis becomes their characteristic form of representation. (Galison,1997. p.80)

En la época victoriana, mucha gente era aficionada de la naturaleza y usaba la pintura, la poesía, la fotografía o el laboratorio para reproducir los fenómenos naturales. C.T.R. Wilson, el inventor de la cámara de niebla, fue precisamente uno de aquellos aficionados de la naturaleza. De joven, le gustaba viajar con su hermano por las montañas escocesas y fotografiarlas. En su época, la erupción del volcán Krakatoa en 1883 y la niebla producida por la industria, entre otros fenómenos, estimularon los estudios de los fenómenos meteorológicos en Inglaterra. En 1890, Wilson estableció, con otros voluntarios, un observatorio en la montaña Ben Nevis, a fin de estudiar la formación de la niebla. En 1895, diseñó y construyó una cámara de niebla para imitar el proceso de formación de la niebla. En 1911, inspirado en la fotografía instantánea usada por otros científicos, decidió utilizar esa técnica para capturar las escenas en la cámara de niebla, logró de ese modo observar las trayectorias de los átomos.

En su libro Galison hace un estudio excelente, y muy detallado, sobre el proceso de la invención de la cámara de niebla. No voy a repetir aquí sus detalles. Lo que es claro por su exposición es que la invención de Wilson es excelente un ejemplo de la tradición experimental basado en imágenes. Se trata de una invención que no provino de la física, pero que se convirtió en un gran instrumento para el desarrollo de la física de partículas elementales y de la física nuclear de este siglo. Galison afirma:

The historical reconstruction of mundane as well as transcendental physics is necessary to set Wilson's cloud chamber work in its properly Victorian context. That world will emerge as a startling source of modern physics, revealing in Wilson's work a coherence that is entirely lost when it is divided into "meteorology" on the one hand and "physics" on the other. Lying between or perhaps in the intersection of the two domains is an area better called "mimetic experimentation", a term that will designate the attempt to reproduce natural physical phenomena, with all their complexity, in the laboratory. (Galison,1997,pp.74-75)⁶

A continuación, Galison estudia cómo la tradición basada en imágenes, al ser una tradición de construcción de instrumentos experimentales, y al confrontar la tradición basada en vínculos lógicos, ha sido una parte crucial de la experimentación durante el desarrollo de la física de este siglo.

La incuestionable existencia de la tradición experimental basada en imágenes manifiesta, una vez más, que la idea de los experimentos como fundamentalmente estadísticos, planteada por Mayo, no captura todo lo que concierne a la experimentación. Existen otro tipo de experimentos, que no están contemplados en la filosofía de la experimentación de Mayo.

En este capítulo, hemos explorado algunos problemas del modelo primario de la jerarquía de los modelos experimentales de Mayo. He señalado que, por un lado, el modelo primario no es plausible, porque la formación de las hipótesis locales, requerida por el modelo primario, está inevitablemente bajo el control de las teorías globales; y por el otro, el modelo primario no es suficiente, porque existen muchos tipos de experimentación que no son estadísticos ni están reflejados en la filosofía de la experimentación de Mayo.

Notas

1, ¿Por qué la jerarquía de los modelos experimentales de Mayo no es la versión débil? En Muchas ocasiones Mayo defiende un estudio local y de grados sobre los problemas experimentales. Estoy de acuerdo con que la metodología estadística defendida por Mayo es una cuestión local, pero el problema está en la intención de Mayo de ampliar esta metodología a toda la filosofía de la ciencia. Como hemos visto Mayo intenta utilizar esta metodología para resolver los dificultades en la inferencia científica, y dedica a este tema la mayor parte del espacio de su libro. El modelo primario es una puente que relaciona los experimentos locales con los problemas del nivel alto. La idea de Mayo consiste en que esta relación se da por las consideraciones estadísticas como los casos de Perrin y Eddington. En este sentido, los modelos de Mayo no son ejemplos de la versión débil sino la fuerte. Mi propósito no es el de oponerme a la relación entre las teorías de nivel alto con los experimentos locales, sino a una relación unívoca y dominada por las teorías.

2, Es correcto que los nuevos experimentalistas deben poner más atención sobre el uso de la estadística en los experimentos. Pero lo que ellos van a sostener son: (1) la inferencia estadística no es el único o el más importante razonamiento en el proceso experimentales; (2) la inferencia científica no se basa puramente en la inferencia estadística.

3, Sin embargo, ambos experimentos, el estadístico y el no-estadístico, comparten un mismo carácter: el razonamiento es inductivo de manera que el resultado no es lógicamente conclusivo. Galileo, usando la voz de Salviati, elogió el experimento de Gilbert, el cual repitió el de Norman acerca de la piedra imán, aunque hubiera deseado que Gilbert "que fuese matemático y en especial bien formado en geometría, cuya práctica le habría tornado menos dispuesto a aceptar como pruebas rigurosas las razones que aduce como causas de las conclusiones que él mismo ha observado. Estas razones, hablando claramente, no tienen el rigor indudablemente requerido para obtener conclusiones naturales, necesarias y eternas." (Galileo,1975,p.432)

4, El concepto de modelo teórico es más o menos el mismo que el que propone van Frassen, pero expresado de una manera más clara: "A theoretical model, then, is a kind of system. But it is a very special kind of system. Such systems are not necessarily found in nature. Rather, they are primarily creatures of language. We create them by formulating statements and arranging the statements into a definition of the kind of systems we intend. Thus, a theoretical model is a kind of system whose characteristics are specified by an explicit definition. By "definition" I mean a stipulation as to how certain terms are to be used." (Giere,1984, p,80)

5, Aunque bajo la metodología bayesiana la distinción entre las hipótesis deterministas y estocásticas no tiene mucho sentido, Howson y Urbach señalan que la distinción existe en la vieja filosofía de la ciencia y la estadística: "Philosophers of science have traditionally concentrated attention primarily on deterministic hypotheses, leaving statisticians to discuss the methods by which statistical or nondeterministic theories should be assessed. Accordingly, a large part of what would more naturally be regarded as philosophy of science is normally treated as a branch of statistics, going under the heading 'statistical inference'. So it is not surprising that philosophers and statisticians have developed distinct methods for their different purposes." (Howson y Urbach,1996,segunda edi. p,115)

6, Galison no omite la influencia de la tradición de lógica sobre la invención de Wilson. La diferencia entre la cámara de niebla de Wilson y la de Aitken era precisamente por la consideración de Wilson sobre la teoría de ionización de la escuela de Cavendish encabezada por Thomson. (Galison,1997,pp.97-104) Además, la teoría de Bragg sobre el rayo de X refuerza la confianza de Wilson para identificar las trayectorias de los átomos. (Galison,1997,pp.113-114) Pero, los trabajos de Wilson en su laboratorio eran, evidentemente, de la tradición de imagen.

Capítulo 3 Las Dificultades para Modelar la Inferencia Causal

A partir de este capítulo, examinaré la manera en que Mayo plantea el problema de la inferencia científica, y haré ver dos limitaciones importantes de su propuesta. En primer lugar, la metodología por partes no es suficiente para modelar toda inferencia causal. En segundo lugar, el problema del nivel de significación, un problema clásico en la estadística, impide una “metodología objetiva” en el sentido que la autora intenta formular.

En este capítulo analizaré el primer punto, señalando el planteamiento de Mayo sobre la metodología estadística para una teoría suficiente de inferencia causal, tanto en el nivel del modelo de los datos como en el diseño experimental. Haré ver que esto proyecto no es plausible, porque no puede afrontar el problema de clases de referencia. Posteriormente, estudiaré cómo las teorías de la inferencia causal de Nancy Cartwright y Peter Lipton tratan de resolver el problema de las clases de referencia. La comparación entre el proyecto de Mayo y las teorías de Cartwright y Lipton nos puede iluminar qué es lo que Mayo deja de lado.

3.1 La estadística del error y los métodos para modelar inferencias causales

Como hemos visto en el capítulo 1, para Mayo, el modo informal del conocimiento experimental es la argumentación a partir de los errores, y el modo formal es la estadística del error. Según la autora, la estadística del error tiene la capacidad de eliminar los siguientes cuatro tipos canónicos de errores en los experimentos científicos:

- (1) el error de los efectos aleatorios y las correlaciones espurias;
- (2) el error de un parámetro;
- (3) el error de los factores causales y
- (4) el error de los presupuestos experimentales. (Mayo, p,18; p,449)

De éstos, los errores (1) y (3) se vinculan con inferencia causal. Por lo tanto, para que el proyecto de Mayo tenga éxito debe, por lo menos, ser capaz de tratar estos dos tipos de errores. Además, si Mayo puede evitar estos dos tipos de errores, entonces estaría muy cerca de poder modelar la inferencia causal.

Cuando analiza el proyecto popperiano de falsificación, Mayo señala acertadamente la problemática de Popper al decidir si el resultado negativo de un experimento se debe a la hipótesis bajo prueba o a las hipótesis auxiliares. Mayo afirma que eso no es un problema para ella, porque las

metodologías por partes y la estadística pueden ser efectivas para evitar los errores sobre los factores causales. Ella dice:

What is overlooked [by Popper] is the way in which experiments may be designed to deliberately isolate the effect of interest so that only a manageable number of causal factors (or types of factors) may produce the particular experimental outcome. Most important, literal control is not needed: one only find ways of arguing so as to avoid the erroneous assignment of the cause of a given effect or anomaly...Another example would be to mimic the strategy of randomized treatment-control studies. The myraid of possible other causes--even without knowing what they are--are allowed to influence the treated and the control groups equally. In other cases, substantive alternative causes cannot be subtracted out in this manner. Then severe tests against hypotheses that these causes are responsible for the given experimental effect must be carried out separately.
(Mayo, pp.15-16)

Como se puede ver, Mayo sostiene que la estadística del error tiene la capacidad de legitimar inferencias científicas, a través de cuatro métodos fundamentales:

- (1) el método del control;
- (2) el método del tratamiento aleatorio (*randomized treatment*);
- (3) el método del balance de los posibles factores causales, y finalmente
- (4) el método de la prueba de la severidad.

Recordemos que estos métodos son una parte importante del modelo de los datos y del diseño experimental. Como ya hemos visto en el capítulo 1, la tarea del modelo de los datos es generar datos que satisfagan los presupuestos experimentales requeridos por el modelo experimental por medio de una prueba estadística. El modelo de los datos se relaciona con el modelo experimental a través de consideraciones estadísticas, pues las hipótesis estadísticas del modelo experimental determinan las características y los requisitos experimentales para la generación de datos. Por ejemplo, si un modelo primario considera que la media de una población es μ , y el modelo experimental hace un diseño de un experimento cuya muestra tiene una media \bar{X} , entonces, el modelo de los datos tendrá que encontrar los datos que tengan este \bar{X} como la media. (sobre los símbolos, véase el apartado 1.3) Lo que Mayo llama “presupuestos experimentales” son los requisitos experimentales que aseguran que las muestras en un experimento representan de la manera más fiel posible a la población. En el apartado 1.3, con el ejemplo del experimento sobre la relación entre el fumar y el temblor de las manos, señalé brevemente los métodos que se hacen sobre el diseño experimental en el proceso de la generación de los datos con el fin de satisfacer los requisitos. Según Mayo, algunos de estos cuatro

métodos mencionados en el párrafo anterior son los recursos más efectivos para verificar los factores causales en la inferencia a las causas. Es decir, para evitar los errores en las inferencias causales, si usamos la terminología del argumento a partir del error como el punto de vista para enfocar este tema.

Para profundizar en estos métodos, retomaré el ejemplo del fumar y el temblor de las manos, pero de manera menos formal. Los médicos quieren saber si fumar es la causa del temblor de las manos. Para determinarlo, el modelo primario plantea la siguiente idea esquemática: primero, los médicos toman dos grupos de personas como muestra, uno compuesto por las que fuman y el otro por las que no fuman. Piden al primer grupo que fume mientras que en el segundo grupo nadie fuma. A continuación, registran el nivel del temblor de las manos de los integrantes de ambos grupos. Si el nivel promedio de temblor del grupo de lo que no fuman es significativamente menor al grupo de fumadores, podemos concluir que fumar es una causa del temblor. El modelo experimental plantea la hipótesis nula que afirma que fumar no es la causa y la hipótesis contraria que afirma que si. Además, el modelo experimental tiene que decidir el nivel de significación para determinar qué diferencia mínima, entre los registros de los dos grupos, será suficiente para permitir una elección entre las dos hipótesis. También conlleva presupuestos experimentales para el modelo de los datos, cuyo objetivo es asegurar que las muestras de estos dos grupos tengan suficiente carácter representativo de la población en cuestión. Por ejemplo, el número de personas en cada grupo, no puede ser demasiado pequeño porque no es representativo, pero tampoco es necesario que sea de millones. Un número razonable y realizable se decide por el modelo de los datos.

Otro problema que se resuelve por medio de presupuestos experimentales es el de los factores extraños, o las variables extrañas, factores no controlados que puedan tener influencia sobre los resultados de una prueba. Por ejemplo, ciertos tipos de enfermedades aumentan el nivel de temblor, y si algunos integrantes del grupo de fumadores sufren alguna de estas enfermedades, la media del registro del nivel de temblor de este grupo será naturalmente mucho más alta que la del otro grupo, pero no podemos deducir de ello que el fumar causa el temblor porque no sabemos si el alto nivel del registro de este grupo es causado por el fumar o por dichas enfermedades. Una manera de evitar estas variables extrañas consiste en los controles mencionado por Mayo en los párrafos anteriores. El control es una metodología que puede corregir, neutralizar o eliminar las influencias de ciertas variables extrañas que pueden cambiar los efectos al no ser controladas durante el proceso experimental. En el caso de nuestro ejemplo, el control consiste simplemente en analizar la historia clínica de cada integrante de los dos grupos antes de la prueba y excluir a aquellos que padecen de estas enfermedades, aunque en el capítulo 1 me he referido a un control un poco más complicado.¹

Para eliminar las influencias psicológicas de cada miembro de los dos grupos, los experimentadores hicieron un tipo de pipa muy parecida a la pipa real, pero que no contenía tabaco, sino sólo aire húmedo y tibio. Los integrantes del grupo que no fumaba usaban este tipo de pipa y creían que estaban fumando tabaco. Mediante este control de “placebo” y de “doble ciegos” (*double-blind*), usado frecuentemente en las investigaciones médicas, se eliminó la influencia psicológica.

Muchas veces no se logra eliminar todas las variables posibles. Por ejemplo, sabemos que la edad, el sexo y la diferencia geográfica pueden ser variables: la gente de edad más avanzada tiene mayores probabilidades de tener temblor de las manos; es más fácil que lo tengan los hombres que las mujeres; o quizás los nortefños tienen mayor nivel de temblor que los sureños. Pero los experimentadores no quieren eliminar estos elementos por medio de los controles, porque si excluyen a las personas de determinadas edades, a los hombres o a los nortefños en el experimento, los que se quedan no tendrán la capacidad de representar a toda la población requerida por los presupuestos experimentales. En este caso, el experimento tiene por objetivo saber si el fumar es la causa del temblor para toda la gente, y no sólo para las fumadoras femeninas, jóvenes, adultas y sureños. Para afrontar esta dificultad, Mayo propone la metodología del tratamiento aleatorio y del balance. La idea es distribuir por igual estas variables en los dos grupos. Ambos grupos tienen el mismo número de hombres, mujeres, sureños y nortefños; también tienen la misma proporción de jóvenes, adultos y personas de edad avanzada. Con este método, las variables posibles tienen el mismo nivel de influencia sobre ambos grupos, de manera que los experimentadores pueden fácilmente formarse un juicio sobre los factores que quieren conocer.

Todavía existe otra posibilidad. Imaginemos que la causa verdadera del temblor de las manos no es el fumar, sino cierto tipo de tabaco que contiene cierta sustancia química. En este caso, Mayo propone otro método que hace pruebas de alto nivel de severidad sobre diferentes tipos de tabacos para ver sus efectos sobre el temblor. Esta metodología ayuda a los experimentadores a distinguir las causas falsas de las reales. Obviamente esta metodología no se limita al modelo de los datos, sino que requiere que los experimentadores planteen otros experimentos que tengan su propio modelo primario, modelo experimental y modelo de los datos. En este tipo de experimento, el juicio sobre cuáles serán las causas reales se forma usando el criterio de la severidad calculada cuantitativamente. Dice Mayo:

Quantitative knowledge teaches not only about the existence of certain entities but also about the properties of the process causing the effect. (Mayo, p.44)

Mayo cree que por medio de la prueba de la estadística del error y de los cuatro métodos para determinar los factores causales, es decir, los controles, el tratamiento aleatorio, el balance, y las

pruebas de alto nivel de severidad, se pueden resolver los problemas relacionados con la determinación de los factores causales y por lo tanto con la caracterización de inferencias causales correctas. Para la autora, estos métodos se convierten en modelos canónicos por su confiabilidad en la inferencia a las causas, o más específicamente, en la eliminación de los errores sobre los factores causales. Tomando el ejemplo de los controles, dice ella:

An experimental argument based on an ideal case of literal experimental control illustrates the notion of a canonical model. It is an exemplary case of how a causal inference may be warranted with high (if not maximal) reliability. A canonical model for a certain type of claim is the basis for two kinds of spin-off strategies: first, arriving at further canonical models, and second, discovering and scrutinizing methods for satisfying the assumptions of these models. As an example of the first strategy, the ideal of perfect control gives rise to exemplary arguments that approximate or simulate the ideal case, such as arguments based on control groups (e.g., randomized treatment-control studies). Having justified the approximation to the ideal case, the stage is set for the second type of strategy. As an example of the second strategy, consider how rules about keeping subjects and experimenters blind, using placebos, and so on, stem directly from the goals of a valid comparison of the treated and untreated groups. (Mayo, p. 453)

Además, para Mayo, la inferencia sobre los factores causales es una parte de la inferencia científica cuya objetividad es garantizada por la estadística del error. Los cuatro métodos tratan de satisfacer los presupuestos experimentales, y como hemos visto en el capítulo 1 sobre el modelo de los datos, estos presupuestos experimentales son requeridos por las consideraciones estadísticas provenientes del modelo experimental, las cuales para Mayo construyen los típicos razonamientos objetivos.

Me parece que son demasiado optimistas las siguientes dos conclusiones de Mayo: La primera afirma que los cuatro métodos en los diseños experimentales y la generación de los datos junto con la estadística del error son suficientes para caracterizar totalmente las inferencias causales, y la segunda considera que la inferencia causal hecha por medio estos métodos y la estadística del error es un inferencia totalmente objetiva, sin ninguna consideración sobre los contextos de fondo. Creo que son demasiado optimistas porque Mayo ha olvidado un problema grave para modelar la inferencia causal: el problema de las clases de referencia. Si el proyecto de Mayo no tiene capacidad de afrontar el problema de las clases, eso significa que no es capaz de evitar los dos tipos de errores, por lo tanto no es suficiente para modelar inferencia causal. Esta limitación del proyecto de Mayo implica dos

puntos importantes. En primer lugar, la inferencia causal, como veremos, es un tema más complicado de lo que Mayo plantea. Como veremos abajo, una consideración importante a tomarse en cuenta son las dificultades generadas por el problema de las clases de referencia. Mayo, como veremos, ignora el problema, y esto se refleja en problemas para su propuesta. En segundo lugar, voy a argüir que contrario a la que Mayo sugiere, la inferencia estadística requiere enmarcarse en contextos. Cuando comparemos las teorías de Nancy Cartwright y Peter Lipton sobre inferencia causal con la de Mayo (cosa que haremos en las secciones 3 y 4) veremos las dificultades de la metodología de Mayo.

En la siguiente sección, vamos a ver cuál es el problema de las clases de referencia, y porqué el proyecto de Mayo no tiene capacidad de afrontarlo.

3.2 El problema de las clases de referencia

El problema de las clases de referencia viene de un requisito de la explicación estadística-inductiva defendida por Hempel. La explicación estadística-inductiva (la explicación EI) es un argumento inductivo que afirma que cuando la probabilidad de un suceso a explicarse (explanandum) E, dado ciertos elementos o factores explicativos (que constituyen el explanans) C, es muy alta, y tenemos "a", que es un caso de C simbolizada como Ca, podemos tener una alta confianza en que "a" también es E simbolizada como Ea. (Hempel, 1965, p,390)

Existe un problema al que Hempel considera como el problema de la ambigüedad de la explicación EI: a partir de las premisas compatibles se puede sacar conclusiones que se contradicen. Por ejemplo, Juan (j) tiene la infección de estreptococo Sj y toma penicilina Pj. Sabemos que la probabilidad de que una persona que toma penicilina se recupere es muy alta, digamos .9($\Pr(R/P)=0.9$), donde R se refiere al suceso de recuperación, podemos concluir, de acuerdo al modelo de la explicación EI, de que Juan va a recuperarse Rj(i.e que el suceso Rj va a tener lugar). Pero Hempel dice que puede suceder algo más. Supongamos que Juan también tiene una enfermedad S*j, para la cual tomar penicilina es contraproducente. En este caso la probabilidad de su recuperación va a ser muy baja aunque tome penicilina. Entonces, tenemos $\Pr(R/P \& S^*=0.1)$, que nos lleva a una conclusión contradictoria. (Hempel, 1965, p,383; p,394) Este es el problema de las clases de referencia.

Para evitar el problema, Hempel impone a una explicación EI el requisito de la especificación máxima: se requiere que las condiciones iniciales en la explicación incluyan a toda la información relevante en una situación dada. En el caso nuestro, podemos dividir toda la clase de la gente que tiene la infección en dos subclases: la que toma penicilina y la que no la toma. Pero, esta clase de

referencia no permite una explicación correcta, porque no cumple el requisito de la especificación máxima, es decir, no toma en cuenta otra información relevante que es la enfermedad S*. Hay que dividir la clase de referencia otra vez entre la gente que tiene S* y la no la tiene. El requisito de la especificación máxima requiere que cada subclase, después de la partición, sea homogénea en que no subdivisión se necesita (no existe otra información relevante). Así, una clase de referencia es homogénea cuando no hay manera de dividirla más.

La consecuencia directa del requisito de la especificación máxima es la relatividad epistémica de la explicación estadística. Porque el nivel de la partición homogénea depende de los conocimientos actuales, es decir, cuando no podemos dividir más las subclases, ello se debe a que, según los conocimientos de que disponemos ahora, no podemos hacerlo, pero esto no significa que en el futuro no podemos lograrlo cuando tengamos más conocimientos nuevos. Dice Hempel:

The preceding considerations show that the concept of statistical explanation for particular events is essentially relative to a given knowledge situation as represented by a class K [Kt is the class of statements asserted or accepted by empirical science at time t. The class of accepted statements will be referred to simply as K when specific reference to the time in question is not required (pp,395-396)] of accepted statements. Indeed, the requirement of maximal specificity makes explicit and unavoidable reference to such a class, and it thus serves to characterize the concept of "I-S explanation relative to the knowledge situation represented by K". (Hempel, 1965, p,402)

Este es el punto que Alberto Coffa y Wesley Salmon quieren atacar. Coffa señala que debida a la relatividad epistémica, la explicación EI no puede ser un argumento inductivo. Según Coffa, la explicación EI como un argumento inductivo debe ser una analogía de la explicación deductiva y nomológica. La diferencia entre las dos es que en lugar de una ley universal, el argumento inductivo toma la ley estadística que es verdadera. Pero debido al problema de las clases de referencia, la relatividad epistémica de la explicación EI tiene como consecuencia la posibilidad de que, desde la premisa verdadera, llegamos a una conclusión falsa, en caso de que existan otros elementos relevantes. (Coffa, p,76)

Algunos filósofos consideran que es necesario que se introduzca el concepto de causalidad en el modelo de la explicación estadística. La explicación EI puede compartir la idea clásica de causalidad: una cosa es la causa de (y explica) la otra, cuando la primera da una alta probabilidad a la segunda. Salmon propone el proyecto de la explicación de referencia estadística (la explicación RE) como una alternativa. A diferencia de la explicación EI, la explicación RE no requiere que el

factor explicatorio tenga una probabilidad alta sobre el factor a explicarse, sino sólo que cambie la probabilidad de este, es decir, C explica E sólo en caso de que $\Pr(E/C) > \Pr(E/-C)$, o $\Pr(E/C) > \Pr(E)$, o $\Pr(E \& C) > \Pr(E)\Pr(C)$. Como consecuencia, requiere que E y C estén estadísticamente correlacionados. Pero la correlación no implica la causalidad. Un factor G, que no es la causa de E, puede ser correlativo de E, en caso de $\Pr(E/C \& G) = \Pr(E/G)$ y $\Pr(E/C \& -G) = \Pr(E/-G)$. Esto también es un problema de las clases de referencia. Para evitar las situaciones como estas Salmon establece la idea de que una causa verdadera "C filtra(*screen off*) el factor irrelevante G para una explicación causal sobre un efecto E". Salmon toma un ejemplo:

Screening off is frequent enough and important enough to deserve further illustration. A study, reported in the news media a few years ago, revealed a positive correlation between coffee drinking and heart disease, but further investigation showed that this correlation results from a correlation between coffee drinking and cigarette smoking. It turned out that cigarettes smoking screening off coffee drinking from heart disease, thus rendering coffee drinking statistically (as well as causally and explanatorily) irrelevant to heart disease. (Salmon, 1984, p.44)

En este caso, el fumar filtra un factor irrelevante que es tomar café para identificar la causa del ataque cardíaco. Este es el tipo de problemas que hoy día preocupan a los filósofos respecto de la causalidad. Por ejemplo, la paradoja de Simpson (Cartwright, 1983, p.24; von Frassen, 1980, pp.148-150; Salmon, 1984, p.39). Hempel no tiene este problema, porque el requisito de la partición homogénea evita estos casos. Pero el costo es la relatividad epistémica. Salmon rechaza el proyecto de Hempel no sólo por la consecuencia de la relatividad epistémica, sino también por los problemas en la partición de las clases de referencia. Salmon llama a la partición de Hempel "partición epistémicamente homogénea". Esta es aquella en que nuestro conocimiento no nos dice nada sobre los elementos relevantes en una partición. Sin embargo, según Salmon, en una explicación causal, sólo debemos tomar en cuenta los elementos relevantes por que los elementos irrelevantes no son homogéneos con la partición relevante que nos interesa. (Salmon, 1989, pp.63-64f) Por eso, Salmon plantea su explicación estadístico-relevante RE que sólo requiere una partición objetivamente homogénea. En una partición de este tipo sólo caben los elementos relevantes que son homogéneos, de manera que los factores irrelevantes ya han sido filtrados (*have been screened off*). Sin embargo, esta solución no carece de problemas.

La explicación RE puede explicar por qué el fumar es la causa del ataque cardíaco y tomar café no lo es, por medio de la partición objetivamente homogénea, pero no nos dice nada sobre de qué manera podemos conseguir esta partición objetivamente homogénea. Antes de tomar en cuenta el

factor del fumar, los científicos habían considerado que ellos habían tenido una partición P1 que era objetivamente homogénea en ese momento: la partición en que tomar café es la causa del ataque cardíaco. Pero, después de incluir el factor del fumar, los científicos tenían otra partición P2 que es totalmente diferente a P1. Hasta hoy, ellos han considerado que P2 es objetivamente homogénea porque todos los factores relevantes han sido incluidos. Pero esto sólo es relativo al conocimiento que ahora tenemos. Pero ¿por qué no hay posibilidad de que en el futuro encontremos otros factores relevantes que ahora no tomamos en cuenta, y tengamos otra partición P3, que sea totalmente diferente a P1 y P2? Es decir, ¿por qué el fumar no puede ser filtrado (*screened off*) como irrelevante por factores de los que sólo sabremos en el futuro cuando tengamos un conocimiento más avanzado?

El requisito de no ser filtrado como irrelevante evita uno de los problemas generados por las clases de referencia, que es el error de las correlaciones espurias. Si la teoría de Mayo pretende ser suficiente para modelar la inferencia causal, tiene que ser capaz de eliminar lo irrelevante, y por lo tanto evitar el error de los factores causales, que es uno de los errores que ella pretende identificar por medio de los cuatro métodos mencionados anteriormente. En realidad los métodos de Mayo no son suficientes para la tarea. Tomemos el caso de Salmon sobre la relación entre ataques al corazón y tomar café. Este es un ejemplo de error al asignar factores causales. De hecho, tomar café es un ejemplo de una correlación causal espuria. Seguramente antes de que los científicos llegaran a la conclusión de que tomar café causa ataques cardíacos, y adoptaran como causal algo que ahora sabemos que no es, ellos hicieron muchos experimentos del tipo de Neyman-Pearson. Esto es, usaron grupo control, descartaron el "efecto placebo" o hicieron pruebas "doble ciegos", plantearon tratamientos aleatorios, balancearon los factores extraños que ellos ya conocían o todavía no conocían, escogieron el nivel de significación y sólo permitieron los resultados corroborados por los experimentos de una alta severidad. ¿Cómo evitar este tipo de error sobre los factores causales reales? Los primeros tres métodos junto con la estadística que Mayo plantea no nos proveen una solución, porque con todos estos maniobras, los científicos todavía pueden llegar a la conclusión que la causa del ataque cardíaco es tomar café, que es una causa falsa.

Pero podría objetarse que el cuarto método de Mayo permite resolver la situación. Creo que no. La severidad de un experimento tampoco es suficiente para descubrir la relevancia de las causas, y sólo sirve, como los otros tres métodos, para decidir cuál es la causa verdadera entre los posibles factores que se consideran relevantes. Es decir, podemos someter a prueba la hipótesis de que el tomar café es la causa del ataque cardíaco, o la hipótesis de que el fumar tabaco es la causa, y averiguar cuál de las dos hipótesis queda corroborada por un experimento de un nivel de severidad más alto. Pero esto se hace sólo cuando los científicos han reconocido como potencialmente relevante

el factor de fumar tabaco. El método de alta severidad mismo no sirve para descubrir nuevos factores posibles. Este es un problema serio que Mayo tiene que resolver pero no lo hace.

En las dos siguientes secciones, estudiaré las teorías de Cartwright y Lipton sobre inferencia causal, y haré una comparación entre estas teorías y el proyecto de Mayo. Tanto Cartwright y como Lipton están interesados en formular una teoría plausible de la explicación científica. En contra de una actitud meramente pragmática sobre la explicación, ambos plantean, aunque de diferentes maneras, la misma idea de que si podemos lograr una caracterización no subjetiva sobre lo que es una inferencia causal, tendremos buenas razones para sostener el valor epistémico de la explicación científica. Por supuesto, ellos tienen diferentes opiniones sobre el proceso de búsqueda de este tipo de inferencia, tema que vamos a ver más adelante. Al comparar las teorías de Cartwright y Lipton con el proyecto de Mayo puede ser muy ilustrativo. Porque, tanto Cartwright como Lipton, como cualquier persona que quiera tener una teoría para modelar la inferencia causal, tienen que afrontar el problema de las clases de referencia. Para ello, ambos autores habían planteado ciertas metodologías que son muy parecidas a las de Mayo, pero rápidamente han conocido sus limitaciones y han propuesto nuevos métodos que, según ellos, son capaces de resolver el problema de las clases de referencia. Estos otros enfoques van a sugerirnos algunos elementos faltantes en el proyecto de Mayo.

3.3 La explicación causal de Cartwright

Cartwright considera correcta la idea de Hempel (también la de Patrick Suppe) que sostiene que el factor explicativo (causa) incrementa la probabilidad del factor a explicarse (efecto). Bajo estas consideraciones, Cartwright formula su propio criterio sobre la relación entre las leyes causales y las leyes de asociación:

CC: C causa E si y sólo si $\Pr(E/C \pm F_1 \dots \pm F_n) > \Pr(E/-C \pm F_1 \dots \pm F_n)$, donde $\{F_1, \dots, F_n, C\}$ es el conjunto de todos los factores causales para E.

Obviamente, el conjunto $\{F_1, \dots, F_n, C\}$ que se trata del problema de las clases de referencia, cumple el requisito de la homogeneidad de Hempel, porque requiere que todos los otros factores F_1, \dots, F_n hayan sido fijados. Pero en el modo de fijar a este conjunto, Cartwright defiere de Hempel. Cartwright sostiene que el criterio CC ni es suficiente ni es necesario para una explicación. No es necesario porque las leyes causales en CC no son transitivas. En este caso, un fenómeno puede explicarse por un factor que se conecte con un serie de procesos, cada uno conlleva su propia ley causal, pero no hay una ley causal que conecta directamente el explanans con el explanandum. No es

suficiente porque un fenómeno singular puede estar dominado bajo varias leyes causales diferentes, y no sabemos, en una situación dada, qué factor causal pueda tener un efecto específico. Es decir, en las diferentes ocasiones, se pueden tener diferentes leyes causales para explicar un fenómeno.

Con relación a esto, Cartwright plantea un proyecto más constructivo, cuya idea consiste en que no se necesita fijar todos los factores relevantes para un fenómeno como piden Hempel y Salmon, sino sólo los factores que tengan influencia en las ocasiones en que nos interesamos. Así, Cartwright introduce un nuevo concepto sobre cómo fijar el conjunto $\{F_1, \dots, F_n, C\}$. No necesitamos saber todas las causas de un fenómeno, sino ciertos factores causales que nos interesan en ciertas situaciones particulares en las que enfocamos nuestra atención. En una situación que nos interesa, tenemos que fijar todos los factores relevantes para que no cometamos el error de “no ser filtrado los factores posibles como irrelevantes”. Eso requiere que hagamos las particiones lo más fina o más homogéneas posibles, pero tampoco es necesario hacer las particiones fuera de esta situación, que es relevante para la explicación, e incluir los factores irrelevantes que no nos interesan. Si fijáramos los factores irrelevantes, podríamos confundir las causas reales con las pseudo-causas. Por ejemplo, la universidad de Berkeley fue acusada por la discriminación sexual en la admisión de los estudiantes, porque el porcentaje de la admisión de los estudiantes masculinos era mucho más alto que la de femeninas. Pero después de la investigación en el nivel de las facultades, descubrieron que el alto porcentaje de los estudiantes masculinos era porque ellos solían tomar examen de admisión en las facultades menos exigentes, pero las estudiantes femeninas solían intentar las facultades más exigentes. Por tanto, cuando hacemos la partición en el nivel de las facultades y no en el nivel de la universidad, podemos rechazar el juicio causal de la discriminación. Obviamente es una teoría más sofisticada que la de Hempel.

Ahora, el problema consiste en cómo decidir si la partición es suficientemente fina o todavía no lo es. Se trata de una pregunta sobre de si es verdadera la causa que hasta ahora conocemos. Tomemos el ejemplo de Salmon otra vez. En un momento t_1 , los científicos hacen su partición y creen que esta partición p_1 incluye a todos los factores relevantes. Con esta partición, ellos llegan a la conclusión de que tomar café causa ataque cardíaco. Pero cierto tiempo después en el momento $t_2 = t_1 + \Delta t$, ellos caen en cuenta de que existen otros factores relevantes que pasaron por alto debido a la partición incorrecta p_1 . Hay un cambio de partición de p_1 a p_2 que involucra un nuevo juicio sobre la causa del ataque cardíaco: ahora es fumar y no tomar café. Pero ¿es correcta la partición p_2 en t_2 ? ¿No existirá el tiempo $t_3 = t_2 + \Delta t$ en que descubramos que la partición p_2 en t_2 no es correcta y el ataque cardíaco no es causado por fumar sino por otros factores que en t_2 no conocemos?

Cartwright conoce muy bien que ésta no es una cuestión trivial. Ella misma pone un ejemplo. La hipótesis clásica de Keynes dice que el dinero causa el ingreso, pero esta conclusión es basada sólo en dos variables, PNB(Producto Nacional Bruto) y la reserva financiera. Pero entre 1972 y 1982, C. Sims hizo una serie de estudios basados en seis variables, a saber, el interés nominal, el gasto federal y el ingreso federal además de los dos mencionados. Descubrió que la tesis de Keynes no era sostenible bajo la nueva partición que incluía a seis variables. Pregunta Cartwright: "is a six-variable model large enough? Or will more variables reverse the findings, and show that money does cause activity in the economy after all?" (Cartwright,1989,p,57) Este problema de las clases de referencia, a saber, problema de cómo saber que la causa que hemos logrado conocer es la causa verdadera constituye una cuestión clave para ella.

La solución, o por lo menos la dirección para intentar resolver este problema, según Cartwright, debe ser la justificación de la metodología para la búsqueda de las causas. Su idea de que una partición debe incluir los factores que son relevantes para la explicación no es suficiente para asegurar que la causa que se encuentre sea verdadera. El remedio está en las manipulaciones experimentales, en ellas se basa la confiabilidad de la inferencia de las causas que hacen los científicos. Dice ella en su libro de 1983:

The fact that the causal hypotheses are part of a generally satisfactory theory is not enough, since success at organizing, predicting, and classifying is never an argument for truth. Here, as I have been stressing, the idea of direct experimental testing is crucial...Engineers at Spectra Physics construct their lasers with the aid of the quantum theory of radiation, non-linear optics, and the like; and they calculate their performance characteristics. But that will not satisfy their customers. To guarantee that they will get the effects they claim, they use up a quarter of a million dollar's worth of lasers every few months in test runs. (Cartwright,1983,p,98)

Aquí, Cartwright expone la idea de que la verdad de una explicación causal se determina por las manipulaciones experimentales. En su libro de 1989, ella analiza explícitamente estas manipulaciones y sostiene que son de dos tipos.

El primero es lo que ella llama experimento aleatorio (*the randomized experiment*, Cartwright,1989,p,62). El ejemplo es el experimento inventado por Fisher sobre el caso de una señora que prueba té.² Una señora dice que al probar una taza de té puede distinguir si se ha vertido primero la leche o el té en la taza. El experimento, que trata de probar si la señora tiene esta capacidad o no consiste en estas maniobras: se prepara cierta cantidad de tazas de té, digamos 100 tazas, unas con leche vertida primero y otras con té primero; luego se pide a la señora que las pruebe

y forme su juicio. Si las respuestas de la señora son correctas hasta cierto nivel, en este caso el nivel de significación, podemos afirmar que lo que dice la señora no está basado en pura adivinación fortuita. Pero los cálculos estadísticos no son suficientes para hacer esta inferencia. Tanto Mayo como Cartwright señalan que existen muchos factores que pueden influir en esta inferencia. Por ejemplo, si los experimentadores preparan 70 tazas vertiendo leche primero y 30 echando té primero, existe la posibilidad de que la señora simplemente diga sin pensar 100 veces que se ha vertido leche primero, y los cálculos estadísticos en este caso reconocen *a fortiori* como verdadera la capacidad de la señora.³ Para evitar esta posibilidad, los experimentadores tienen que preparar 50 tazas vertiendo leche primero y otras 50 echando té primero. Pero si las primeras 50 tazas en el experimento son con leche vertida primero y las siguientes con té echado primero, la señora todavía puede decir simplemente que las primeras 20 tazas en el experimento son con leche vertida primero y las siguientes 80 con té echado primero por una adivinación puramente arbitraria, y la hipótesis de que la señora posee la capacidad considerada pasa el experimento con un alto nivel de severidad estadística. Para evitar ello, los experimentadores tienen que preparar las 100 tazas, 50 con leche vertida primero y 50 con té echado primero, sin ningún orden. Otros elementos señalados por Fisher pueden ser los colores de té, las temperaturas o los materiales de taza. Por ejemplo, la señora puede adivinar correctamente porque ella descubre que el color del té con leche vertida primero es un poco más oscuro que el de té que se ha echado primero. Fisher insiste en que de hecho son inagotables semejantes factores causales incontrolados que puedan influir en el resultado. Cartwright propone el experimento aleatorio del grupo controlado para evitar las influencias de tales factores:

An ideal randomized treatment-and-control group experiment must satisfy two related conditions. It will consist of two groups, the treatment group and the control group; in Fisher's experiment, the cups with milk in first and those with tea in first. The first requirement is that all other causes that bear on the effect in question should have the same probability distribution in both groups...The second requirement is that the assignment of individuals to either the treatment or the control group should be statistically independent of all other causally relevant features that an individual has or will come to have. The random selection of individuals for one group or the other is supposed to be a help to both ends. But it is not the full story. Consider the placebo effect. How is the medicine to be introduced without also introducing some expectation for recovery, or perhaps, in the case of counter-suggestibility, some dread? This itself may be relevant to whether one gets better or not. There are a number of clever and well-known devices, for example blinds and double blinds, to deal with problems like

these, and a vast amount of accompanying literature, both philosophical and practical, discussing their effectiveness. (Cartwright, 1989, p,64)

El segundo tipo de las manipulaciones experimentales que Cartwright propone es experimento totalmente controlado (*The Totally Controlled Experiment*, Cartwright, 1989, p,66). En el experimento de Fisher, se requieren diferentes métodos para lograr cierto control sobre los inagotables factores relevantes y muchas veces no-conocidos. Por el contrario, el experimento totalmente controlado sí puede controlar todos los factores relevantes o eliminarlos. Cartwright toma el experimento de Everitt como ejemplo. Este experimento tiene por objeto probar la teoría de la relatividad de Einstein, que predice que la curvatura de espacio y tiempo producirá dos efectos en un giroscopio, consistentes el uno en la velocidad geodésica que teóricamente debe ser 6.6 arcosegundos por año, y el otro en la velocidad de movimiento que debe ser 0.042 arcosegundos cada año. Pero ¿cómo se garantiza que el resultado del giroscopio es la velocidad geodésica y la velocidad de movimiento y no es producido por los otros factores? Everitt no investiga todos los factores posibles ni calcula sus cantidades para abstraer luego esta cantidad del resultado, como se hace en la mayoría de los procesos experimentales. Lo que él intenta es disminuir, por medio del mejoramiento de los instrumentos, todos los errores posibles hasta el nivel más pequeño posible en el margen del error. En este caso 0.3 milia-arcosegundos por año, y espera que el resultado no sea lo que la pura teoría predecía sino una cantidad presupuesta por este nivel de errores posibles.

Con esta maniobra, los experimentadores pueden controlar los factores extraños sin saber cuáles son estos errores. Por medio del tipo de los experimentos totalmente controlados, junto con el tipo de los experimentos aleatorios, Cartwright nos dibuja un esquema del proceso de fijar los factores relevantes del conjunto $\{F_1, \dots, F_n, C\}$. Dice ella:

I have given a rough sketch of the most dominant kinds of solution: try to construct an experiment that gets the conclusions indirectly without learning the cause one by one, or try to figure out exactly what they are and control for them. In practice both kinds of experiments borrow a little of both methods. Usually social scientists first control for the causes they know about, and then randomize, and in the end the gravity-probe experiment is going to roll the spacecraft in the hope of averaging out any cause they did not know about. (Cartwright, 1989, p,71)

Ahora la pregunta es: ¿son suficientes estos métodos para la explicación causal que Cartwright quiere? Tomamos nuevamente el ejemplo de Salmon como piedra de toque, y nos preguntamos ¿son suficientes estos métodos para que los científicos discernan la falsedad de la relación correlativa entre el tomar café y el ataque cardíaco? Creo que la respuesta es negativa, porque con estos métodos

los científicos podrán discernirlo sólo cuando se haya conocido la posibilidad de que exista otro factor, a saber, el fumar tabaco, que ya pueda eliminar el factor falso, tomar café. Y los métodos por sí mismos no están en condiciones de captar esta existencia. Los científicos sacan la conclusión de que el tomar café causa el ataque cardíaco después de haber usado el experimento aleatorio y el experimento totalmente controlado, por ejemplo, formando dos grupos, el grupo de fumar y el grupo controlado, usando “efecto placebo” o “double-blind”, balanceando la edad, el sexo, el estado de salud, etc., en dos grupos. Además, el mejoramiento de las técnicas experimental (los instrumentos y las experiencias de los experimentadores) pueden disminuir el margen del error dentro un ámbito muy pequeño, a fin de eliminar ciertos errores posibles. Pero todo esto no conduce necesariamente a descubrir que el fumar tabaco también tiene la relación correlativa con el ataque cardíaco antes de que los científicos la sepan. Sólo después de que, quizá por casualidad o por el desarrollo de cierta parte de conocimiento, los científicos se hayan dado cuenta de que el fumar tabaco podría ser una factor relevante, podrán usar estos métodos para decidir cuál es la causa verdadera de ataque cardíaco: el fumar o el tomar café. Esto significa que el proyecto de Cartwright (CC y dos tipos de manipulaciones experimentales), como el de Mayo (cuatro métodos y la estadística del error), aunque puede encontrar la relación estadísticamente correlativa entre una causa y su efecto, no tiene la capacidad de decidir que si esta cause es verdadera o es solamente una pseudo-cause, debido a su incapacidad de afrontar el problema de las clases de referencia.

Cartwright conoce muy bien la limitación del principio CC junto con los dos tipos de manipulaciones experimentales. Ella analiza la causa de esta limitación así: El principio CC requiere que una causa incremente la probabilidad de su efecto en las poblaciones en que todos otros factores causales y relevantes sean fijados. En las prácticas científicas, los científicos no pueden, tampoco es necesario, investigar todos estos factores. Lo que ellos hacen es usar estos dos tipos de las manipulaciones experimentales, para limitar la necesidad de buscar información relevante. Sin embargo, el problema es que la justificación de estas manipulaciones experimentales depende del conjunto completo de todos los otros factores causales. Es decir, podemos usar el principio CC y estas manipulaciones experimentales para hacer una inferencia causal, pero hay un requisito: tenemos que haber sabido que la causa que consideramos como verdadera es la causa genuina, a saber, es la que ha filtrado (*has screened off*) todos los factores irrelevantes.⁴ (Cartwright, 1989, pp.86-87)

Eso significa que una inferencia causal que solamente usa el principio CC junto con estas manipulaciones experimentales no puede ser suficiente. Para hacer suficientemente una inferencia causal, CC junto con las manipulaciones experimentales tienen que relacionarse con los contextos de fondo. Cartwright dice:

...[R]andomized experiments go a long way toward eliminating the need for background causal knowledge. In particular, since the treatment is supposed to be introduced independently of any of the processes that normally occur, problems of spurious correlation can never arise. But the probabilities that show up in a randomized experiment, even in a model experiment where all the ideal specifications are met, will not reveal the true capacities which cause may have. (Cartwright, 1989, p. 103)

Si Comparamos este proyecto de Cartwright con el de Mayo, el resultado de la incapacidad del proyecto de Cartwright para resolver el problema de las clases de referencia, no será extraño. Porque podemos ver que los dos tipos de manipulaciones experimentales de Cartwright, aunque se toman desde otra perspectiva, cumplen la misma función que los tres métodos que Mayo plantea (a saber, el control, el tratamiento aleatorio y el de balance). Los conceptos sobre los controles experimentales y los métodos de “efecto placebo”, “ciego” y “doble ciegos” entendidos por las dos autoras son idénticos. Cartwright clasifica los dos tipos de las manipulaciones experimentales según si los científicos pueden saber o no cuáles son los factores extraños. En los experimentos aleatorios, los científicos podrían llegar a saber cuales son estos factores extraños, y en los experimentos controlados totalmente no necesariamente. La clasificación de Mayo de los cuatro métodos es totalmente compatible con estos dos tipos de manipulaciones experimentales de Cartwright. Generalmente, cuando los científicos saben cuales son los factores extraños, ellos usan la maniobra del control para eliminar estos factores, y cuando los científicos no los saben, ellos usan las maniobras del tratamiento aleatorio o el balance que pueden distribuir los factores posibles igualmente en ambos grupo en prueba y grupo controlado, a fin de neutralizar las influencias de estos factores posibles. Una diferencia grande entre el proyecto de Mayo y el de Cartwright es que Mayo considera que el hacer los experimentos de alta severidad puede servir como un método. Pero como hemos visto en la sección anterior, este método tampoco es capaz de afrontar el problema de las clases de referencia. Todo esto significa que el principio de CC junto con los dos tipos de manipulaciones experimentales de Cartwright tiene la misma incapacidad de afrontar el problema de las clases de referencia que los cuatro métodos junto con la estadística del error de Mayo.

La reflexión de Cartwright sobre la limitación de su principio CC junto con los dos tipos de manipulaciones experimentales demuestra indirectamente, la limitación de los cuatro métodos y la estadística del error que Mayo propone para la inferencia causal. Como hemos visto, los proyectos de las dos autoras tienen muchos elementos similares, y la severidad de experimento, que Mayo propone y que es el único método que Cartwright no toma en cuenta en su manipulaciones experimental, no nos provee manera efectiva para resolver el problema de las clases de referencia. Esto significa que el

proyecto de Mayo tampoco es suficiente para inferencia causal, como lo es el proyecto de Cartwright en el nivel de CC junto con las manipulaciones experimentales. Mayo tiene que mejorar su proyecto incluyendo los contextos de fondo, como Cartwright señala, o de buscar otras maneras que sean suficientes para resolver el problema de las clases de referencia.

Ahora, hay un problema grave para Cartwright. Puesto que la justificación de las manipulaciones experimentales depende del conocimiento de contexto sobre el conjunto de todos los factores causales, nos obliga tener una manera accesible de investigar este conjunto de los factores causales, a fin de poder hacer una inferencia causal. Cartwright dice que los científicos para investigar este conjunto no tienen otra manera más que estos dos tipos de manipulaciones experimentales.⁵ Aquí parece que existe un círculo de justificación. Para romper este círculo, o por lo menos para encontrar una manera que es accesible para poder investigar este conjunto de todos los factores causales, Cartwright propone el principio CC* en el que se distinguen la causa singular y la general. (Cartwright,1989,p,96) Al disponer de suficiente información contextual lograda por la experiencia y la técnica, según Cartwright, los científicos podrán decidir si una causa singular es correcta o no. Las causas singulares tienen la prioridad en la causalidad, porque las generales se pueden modelar en forma de causas singulares. Sin los conocimientos de contexto, no podemos establecer la relación causal por medio de la regularidad que hemos conseguido por medio de las manipulaciones experimentales. Los conocimientos de contexto no sólo incluyen las leyes generales, tales como “aspirina alivia el dolor de cabeza”, sino también los hechos singulares, tales como “una cápsula de aspirina que Pedro tomó ayer alivió su dolor de cabeza”. Según Mayo, este último caso es la base sobre la que los científicos pueden alcanzar para investigar el conjunto de los factores causales que se requiere para una inferencia causal.

Otro punto interesante que vale la pena mencionar es que los experimentos para resolver el problema de la causa singular generalmente no son estadísticos, sino que constituyen casos donde un sólo fenómeno experimental es suficiente. Dice Cartwright:

The bulk of experiments that support the gigantic edifice of twentieth-century physics are never repeated, and they involve no statistics, The trick of the outstanding experimenter is to set the arrangements just right so that the observed outcome means just what it is intended to mean; and that takes repeated efforts, usually over months and sometimes over years. But once the genuine effect is achieved, that is enough. The physicist need not go on running the experiment again and again to lay bare a regularity before our eyes. A single case, if it is the right case, will do.
(Cartwright,1989,p,92)

Esta es una observación que refuerza mi argumento expresado en el capítulo anterior. Además, si esta tesis sobre la prioridad de la causa singular es correcta, la consecuencia consistirá en que los experimentos no-estadísticos no serán menos importante que los experimentos estadísticos en el proceso de la búsqueda de las causas.

La plausibilidad de este nuevo proyecto de Cartwright de causalidad es un tema polémico de la filosofía contemporánea de la ciencia. Aquí no voy a profundizar más, porque está totalmente fuera el alcance de la teoría de Mayo y se pierde el punto comparativo entre las dos teorías. Lo que sí quiero hacer a través de la comparación entre las teorías de Mayo y Cartwright es sostener el siguiente punto de vista: las metodologías estadísticas defendidas por Mayo no son suficientes para modelar la inferencia causal como ella intenta. También la teoría de Cartwright deja en claro por qué la teoría de Mayo no es suficiente: la incapacidad para afrontar los problemas de las clases de referencia.

En la siguiente sección, exploraré de una manera muy breve la teoría de Lipton, que es otra teoría interesante de la explicación causal, y la compararé también con el proyecto de Mayo. Esta comparación nos ilustra, desde otra perspectiva, la misma limitación del proyecto de Mayo para afrontar el problema de las clases de referencia.

3.4 La explicación causal de Lipton

Peter Lipton (1991) ha formulado su proyecto usando la idea de la inferencia a la mejor explicación. Un ejemplo de este tipo de inferencias es la inferencia a las causas, en un modelo de explicación causal. Los métodos para investigar inferencia causal son los que John Stuart Mill plantea y nombra como “métodos de las investigaciones experimentales”. Los métodos millianos son cinco: el método del acuerdo, el método de la diferencia, el método del acuerdo y la diferencia, el método de la variación concomitante y el método de los residuos. Los dos métodos centrales son el método del acuerdo y especialmente el método de la diferencia, porque los otros tres métodos o bien son una combinación de, o casos específicos de, estos dos. El método del acuerdo consiste en sostener los efectos sin cambio y variar las circunstancias de este efecto para ver cuál elemento es el que se mantiene sin cambio durante la variación de las circunstancias. Un elemento que se mantiene sin cambio durante este procedimiento es un candidato a ser la posible causa de este efecto. Por ejemplo, para investigar cuál es la causa del temblor de las manos, los científicos pueden buscar los pacientes que sufren alto nivel de temblor y investigar cuál es el elemento común entre ellos. Si el porcentaje de fumadores entre ellos es mucho más alto que el de las gentes en la sociedad, los científicos pueden considerar que el fumar es una posible causa. El método de la diferencia consiste en sostener las

circunstancias sin cambio y variar los efectos para ver cuál elemento influye el cambio de efecto. Un elemento que influye el cambio de efecto durante este procedimiento es un candidato a ser la posible causa. Los primeros tres métodos de Mayo, a saber, el control, el tratamiento aleatorio y el balance, y las dos manipulaciones experimentales de Cartwright, son precisamente aplicaciones del método de la diferencia. Usando la prueba de Neyman-Pearson, los científicos preparan ambos grupos, el de control y el de prueba, bajo la misma condición. La única diferencia entre dos grupos es que en el grupo de prueba la gente fuma y en el grupo de control la gente no lo hace. Los científicos investigan si esta diferencia va a producir una diferencia significativa del nivel de temblor de las manos.

Aunque ambos métodos, el del acuerdo y el de la diferencia son dos métodos centrales de la metodología de Mill, el método de la diferencia de hecho es más fundamental que el método del acuerdo para la inferencia causal. Porque el método del acuerdo no puede resolver el problema de las causas comunes o la "pluralidad de las causas", en la terminología de Mill. Lipton toma un ejemplo. Dos personas han muerto, una por envenenamiento y la otra por herida de espada, pero ambas fuman. (Lipton,1991,p,110) En este caso, al usar el método del acuerdo, se puede tener una conclusión falsa: el fumar causa las muertes. En nuestro caso sobre la causa del temblor de las manos, los científicos investigan los pacientes que sufren alto nivel de temblor de manos, y descubren que no sólo todos ellos fuman, sino que la mayoría de ellos también tienen edad avanzada. Usando sólo el método del acuerdo, los científicos no pueden distinguir cuál es la causa real: el fumar o tener edad. Mill propone dos maneras de resolver este problema de causas comunes. La primera es aplicar el método de la diferencia cuando es posible. La segunda afirma que lo que se mantiene durante la variación de las circunstancias, o bien es la causa real, o bien es un conjunto de las causas comunes. Lipton señala que algunas veces los conocimientos de fondo nos pueden ayudar resolver el problema de las causas comunes, pero no en todos los casos. (Lipton,1991,p,110) La segunda propuesta de Mill simplemente abandona el intento de hacer la distinción entre la causa real y las causas comunes. La primera acude al método de la diferencia. En las prácticas científicas, los científicos generalmente usan el método del acuerdo para especificar las posibles direcciones de investigación, porque el método del acuerdo puede señalar cuáles son las causas posibles. Pero no pueden utilizar solamente este método para hacer inferencias causales por su incapacidad de resolver el problema de las causas comunes. Por lo tanto, el método de la diferencia es el núcleo de los métodos de Mill para la inferencia causal.

Todo esto es muy interesante para nuestro caso, porque, como hemos visto, los primeros tres métodos de Mayo (también los dos tipos de las manipulaciones experimentales de Cartwright) en la última instancia son las aplicaciones del método de la diferencia que es el núcleo de los métodos

millianos. Por eso, las análisis de Lipton sobre los métodos millianos pueden iluminar de un modo productivo el proyecto de Mayo. Según Lipton, los métodos de Mill también sufren el problema de las clases de referencia, o el problema de la diferencia múltiple, en los términos de Lipton. Explica Lipton esta situación así:

Although Mill's strict statement of the Method of Difference sanctions an inference only when we know that there is a sole difference..., Mill recognizes that this is an idealization. However, ...there will always be more than one difference between their antecedents. Some of these will be causally relevant, but others not. The problem of multiple differences is the problem of making this discrimination. (Lipton, 1991, p. 115)

Obviamente es el mismo problema del ejemplo de Salmon sobre la relación entre el tomar café y el ataque cardíaco, ejemplo en que la causa real, el fumar tabaco, no había sido descubierta. Para hacer una inferencia correcta y suficiente sobre las causas, el método de la diferencia tiene que ser capaz de resolver este problema. Mill propone algunas maneras de hacerlo, pero según Lipton, no son suficientes. Dice Lipton:

Mill proposes two solutions. First, we may ignore differences "such as are already known to be immaterial to the result". Second, while passive observations will seldom satisfy the requirements where a precise change is introduced into a system will often leave that change as the only possibly material difference between the situation before the change and the situation afterwards. Neither of these solutions is adequate. The first allows us to discount some differences, if we know they are irrelevant, but does not tell us how we determine irrelevance. The second shows that Mill is too sanguine about the powers of experimental technique to eliminate all but one possibly relevant difference. (Lipton, 1991, pp. 115-116).

La incapacidad del método de la diferencia para resolver el problema de las clases de referencia no es una conclusión sorprendente para nosotros. Hemos visto en la sección anterior que los primeros tres métodos de Mayo, junto con la estadística del error, que son aplicaciones del método de la diferencia, no son capaces de resolver este problema. También hemos visto que el método de los experimentos con alto nivel de severidad que Mayo propone no nos ayuda superar este problema. Esto significa que los cuatro métodos junto con la estadística del error que Mayo plantea, igual a estos métodos de Mill, no son suficientes para afrontar el problema de las clases de referencia. Por eso, la análisis de Lipton sobre la limitación de los métodos millianos en el problema de las clases de referencia sirve como otro argumento (o el mismo argumento desde otra perspectiva) que señala que el proyecto de Mayo no es suficiente para modelar inferencia causal.

Para superar esta limitación, Lipton acude a su propia teoría de la inferencia a la mejor explicación. En resumidas cuentas, según Lipton, si los métodos millianos quieren ser suficientes para hacer inferencia causal, tienen que añadir las siguientes métodos nuevos de la inferencia a la mejor explicación. Primero, es el método de “hecho y envoltura” (*fact and foil*), que requiere que una explicación causal no necesite contestar la pregunta ¿por qué P?, sino una fórmula comparativa de “hecho y envoltura” que es ¿por qué P mas bien que Q? Por ejemplo, nos preguntamos por qué Pedro contrae parálisis y Roberto no. Al contestar este pregunta, tenemos varios métodos para una inferencia sobre las causas. En nuestro caso, podemos usar el método de la diferencia y decir que Pedro contrae parálisis y Roberto no porque aquél tuvo sífilis y éste no. La sífilis es la diferencia entre Pedro y Roberto. Hempel y van Frassen (Hempel,1965,pp.369-370; van Frassen,1980, p,128) preguntan: ¿por qué José quien también tuvo sífilis, no contrae parálisis? Podemos usar el método de la diferencia otra vez y decir que José tuvo sífilis pero ha sido tratado por un médico a tiempo y Pedro no. Así la causa de la parálisis de Pedro no es la sífilis en la primera inferencia, sino la sífilis sin tratamiento a tiempo. La ventaja de este concepto de “hecho y envoltura” es que nos podemos limitar al ámbito de las diferencias, en lo que sólo estudiamos las diferencias relevantes y excluimos otras irrelevantes. (Lipton,1991,pp.116-117) El segundo método es “hermosura” (*loveliness*) de la explicación. Llamamos una explicación “hermosa” cuando podemos darnos el mejor entendimiento en caso de que esta explicación sea verdadera. Por ejemplo, cuando tenemos varios factores que son causas posibles, podemos preferir lo que pueda relacionarse con efecto por medio de un mecanismo causal, pues esto facilitará nuestro entendimiento. El tercer método es el conocimiento de fondo. Las experiencias y las técnicas muchas veces ayudan a los científicos en determinar las causas verdaderas y en percibir la existencia de las causas espurias (Lipton,1991,pp.118-119).

Por supuestos, estos tres métodos, que sirven como complementos de los métodos de Mill, no se constituyen en un algoritmo para siempre hacer correctamente las inferencias causales, sino que son reglas heurísticas que pueden ayudar a los científicos afrontar el problema de las clases de referencia. También son métodos falibles y dependientes de los contextos de fondo.

En este capítulo, hemos analizado primero la inferencia de los factores causales defendida por Deborah Mayo, quien sostiene que su teoría de la inferencia estadística, junto con los cuatro métodos del diseño experimental y la generación de los datos, son suficientes para modelar objetivamente la inferencia causal sin someter las consideraciones de contexto de fondo. Después del estudio del problema de las clases de referencia, mostramos que el proyecto de Mayo no es suficiente para afrontar este problema, por lo tanto, es decir no puede modelar adecuadamente la inferencia causal. Después, comparamos este proyecto de Mayo con las teorías de la explicación causal de Cartwright y

Lipton, quienes han analizado este problema desde diferentes perspectivas y han tratado de resolverlo diferentes maneras. Estas dos maneras comparten ciertos caracteres comunes. Primero, afirman que la dirección de resolver el problema de las clases de referencia tienen que ver con los contextos de fondo. Segundo, esta dirección no tiene necesariamente que ser estadística. Estos dos caracteres adoptados por dos autores son precisamente los que Mayo deja de lado.

Sin embargo, el problema de la inferencia científica en la teoría de Mayo no radica sólo en la inferencia de los factores causales, sino también en su concepto de la severidad y el nivel de significación, temas que analizaré en el siguiente capítulo.

Notas

1. Algunas veces se clasifica este control según las muestras que se escogen, como en el caso de que Perrin quien preparaba las partículas de polen. Pero está claro que los problemas de escoger las muestras y los controles no son exclusivos. Una consideración en el proceso de escoger las muestras es el problema de cómo encontrar las muestras que tengan la menor cantidad de variables extrañas.
2. Mayo también toma este ejemplo para explicar la jerarquía de los modelos experimentales, porque es un caso clásico experimento del tipo de Neyman-Pearson. Sin embargo, a Cartwright no le preocupan mucho los problemas de la hipótesis nula y contraria, y ni el nivel de significación; más bien lo que le interesa es cómo pueden los experimentadores controlar los factores relevantes, o las variables extrañas en términos de Mayo, para los efectos.
3. En este experimento, la señora sólo tiene que acertar 60 de las 100 pruebas para pasar un experimento estadístico con una severidad mayor que 0.97. Es decir, la posibilidad de que la señora pase el experimento por adivinación puramente fortuita es menos que 0.03. (El cálculo está en la página 159 de Mayo)
4. Aquí sólo expongo esta idea de Cartwright por una manera intuitiva. En el libro de 1989, Cartwright usa el modelo lineal de causalidad para explicar esta idea. Dice ella: "It may seem that the use of these formal techniques avoids the need for conditioning on other causal factors. But it is not so...The reason lies in the open back-path-assumption. In order to draw causal conclusions from an equation, and thereby from the probabilities that identify the parameters of the equation, one must be assured that each factor in the equation has an open back path; and what that guarantees is that each of the factors is a genuine cause after all. So, in fact, the partial conditional probability that tells whether a particular parameter is zero or not---and thus tells whether a putative cause is genuine---is after all a probability that conditions on a complete set of other causal factors."(p,86) (Sobre "back-path-assumption", véase página 31-32 en el mismo libro.)
- 5, Cartwright expresa esta idea también usando el modelo lineal de causalidad. Dice: "In a sense the open-back-path condition is doing the same kind of work that is done by the randomizations and controls of section 2.4 (where introduce the two experimental manipulation): it side-steps the need for direct knowledge of the other causes by taking advantage of other knowledge that is more readily accessible." (Cartwright, 1989, p,86. también vease la nota 4)

Capítulo 4 El cálculo de la severidad y la selección del nivel de significación

En el capítulo 3, propuse la idea de que la inferencia estadística defendida por Mayo no es adecuada para modelar a las inferencias causales, y que la metodología que ofrece Mayo para detectar el error no es suficiente para determinar de manera sistemática todos los tipos de errores que ella misma propone como posible. En este capítulo, estudiaré otro aspecto de la inferencia estadística que Mayo trata de defender: la idea de que la objetividad de las inferencias estadísticas garantiza la objetividad de las inferencias científicas. Mayo cree que el cálculo matemático de la severidad de la experimentación puede garantizar la objetividad de las inferencias estadísticas. Por lo tanto, las inferencias científicas, que según Mayo están basadas totalmente sobre inferencias estadísticas, pueden encontrar la base objetiva que permite hacer inferencias sin ninguna consideración de los contextos de fondo. En este capítulo, primero expondré esta idea de Mayo. Luego, haré ver que el problema del nivel de significación impide una construcción de la objetividad científica como Mayo la plantea. El tema de este capítulo, el cálculo de la severidad y el tema de la objetividad, en realidad, es la otra cara de la moneda del problema que hemos estudiado en el capítulo anterior. Los dos capítulos nos permiten ver, desde diferentes perspectivas, el mismo problema del proyecto de Mayo: la metodología basada sobre la estadística no-bayesiana no nos suministra una teoría suficiente, que ella intenta, para modelar toda inferencia científica.

4.1 Hacia una inferencia objetiva

Generalmente llamamos a la prueba de Neyman-Pearson estadística objetiva para distinguirla de la dirección subjetiva bayesiana. Pero al principio, ni Neyman ni Pearson consideraban que su teoría podría servir de base a una teoría de la inferencia, ni mucho menos a una teoría objetiva de la inferencia.

Como lo hemos analizado en el capítulo 1, la metodología de Neyman-Pearson requiere que los científicos asignen el nivel de significación o tamaño (*size*) α y el poder (*power*) β , antes del comienzo de un experimento, para indicar la probabilidad que los científicos atribuyen a los errores del tipo I y II. Al usar α y β en los cálculos estadísticos y comparar los resultados de estos cálculos con la distribución normal, según Neyman y Pearson, los científicos pueden formar un juicio con buenas razones sobre el resultado del experimento, para rechazar o aceptar la hipótesis nula H_0 .

Pero el propio Neyman no considera que esta metodología pueda suministrar a los científicos una teoría de inferencia científica, porque la aceptación o rechazo de una hipótesis por medio de esta

metodología, a juicio de Neyman, no tiene nada que ver con que esta hipótesis sea verdadera o no, sino que es sólo algo que nos indica que es preferible tomar una acción A a otra acción B. Por ello, cree Neyman que su metodología no es más que un regla de conducta:

Here, for example, would be such a "rule of behaviour": to decide whether a hypothesis, H, of a given type be rejected or not, calculate a specified character, x, of the observed facts; if $x > x_0$ reject H; if $x \leq x_0$ accept H. Such a rule tells us nothing as to whether in a particular case H is true when $x > x_0$ or false when $x \leq x_0$. But it may often be proved that if we behave according to such a rule...we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false. (Neyman and Pearson "On the problem of the most efficient tests of statistical hypotheses" in Joint Statistical Paper, 1967, citado por Mayo, p,368)

Neyman llama a esta regla estadística en los experimentos "regla de la conducta inductiva (inductive behavior)". Inductiva porque aunque α (la probabilidad del error de tipo I) sea muy pequeño y β ($1-\beta$ es la probabilidad del error de tipo II) sea muy grande, las probabilidades de error de ambos tipo I y II no podrán eliminarse completamente. Mayo califica esta posición de conductista, porque, "on Neyman's view, when evidence is inconclusive, all talk of 'inferences' and 'reaching conclusions' should be abandoned. Instead, Neyman sees the task of a theory of statistics as providing rules to guide our behavior so that we will avoid making erroneous decisions too often in the long run of experience." (Mayo,p,368)

La idea de Neyman sobre la implausibilidad de una inferencia en que la verdad esté garantizada por los datos incompletos, y sobre la reducción de esta regla estadística a las decisiones con respecto a las conductas, sirve como una de las razones de los bayesianos para defender la superioridad de su interpretación de la inferencia científica. Este es precisamente un punto con que Mayo no puede estar de acuerdo.

Sin embargo, la metodología de Neyman-Pearson puede entenderse de dos maneras muy diferentes. Una es el conductismo defendido por Neyman. La otra considera que los resultados experimentales son evidencias para las hipótesis bajo prueba. Pearson, el otro autor de la prueba estadística en cuestión, adoptó la segunda posición. Para Pearson, el objetivo de la estadística no consiste en proveernos de una estrategia para escoger una acción, sino que en lo siguiente:

To the best of my ability I was searching for a way of expressing in mathematical terms what appeared to me to be the requirements of the scientist in applying statistical tests to his data. (citado por Mayo, p,381)

La estadística en los experimentos es, para Pearson, una inferencia desde los datos o las muestras a la realidad de las poblaciones en cuestión. Esta suministra a los científicos de una base de razonamiento epistémico para llegar a un juicio, basado en esta inferencia. Pearson lo explica más claramente:

Statistical theory which is not purely descriptive is largely concerned with the development of tools which will assist in the determination from observed events of the probable nature of the underlying cause system that controls them...We may trace the development through a chain of questionings: Is it likely, (a) that this sample has been drawn from a specified population, P; (b) that these two samples have come from a common but unspecified population; (c) that these k samples have come from a common but unspecified population? (citado por Mayo, p,384)

De esta manera, Pearson afirma que la inferencia estadística puede ser una inferencia científica. A fin de lograrlo, los científicos no deben fijarse sólo en la probabilidad, o la comparación entre la probabilidad de una hipótesis H y la de la hipótesis alternativa J, como piden algunos bayesianos y estadísticos, sino que también tienen que estudiar sobre los errores posibles. Para Pearson, el razonamiento experimental no radica en la idea de aceptar una hipótesis H cuando se cree que H está más cerca de la verdad, sino más bien en la idea de que es muy improbable que se cometan errores cuando se acepta una hipótesis H. Dice Pearson:

The tests should only be regarded as tools which must be used with discretion and understanding...We must not discard the original hypothesis until we have examined the alternative suggested, and have satisfied ourselves that it does involve a change in the real underlying factors in which we are interested;...that the alternative hypothesis is not error in observation, error in record, variation due to some outside factor that it was believed had been controlled, or to any one of many causes. (citado por Mayo, p,387)

Claramente esto respalda la idea de que cuando se acepta una hipótesis H, no es suficiente sólo fijarse en su probabilidad: hace falta también verificar si los datos en el experimento pueden representar a la población en cuestión. Pero Pearson no explica de manera clara y concreta, por lo menos de acuerdo con la narrativa de Mayo, en qué consisten los errores, de qué tipos son y con qué métodos pueden los científicos evitarlos. Los errores mencionados en el párrafo citado arriba, a saber, los errores en la observación, en el registro y en el juicio sobre las variables extrañas, son los que se introducen en el diseño experimental, en la generación de los datos o en las condiciones *ceteris paribus*, y no se pueden corregir recurriendo a una maniobra puramente estadística.

Mayo hereda la idea de Pearson en el sentido de que la prueba de Neyman-Pearson sirve como un instrumento para la inferencia científica. Pero su posición es más fuerte, porque ella cree no sólo que los errores experimentales pueden detectarse con los métodos estadísticos, sino también que la inferencia estadística puede constituirse en una teoría objetiva de la inferencia científica.

Ello queda expuesto en su teoría según la cual, por un lado, el cálculo estadístico de la severidad señala la probabilidad cuantitativa de los errores; y por el otro, el criterio mismo de la severidad nos da el conocimiento experimental. Además, hemos mencionado en el capítulo 1 que para Mayo, el conocimiento experimental tiene dos modos, a saber, el modo formal y el modo informal. El modo formal se expresa por la probabilidad o la frecuencia relativa de los resultados experimentales. Por eso, este modo formal puede expresarse como un proceso de satisfacción del criterio de severidad. Siendo cuantitativa y calculada por la estadística, la severidad de un experimento constituye una garantía de la objetividad de la inferencia experimental si el experimento usa la inferencia estadística. Dice Mayo:

Utilizing a test's error probabilities in this manner, customizing even further to take account of the particular result, enables distinguishing warranted from unwarranted interpretations of the results, and it enables doing so objectively. The objectivity of the assessment is afforded by the objectivity of the error probability properties of the test. (Mayo,p,408)

The control of error probabilities has fundamental uses in learning contexts. The link between controlling error probabilities and experimental learning comes by way of the link between error probabilities and severity. The ability to provide methods whose actual error probabilities will be close to those specified by a formal statistical model, I believe, is the key to achieving experimental knowledge. (Mayo,p,411)

Mayo piensa que la severidad de un experimento, que se calcula matemáticamente, nos puede indicar, por medio de la estadística del error, el resultado correcto (*distinguishing warranted from unwarranted interpretations of the results*), y que este procedimiento es objetivo. La objetividad científica se basa en la estadística del error, cuya objetividad está a su vez garantizada por la severidad de la experimentación. Para Mayo, no hay duda sobre la objetividad de la severidad de la experimentación, pues resulta de cálculos matemáticos estrechamente vinculados a las investigaciones científicas.

Yo pienso sin embargo que la objetividad en la experimentación es algo más problemático. Para entender esto, veamos más claramente el procedimiento de cálculo de la severidad de un experimento.

En el primer capítulo describí el procedimiento para el cálculo de la severidad por medio del caso concreto de la relación entre el fumar y el temblor de las manos. Aquí me limito a exponer la razón que Mayo da sobre la validez de este cálculo. Mayo acepta el concepto de severidad presentado por Popper, concepto según el cual los científicos tienen que plantear, con relación a la falsedad de las hipótesis, experimentos de alta severidad:

The theoretician will therefore try his best to detect any false theory...he will try to "catch" it. That is, he will...try to think of cases or situations in which it is likely to fail, if it is false. Thus he will try to construct severe tests, and crucial test situations.
(Mayo,p,14)

Aunque esta idea sobre la severidad experimental de Popper es muy parecida a la de Mayo, existen tres diferencias importantes entre la una y la otra. Primero, el concepto de severidad según Popper depende del riesgo que tenga una hipótesis de fallar. Este riesgo será mayor cuando más falsable sea la hipótesis (es decir cuando mayor sea su contenido empírico, cuantas más consecuencias observacionales posea). Segundo, la severidad también depende de la relación de la hipótesis con otras hipótesis alternativas. H pasa una prueba de alta severidad con e (un enunciado de evidencia) si todas las hipótesis alternativas predicen $\neg e$ y el resultado de la prueba es e . Tercero, esta noción de severidad no puede calcularse por medio de la teoría de probabilidades. El concepto de Mayo sobre la severidad es muy diferente.

Sería más fácil entender la idea de Mayo, si entendiéramos primero los dos casos límites de la severidad:

La severidad mínima (cero): H pasa una prueba T de severidad mínima con la evidencia e si y sólo si, cuando H es falsa, T también la deja pasar.

La severidad máxima (uno): H pasa una prueba T de la severidad máxima con la evidencia e si y sólo si, cuando H es falsa, T nunca la deja pasar.

Una vez que tengamos estos casos límites, podemos tener una definición de severidad, que Mayo llama criterio de severidad:

La probabilidad de que H con la evidencia e pase la prueba T, es muy baja, si H es falsa.

En el capítulo 1, hemos expuesto que la severidad es, según la prueba de Neyman-Pearson, una función del nivel de significación α y el poder β , y que α y β son las probabilidades asignadas para dos tipos de errores: el error de tipo I y el de tipo II, consistentes respectivamente en rechazar H cuando H es verdadera y aceptar H cuando H es falsa. Se escriben así:

La probabilidad de error de tipo I = $\Pr(T \text{ rechaza } H/H \text{ es verdadera}) \leq \alpha$.

La probabilidad de error de tipo II = $\Pr(T \text{ acepta } H/J \text{ es verdadera}) \leq \beta$.

En estas fórmulas, J es la hipótesis contraria. Como el error de tipo I y el de tipo II comparten la misma distribución de frecuencia, β también es una función de α . Por eso, a continuación sólo voy a analizar α .

Una vez que tengamos el nivel de significación asignado, podemos calcular la severidad cuantitativamente. Por ejemplo, en el caso de que H sea rechazada y J aceptada, la severidad será la probabilidad mínima de no cometer el error de tipo I, o sea, será $1-\alpha$. Como queda señalado en el capítulo 1, eso significa que la media observada \bar{X} es mayor que el punto límite del rechazo de H y la aceptación de J, el punto \bar{X}^* . La severidad representa la situación de $\bar{X} = \bar{X}^*$, y garantiza que J pasa la prueba T con un grado de severidad de $1-\alpha$, o dicho de otra manera, con la probabilidad $1-\alpha$ de no cometer el error de tipo I. En el caso del fumar y el temblor de las manos, α es 0.03, y la severidad entonces es 0.97. Se puede expresar el razonamiento más formalmente así:

$$(1) \Pr(\text{la prueba T tiene una media mayor que } \bar{X}^* / H \text{ es verdadera}) \leq \alpha$$

Ello implica que al pasar J una prueba T, J soporta una severidad $1-\alpha$, por las siguientes razones:

$$(2) \Pr(\text{pasa J en T/H es verdadera}) \leq \alpha$$

$$(3) \Pr(\text{no pasa J en T/H es verdadera}) > 1-\alpha$$

Entonces:

$$(4) \text{ Al pasar la prueba T, J ha pasado una prueba de alta severidad } 1-\alpha$$

A juicio de Mayo, la objetividad tanto en la inferencia estadística como en la experimental, sirve para eliminar todos los elementos pragmáticos y contextual de fondo en la justificación de estas inferencias. Esta idea resulta ser muy entendible si tomamos en cuenta las tres propósitos de todo el proyecto de Mayo, propósitos que ya analicé en el primer capítulo. Uno de ellos consiste en encontrar, como lo intentaron antes Hempel, Carnap y Salmon entre otros, una inferencia inductiva totalmente objetiva, o caracterizable con independencia de la estructura de los contextos de fondo. A diferencia de ellos, Mayo utiliza la teoría estadística de Neyman-Pearson para representar el razonamiento de la inferencia científica, que según Mayo se puede reducir a la inferencia experimental. Puesto que el cálculo de la estadística es un proceso puramente matemático, parece que Mayo ha conseguido realizar con éxito su proyecto. Para Mayo, ésta es una manera muy plausible y quizá la única para evitar la dirección relativista representada por el Programa Fuerte en la filosofía de la ciencia contemporánea.

Nadie va a dudar de que el procedimiento del cálculo matemático mismo es objetivo. Pero, yo pienso que al introducirlo, los científicos no están interesados en el cálculo mismo, sino en hacer uso de él como instrumento que ayuda a formar juicios. La decisión de cómo y cuándo usarlo implica una subjetividad ineludible en una inferencia experimental. Mayo insiste en que la objetividad de una

inferencia experimental con el uso de la estadística del tipo de Neyman-Pearson, está garantizada por el cálculo de la severidad, y el cálculo mismo es objetivo. Pero el problema no está en el cálculo mismo.¹

Al leer con cuidado el proceso de este cálculo en un experimento, podemos darnos cuenta de que la severidad es una función del nivel de significación α . Antes de un cálculo sobre la severidad, los científicos tienen que escoger y postular α , generalmente 0.05, 0.01 y 0.001. Mayo cita directamente en su obra los niveles de significación de experimentos específicos sin ninguna explicación sobre por qué los científicos los escogen, y con qué procedimiento lo hacen. Parece que, para ella, el nivel de significación α es un elemento ya dado y sin problema, y por eso no tendrá ninguna influencia sobre la objetividad de la inferencia. Esta actitud sobre el nivel de significación de hecho es muy común dentro de los teóricos estadísticos, quienes creen que un nivel de significación no es problemático porque se da en un experimento específico a través de una convención no problemática. En general, los teóricos estadísticos toman el nivel 0.05, 0.01 y 0.001 en diferentes grados de juicio como una convención. Pero hay una diferencia entre Mayo y estos teóricos: ellos no pretenden fundamentar la objetividad de la inferencia científica por medio de estos procedimientos.

Aquí es donde aparece un problema para Mayo. Según ella, la objetividad de la inferencia científica está garantizada por la objetividad de la inferencia estadística, y ésta está garantizada por la severidad de la experimentación. Pero, la severidad no es está exenta de problemas como ella piensa, pues al ser función del nivel de significación, los científicos tienen siempre que asignar un nivel de significación antes de poder calcular la severidad. Esta asignación de un nivel de significación no es nada objetiva, y tiene problemas. En las siguientes dos secciones, haré ver que, primero, el nivel de significación está relacionado con el tamaño de las muestras, y segundo, la asignación de un nivel de significación es un procedimiento pragmático que tiene ineludiblemente que ver con los contextos de fondo.

4.2 El nivel de significación y el tamaño de las muestras

El problema de nivel de significación es un punto clave para el proyecto de Mayo, y no puede pasárselo por alto como ella lo hizo. Primero, como hemos visto, el nivel de significación es una función de la severidad de experimento, la cual es la base de la objetividad de la experimentación. Segundo, el nivel de significación no es nada fijo y absoluto, sino por el contrario, está muy relacionado con, e influido por, los contextos de los experimentos específicos. En esta sección, analizaré la influencia del tamaño de las muestras sobre la selección de un nivel de significación

Hasta hoy en día, muchos estadísticos y filósofos han señalado que el tamaño de las muestras tiene influencias sorprendentes sobre el nivel de significación. Ronald Giere apunta que mientras mayor sea la cantidad de las muestras, menos diferencias habrá entre la frecuencia observada y la esperada, de manera que el juicio sobre rechazar o aceptar la hipótesis nula es cambiante. Veamos un ejemplo. La Universidad de Berkeley fue acusada de discriminación sexual en la admisión de los estudiantes y quería hacer una prueba de Neyman-Pearson para investigar este asunto. La hipótesis nula H_0 suponía que no existía tal discriminación de la que se la acusaba. Pero la Universidad no tenía suficiente dinero para realizar una prueba de gran magnitud, y se contentó con una cobertura de 100 muestras. En este caso, H_0 , que dice que no hay discriminación, requiere que la frecuencia observada \bar{X} sea igual que la frecuencia esperada μ_0 , que es 50%. Los investigadores, según la convención, escogieron el nivel de significación $\alpha=0.05$. Siguiendo la distribución normal, si H_0 es correcta, la frecuencia observada estará dentro de la variación entre 31% a 69%. Es decir, si el porcentaje de las mujeres estudiantes en la totalidad de los 100 individuos en muestra es mayor que 31%, por ejemplo, 42%(42 mujeres en 100), la acusación de discriminación no se fundamenta. Pero si el porcentaje es 29%, la acusación es correcta y H_0 debe ser rechazada. Ello sucede cuando el nivel de significación es 0.05 y el tamaño de la muestra es 100. Sin embargo, si la universidad hubiera tenido más dinero y hecho una prueba con más muestras, bajo el mismo nivel de significación 0.05, el resultado podría haber sido muy diferente. Porque bajo las mismas condiciones de nivel de significación 0.05, la frecuencia del intervalo de la distribución normal cambia según las diferentes magnitudes de las muestras. Este cambio se ve en la siguiente tabla:

el tamaño de muestra	la frecuencia del intervalo
10	0 a 100%
25	14% a 86%
50	26% a 74%
100	31% a 69%
250	40% a 60%
500	42% a 58%
1000	44% a 56%
1500	46% a 54%

Esta tabla señala que si la universidad sólo hubiera tomado 50 casos como muestra, en el nivel de significación 0.05, el porcentaje 28% de las estudiantes mujeres (14 en 50) no habría sido

suficientemente significativo para rechazar H_0 , pero en caso de una muestra de 100 individuos. Al contrario, si la universidad hubiera tenido más dinero para efectuar una prueba de 250 individuos, el porcentaje 42% (105 en 250) habría sido suficiente para sostener la acusación, pero este porcentaje, como hemos analizado, no es significativo en el caso de tal muestra de 100.

Giere nos advierte que la validez de una inferencia estadística está íntimamente relacionada con la forma en que los investigadores especifican el tamaño de las muestras:

What this tells you is a report of "statistical significance" without mention of sample size is MEANINGLESS. Any difference between the observed frequency and the mean value predicted by the test hypothesis might or might not be statistically significant. It all depends on the size of the sample. Thus, if you read a report claiming either "statistical significance" or "no statistical significance," but you cannot tell whether the sample included 25, 250 or 2,500, forget it. The report is worthless. (Giere, 1984, p. 249)

Hanan Selvin señala que la importancia de la relación entre el nivel de significación y el tamaño de las muestras ha sido olvidado frecuentemente por los investigadores que se valen de la prueba de significación. De hecho, se trata de un problema clave para una inferencia estadística. Dice él:

A related difficulty is the confusion of statistical significance with substantive importance. Some researchers give more prominence to the level of significance than to the size of the difference; often only the level of significance and the direction of the finding are reported, and occasionally the level of significance appears alone. High levels of predictability, explanation, and association are legitimate goals for social scientists; they are not the same as a high level of significance, nor is statistical significance a substitute for them. A 1 percent difference may be significant at the .001 level if the sample is large enough, yet such a small difference is essentially meaningless for sociology at present. Correspondingly, a large difference that is not significant at the 5 percent level, simply because it was based on a small sample, may be of major theoretical importance. (Selvin, 1970, pp. 100-101)

El cálculo de la severidad experimental es una función del nivel de significación, y antes de acudir a este cálculo, los científicos tienen que tener el valor de α . Pero el nivel de significación cambia con el tamaño de las muestras. Ahora, el problema está en que la decisión sobre el tamaño de las muestras no es objetiva. Por un lado, bajo un nivel de significación fijo, cualesquiera hipótesis pueden ser rechazadas si el tamaño de las muestras es suficientemente grande. Los científicos tienen

que seleccionar un punto para detener una aumento potencialmente infinito del tamaño de las muestras. La cuestión de dónde está este punto, de no ser totalmente arbitraria, debe depender de la experiencia y el buen juicio (*good sense*) de los científicos que efectúan un experimento específico. Por otro lado, los elementos pragmáticos, por ejemplo, el tiempo y el costo, influyen en la decisión del tamaño de las muestras. Si el nivel de significación no es objetivo y es imprescindible para el cálculo de la severidad experimental, no veo por qué la severidad experimental, que es relativa a los elementos no objetivos, pueda ser la garantía de la objetividad de la inferencia experimental como Mayo plantea.

Howson y Urbach también analizan el mismo problema pero desde otro punto de vista. Dicen:

In any kind of significance test, a hypothesis is assessed according to the probability of the observed outcome relative to the probabilities of other possible outcomes. In other words, whether an observation favours a hypothesis or not depends upon what observations might have been made if the trial had not turned out as it did. This means that the significance of a result of, say, 1 head and 3 tails in an ordinary coin-tossing trial would depend on how the experimental apparatus dealt with other possible outcomes. (Howson y Urbach, 1996, segunda edi. p, 210)

Por ejemplo, tiramos una moneda para probar si ésta es una moneda justa. Imaginemos que ya hemos tenido 6 águilas e i (i es un número natural entre 0 e infinito)soles. Pero, si tenemos un nivel de significación fijo, la aceptación y el rechazo de la hipótesis de que la moneda es justa depende totalmente del número i . Supongamos que tenemos una i bajo la cual debemos aceptar esta hipótesis; pero siempre existe la probabilidad de que estamos en una situación en la que al tirar otra vez la moneda caiga un sol, y eso nos obligue, según la estadística, a rechazar la hipótesis (Howson y Urbach, 1996, p,210).

No es difícil percibir que se trata del mismo problema del tamaño de las muestras, pues cuando cambia la cantidad de individuos, cambia la inferencia sobre el resultado experimental. Lo único que se puede hacer consiste en detener este cambio de i en cierto punto determinado. En la práctica científica, los científicos tienen diferentes maneras de hacerlo. Howson y Urbach llaman estas maneras Reglas de Detención (*stopping rule*), y una de ellas es la prueba de Neyman-Pearson. Pero existen también muchas otras reglas cuyo uso depende de elementos totalmente subjetivos según Howson y Urbach:

...[T]here is no limit to the number of possible stopping rules...and not all of them make the decision to stop the trial depend on the outcome, which some statisticians think is not quite legitimate. For instance, suppose that after every toss of the coin, the

experimenter had pulled a playing card at random from an ordinary pack with intention of calling off the coin trial as soon as the Queen of Spades appeared. This stopping rule induces a quite different outcome space from the two we have considered, which is bound to lead to different conclusions in certain cases.

...The particular experimental result could even be the product of many, conflicting stopping rules. For a scientist might have acted on one such rule, while his collaborator had privately decided to conduct the trial in accordance with another, and yet the outcome of the trial might have happened to satisfy both rules. The outcome space in such a case might be hard to determine, for you would need to discover what each experimenter would have done if the actual outcome had not accorded with his own stopping rule. Would he have tried to overrule his colleague or given in quietly? And if the former, would he have succeeded? Intuitively, such questions do not bear on the testing process, and in practice they are rarely if ever dealt with, yet for the classical statistician, they are vital. (Howson y Urbach, 1996, segunda edi. pp,211-212)

Al criticar la dirección de la estadística clásica, a saber, la de Fisher y la de Neyman-Pearson, Howson y Urbach persiguen el propósito de demostrar que la estadística clásica no es más objetiva que la dirección bayesiana y subjetivista. Quizás algunos ejemplos citados por ellos sean demasiado exagerados, como el caso de la Regla de Detención mediante la aparición de la Reina de Espadas, pues son casos realmente muy raros en la práctica científica. No quiero intervenir en la polémica entre bayesianos y no-bayesianos, ni tampoco discutir la legitimidad de estos casos raros. Pero creo que es correcta la razón que dan Howson y Urbach en el sentido de que no hay una Regla de Detención universalmente aceptada por todos los científicos, sino que cada científico podría tener su propia regla.

Todos estos autores señalan que un nivel de significación no puede servir para una inferencia estadística sin la especificación del tamaño de las muestras. Esta especificación depende a su vez de los contextos de cada experimento específico. Por lo tanto, el cálculo de la severidad de experimentación también tiene que ver ineludiblemente con estos contextos.

4.3 Selección del nivel de significación

La severidad de la experimentación no sólo se relaciona con contextos de fondo por medio de la especificación del tamaño de muestra, sino que también a través de consideraciones pragmáticas en la selección del nivel de significación.

Hemos visto que la mayoría de los teóricos estadísticos consideran que el nivel de significación se da por convención y no es nada problemático. Pero esta convención, ha sido criticada por algunos filósofos. James Skipper, Anthony Guenther y Gilbert Nass(1970,pp.155-160) destacan muchos casos en que los sociólogos, en sus investigaciones estadísticas, aceptan el nivel 0.05 como una convención usada acríticamente. Sin embargo, esta convención nació por una casualidad. Fue Ronald Fisher, afirman ellos, el primero en proponer este nivel de significación para los ejemplos y los ejercicios de su libro "The Design of Experiments"(1935). Posteriormente los investigadores, que adoptaron la prueba de Fisher y los filósofos fisherianos, la asumieron como natural sin ninguna preocupación por su validez (p,156, nota. 2). Pero el hecho es que la asignación de 0.05, 0.01 o 0.001 como nivel de significación es completamente arbitraria. Como dice B.J. Winer:

The frequency of the .05 and .01 levels of significance is a matter of convention having little scientific or logical basis. When the power of tests is likely to be low under these levels of significance, and when Type I and Type II errors are of approximately equal importance, the .30 and .20 levels of significance may be more appropriate than the .05 and .01 levels. (Winer, 1962, p. 13)

Por eso, Skipper, Guenther y Nass llaman a esta convención arbitraria "el sacramento del 0.05". También Howson y Urbach sostienen que es precisamente debido a este aspecto subjetivo de la prueba de Fisher, la inferencia bayesiana no tiene la menor validez:

...[I]n order to derive a unique conclusion from a Fisherian test of significance, arbitrary decisions need to be taken, for instance, as to the appropriate test-statistic. These decisions are arbitrary since the various rules recommended for fixing the test-statistic have been given no rational defense, seeming to rest on personal judgements alone. As a leading advocate of Fisherian methods admitted, "there is no answer to [the question 'which significance test should one use?']...except a subjective one", adding in parenthesis that it was "curious that personal views intrude always". Indeed, it is curious, when one considers that Fisherian methods arose from a dissatisfaction with the Bayesian approach on account of its supposed subjectivity (Howson y Urbach, segunda edi. pp,192-193)

La especificación del nivel de significación, según Mayo, no se limita, sin embargo, a esta convención. Para evitar el elemento no objetivo durante la decisión sobre el nivel de significación, Neyman y Pearson prescriben el siguiente criterio: escoger un nivel de significación donde el tamaño (*size*) de la hipótesis nula es más pequeño y el poder de la hipótesis contraria es más grande. Es decir, escoger un nivel de significación en la que la probabilidad del error tanto de tipo I como de tipo II sea

la más pequeña. Parece que este criterio es suficiente para garantizar la objetividad de la selección del nivel de significación, porque, al evitar la acusación de Howson y Urbach contra Fisher, este criterio no sólo puede ofrecer una defensa racional de la especificación de un nivel fijo. Se trataría así de una decisión en la que la probabilidad de los dos tipos de errores sería la menor, y que no depende de ninguna consideración subjetiva.

Pero el problema es que no siempre tendremos casos en los que las probabilidades de los dos tipos de error sean, simultáneamente, suficientemente pequeñas. Porque muchas veces, al tomar un nivel de significación o tamaño α del error de tipo I suficientemente pequeño, carecemos de un poder β suficientemente grande. Es decir, los dos tipos de errores están vinculados. Cuando disminuye la probabilidad de uno puede aumentar la de otro. Kendall explica esta situación así:

These tests (significance tests) are designed to keep one from making statements about percentage differences (or other difference) when there is too little evidence to justify the statement. In order to do this, in order to avoid making unjustified claims about the magnitude and importance of an observed difference, one sets a rather high level of significance; one says, in other words, that, if there is more likelihood than 1 in 100 (or 5 in 100) that the observed difference could have come about by chance, one must conclude that it is not a "real" difference. By insisting on a high level of significance (a low probability that there is no real difference) one increases the possibility of rejecting as insignificant differences which actually do exist. And this is a serious error in exploratory studies of the kind which we have been conducting. (Kendall, 1970, p.88)

Este problema se resuelve, según Kendall, relacionando el juicio estadístico con el contexto de fondo. Es decir, con las informaciones de que disponen los investigadores antes de la ejecución de un experimento, y con las informaciones nuevas que aparecen durante el proceso del experimento. Por ejemplo, si queremos saber si el medicamento A es más efectivo que el B para una enfermedad, y acudimos para ello a la prueba estadística, el proceso de hacer una inferencia experimental debe ser así:

We want to gather together as much information as we can about the relations between experiences in medical school and the development of particular attitudes and values. Later on, some of the hypotheses which we set up before starting the research, and others which have emerged during the course of our investigation, can be submitted to more definitive and rigorous tests. Until that time, however, we do not want to cut short possibly productive lines of investigation by insisting now that our preliminary results prove themselves significant. We do not want to say now that, because some of our

early results do not meet stringent criteria of significance, they should be disregarded, or discounted, or not reported. (Kendall, 1970, p,88)

Al fin de cuentas, un nivel de significación en sí mismo no es suficiente para definir independientemente una inferencia experimental como Mayo plantea, sino que ha de relacionarse con todo el contexto de fondo de un experimento específico.

Skipper, Guenther y Nass expresan esta misma idea en muy pocas palabras:

Since the two types of errors are inversely related to each other, it is impossible to minimize both of them at the same time without increasing the sample size. Therefore, it is the nature of the problem under study which ought to dictate which type of error is to be minimized. (Skipper, Guenther y Nass, 1970, p,157)

Además, estos filósofos no sólo están de acuerdo con la conclusión de Kendall sino que sostienen también que la magnitud de los datos está relacionada con la asignación del nivel de significación. Pero, por las razones que he expuesto en la sección anterior, la magnitud de las muestras también es un elemento relacionado con el contexto de fondo. Por eso no veo ninguna manera de sostener la existencia de una objetividad total en el procedimiento de la asignación de un nivel de significación.

Al criticar el "sacramento de .05", Skipper, Guenther y Nass proponen un procedimiento pragmático de asignación del nivel de significación en los estudios sociológicos. Primero, al seleccionar un α entre 0.05, 0.01 y 0.001, los investigadores en las ciencias sociales tienen que cuidarse mucho del carácter arbitrario de la convención que están usando. La adopción de la convención no es una panacea universal de toda interpretación de la evidencia. Diferentes tipos de investigaciones requieren diferentes α , según las diferentes interpretaciones en los diferentes experimentos específicos. Segundo, al escoger un α , los investigadores tienen que incluir las consideraciones éticas y morales, especialmente en caso de que, al pasar y no-pasar la prueba, una hipótesis tenga diferentes consecuencias sociales. Es decir, cuando al pasar la prueba una hipótesis tiene una consecuencia social más grave, se requiere que se tome un α más pequeño; en cambio, al no pasar la prueba, la consecuencia social es más grave, se requiere que tome un β más pequeño, o un poder más grande. Tercero, al dar a conocer el resultado de una investigación sociológica, hay que explicar a los lectores el carácter arbitrario del nivel de significación que se ha usado, para que los lectores se formen su propio juicio. Ello implica que los investigadores no deben escoger un nivel de significación estándar, sino que deben dar a conocer a los lectores una explicación de cualquier decisión. Skipper, Guenther y Nass creen que este proceso intersubjetivo constituye un método, quizá

el único, para lograr una inferencia experimental más racional y objetiva. (Skipper, Guenther y Nass, 1970, pp, 159-160)

Labovitz (1970,pp,166-171) está de acuerdo con la crítica al "sacramento del nivel de significación", pero insiste en que, a pesar de poder evitar ciertos problemas, las tres estrategias de Skipper, Guenther y Nass son suficientes para los sociólogos al utilizar la metodología estadística en sus investigaciones, y especialmente no ponen el problema en el contexto de la relación entre conocimiento, teoría y evidencia.² Por eso, Labovitz plantea once criterios para que los investigadores seleccionen el nivel de significación. Dice Labovitz que, si bien estos once criterios no son completos, ofrecen por lo menos las dimensiones más importantes que un investigador tiene que enfrentar cuando toma una decisión sobre el nivel de significación. (Labovitz,1979,pp,166-167) Yo creo que estos criterios también son importantes para entender cómo el problema del nivel de significación es contextualmente dependiente. Por eso, describiré brevemente estos once criterios:

- 1) Las consecuencias prácticas Al escoger un nivel de significación los investigadores tienen que pensar en las consecuencias prácticas, porque ciertos tipos de errores son más graves. Por ejemplo, la utilización de la prueba de Neyman-Pearson para examinar cuál es más efectiva entre la lobotomía prefrontal o la sedación, será un experimento con consecuencias prácticas muy graves, porque un error muy pequeño puede matar a un paciente. En este caso, se requiere un nivel de 0.001 o más pequeño aún. Por otro lado, si el experimento está destinado a probar entre los dos tipos de sedación, el nivel de significación puede ser más relajado, por ejemplo 0.05, porque la consecuencia de un fallo es menos grave en comparación. De hecho, este criterio es muy parecido al requisito de las consideraciones morales y éticas que Skipper, Guenther y Nass proponen.
- 2) La plausibilidad de las hipótesis alternativas Generalmente las hipótesis bajo prueba no están aisladas, sino que están conectadas con los conocimientos de fondo que los investigadores poseen. En caso de que el resultado esté en contra de las teorías existentes o el sentido común, en otras palabras, sin ningún apoyo teórico, se requiere que los investigadores seleccionen un nivel muy pequeño (0.01, 0.001 o menos) sobre la hipótesis que sostenga este resultado, porque los científicos tienen que poner mucho cuidado al refutar las hipótesis que han sido establecidas y que conectan con apoyo teórico. En caso contrario, cuando el resultado esté en conformidad con las hipótesis que cuentan con apoyo teórico, se puede escoger un nivel de significación normal y suficiente (0.05, por ejemplo).
- 3) La relación entre el poder β y el tamaño α Tanto el poder β como el tamaño α se relacionan con la magnitud de las muestras. Cuando la cantidad de muestras se incrementa, la probabilidad de rechazar correctamente la hipótesis nula aumenta. Mientras tanto, con el incremento de la cantidad de muestras, tenemos menos desviación estándar, es decir, el tamaño α disminuye. Por eso, un pequeño

nivel de significación (0.01, 0.001) y un gran poder β corresponden generalmente a una gran cantidad de muestras, y un nivel de significación más grande (0.1, 0.05) corresponde a una menor cantidad de muestras.

4) La relación entre el poder β y la diferencia D entre las medias Hemos definido en el capítulo 1 la diferencia D como diferencia entre la media μ_0 de la población y la media de las muestras \bar{X} . La variación del poder β no sólo está relacionada con el cambio del tamaño α , sino también con la diferencia D . Si la magnitud de las muestras no es demasiado pequeña, con una diferencia D más grande tenemos más probabilidades de rechazar correctamente la hipótesis nula. Los criterios 3 y 4 exigen que los investigadores escojan un nivel de significación según sus relaciones con el poder β y el tamaño de muestras en los experimentos específicos.

5) La relación entre el error de tipo I y el de tipo II Hemos mencionado que Kendall, Skipper, Guenther y Nass han criticado la plausibilidad de buscar simultáneamente un tamaño α pequeño y un poder β grande, porque generalmente la relación entre α y β está inversamente relacionada con la magnitud de las muestras. Esto significa que cuando disminuye la probabilidad del error de tipo I, tiene que aumentar la probabilidad del error de tipo II. Labovitz insiste en que la selección del nivel de significación tiene que ver por ello con qué tipo de error vale más la pena evitar según las decisiones de los científicos.

6) La convención Skipper, Guenther y Nass se oponen con vigor a la convención ("sacramento") de un nivel de significación fijado *a priori* en un experimento. Labovitz señala que si un nivel de significación ha sido usado en una misma esfera de estudios repetida y exitosamente, puede ser una convención local para otros experimentos. Pero, esta convención sirve tan sólo para facilitar las investigaciones a cierto nivel sobre ciertos temas, y carece de toda base epistémica y lógica.

7) El grado de control en el diseño experimental Al usar en su libro el nivel de significación 0.05 en lugar de 0.01 para el experimento agrícola que trata como un ejemplo, Fisher planteó además muchos controles experimentales para eliminar los posibles factores extraños. En ese caso, Fisher estaba muy confiado en que muy pocos factores extraños al protocolo de su experimento podrían afectar al resultado. Por eso, un nivel normal 0.05 es suficiente. En general, un experimento cuyo diseño experimental tiene poco control sobre los factores extraños requiere un nivel de significación más estricto, como 0.01 o 0.001 e incluso menos. El problema está en cuántos factores extraños existen y si son suficientes o no los controles. Esto depende de los conocimientos de fondo que tengan los científicos.

8) La rigurosidad de la prueba La rigurosidad es la capacidad de un prueba estadística para mantener su conclusión deducida en caso de que los presupuestos experimentales hayan sido violados.

Obviamente este punto está relacionado con el punto 7, porque cuando una prueba es bastante rigurosa, tendrá menos factores extraños que no estén bajo control. Por eso, una prueba de poca rigurosidad requiere un nivel de significación bajo (0.01, o 0.001 o menos).

9) La relación entre la prueba de “una dirección” y la de “dos direcciones” En la primera a los científicos sólo les preocupa la diferencia positiva D o la diferencia negativa $-D$, pero no ambas. En la segunda les preocupan ambas diferencias, positiva y negativa. No voy a estudiar en mi trabajo la relación entre los dos tipos de prueba, porque sus cálculos estadísticos son los mismos. Sin embargo, Labovitz señala que muchos textos sobre la estadística insisten en que en la prueba de “una dirección” la hipótesis nula es más fácil de ser rechazada que en la de “dos direcciones”. La razón de esta relación radica en que por regla general los científicos suelen tener más confianza en una dirección que en dos direcciones.

10) La confianza de intervalo La selección del nivel de significación también equivale a la selección del intervalo de confianza. Para un intervalo de confianza más pequeño, se requiere un nivel de significación más grande.

11) La diferencia entre el propósito de probar una hipótesis y el de desarrollar las hipótesis Cuando el propósito del experimento consiste en probar una hipótesis, se requiere que el nivel de significación sea más pequeño para que la hipótesis no sea rechazada erróneamente. Pero cuando el propósito es desarrollar hipótesis que todavía no han sido establecidas, un nivel menos estricto (0.1 o 0.2) es suficiente. (Labovitz, 1970, pp,167-171)

Los once criterios arriba mencionados exigen que la selección de un nivel de significación incluya diferentes consideraciones sobre los diferentes contextos, en experimentos muy específicos. Puesto que el cálculo de la severidad que propone Mayo es una función del nivel de significación, la inferencia experimental, cuya validez es garantizada por la severidad tal como Mayo insiste, finalmente también depende de los contextos de fondo. Lo que he mostrado en este capítulo señala que noción de inferencia científica que defiende Mayo no puede ser totalmente objetiva independientemente de estas consideraciones sobre los contextos de fondo. Al plantear una inferencia experimental que deja a un lado los contextos de fondo, Mayo inevitablemente está simplificando un proceso experimental realmente muy complicado y olvidando los elementos importantes en este proceso de la práctica científica.

Notas

1. Para evitar las críticas que hacen Howson y Urbach a la prueba Neyman-Pearson, Mayo no quiere que la aceptación o el rechazo de una hipótesis basada en el cálculo estadístico sea una cuestión de verdad o falsedad, sino que plantea las reglas más sofisticadas, a saber, la de aceptación (RA) y la de rechazo. La regla de rechazo es así:

RRi: Una diferencia tan grande como D_{obs} es una buena indicación de que μ (la media real) excede a μ' exactamente en la medida en que muy probable que si μ fuera tan pequeña como μ' , la prueba T^* habría resultado con una diferencia menor.

RRii: D_{obs} es una indicación pobre de que μ excede a μ' si es muy probable que la prueba T^* arroje una diferencia tan grande aún, si μ no es mayor que μ' . (Mayo,p,398)

La regla de aceptación es así:

RAi: Una diferencia tan pequeña como D_{obs} es una buena indicación de que μ es menor que μ' si y sólo si es muy probable que si la media real hubiera sido tan grande como μ' habría resultado una gran diferencia de la prueba T^* .

RAii: Una diferencia tan pequeña con D_{obs} es una indicación pobre de que μ es menor que μ' si es muy improbable que hubieran resultado en una diferencia más grande aún si la media hubiera sido tan grande como μ' . (Mayo,p,404)

No voy a discutir si estas maniobras evitan realmente los problemas que plantean Howson y Urbach. Pero está claro que no puede evitar el problema del nivel de significación que estudiaré más adelante. Antes de ejecutar cualquiera regla, o sea, RR o RA, es obligatorio asignar un nivel de significación, que es 0.03 en el ejemplo de libro de Mayo.(p,389)

2. Labovitz señala que hay tres puntos que Skipper, Guenther y Nass no analizan: "First, the authors do not place the arbitrariness of a significance level within the perspective of the general state of theory, knowkedge, and evidence. Instead, they emphasize the single test, whether or not it reaches a predetermined significance level, leads to no major decision. Few, if any, reserchers would accept or reject any statement on the basis of a single test. Second, Skipper *et al.* ignore cross-classification versus tests of significance arguments. While this is not their concern, their whole article is essentially meaningless if tests of significance are not applicable. Finally, the authors emphasize applied research and the lay, statistically unsophisticated audience. If statistically adroit colleagues are the prospective audience perhaps their suggestions are less useful." (Labovitz, 1970, p,167, note, 1)

Conclusión

En el capítulo uno de este tesis, vimos que Deborah Mayo en su libro "Error and the Growth of Experimental Knowledge" tiene tres propósitos interconectados. A saber, (1) hacer ver el papel importante que debe jugar la probabilidad no bayesiana en la filosofía de la ciencia; (2) reconstruir la metodología racional de las ciencias a partir de la justificación de reglas objetivas para la discriminación de las hipótesis (reglas garantizadas por la estadística del error); y finalmente (3) fundamentar una filosofía de la experimentación a través de un enfoque "por partes" y la reconstrucción de la metodología racional mencionada en (2). A partir de los análisis expuestos en los capítulos anteriores, podemos evaluar si Mayo logró lo que ella se propuso, y lo que es más importante, podemos aprender de las limitaciones que hemos detectado en su propuesta.

Creo que Mayo ha logrado mostrar cosas importantes en relación a su primer propósito. Primero, nos ha hecho ver que la probabilidad no bayesiana es por lo menos tan importante como la bayesiana para resolver ciertos problemas en la filosofía de la ciencia. En la práctica científica la estadística, especialmente la prueba de Neyman-Pearson, sirve como una herramienta muy importante para analizar los datos experimentales. Segundo, Mayo está en lo correcto cuando utiliza un enfoque "por partes" en su estudio de la experimentación, porque en cada etapa del desarrollo de un experimento, se pueden tener diferentes problemas. Cada uno de ellos debe resolverse de diferentes maneras en diferentes situaciones.

A pesar de esto, creo que Mayo no ha logrado cabalmente su propósitos (2) y (3). En el capítulo tres y el cuatro, hemos visto que la inferencia estadística que Mayo propone no puede resolver todos los problemas de la inferencia científica, ni puede garantizar reglas objetivas para la discriminación de hipótesis. Por un lado, la inferencia estadística no puede afrontar el problema de las clases de referencia, y por lo tanto, no puede modelar toda inferencia causal. Por otro lado, el criterio de la severidad del experimento (o prueba) no puede garantizar la objetividad de la inferencia estadística como Mayo piensa, ya que es necesario tomar en cuenta los contextos de fondo. Además, en el capítulo dos, hemos visto que no todas las hipótesis locales se relacionan de una manera unívoca con los experimentos por medio de consideraciones estadísticas, y que no todos los experimentos pueden reducirse a experimentos estadísticos. Todo esto nos obliga a concluir que el proyecto de Mayo no cumple con lo que promete. Más en general, las críticas que he formulando al proyecto de Mayo sugieren fuertemente que una filosofía de la experimentación sustentada en la búsqueda de una metodología unificada por medio de un tipo de argumentación estadística no puede tener éxito.

En el capítulo 1, mostramos que el propósito (2) de Mayo coincide con el intento de algunos filósofos, a los que podemos llamar objetivistas metodológicos, como Laudan y Kitcher. Estos filósofos buscan una metodología objetiva que sirva de norma universal para la racionalidad científica. Lo que estos objetivistas metodológicos proponen es un tipo de fundamentalismo metodológico, si bien cada uno lo elabora de una manera distinta. A manera de conclusión quiero sugerir que los problemas que he detectado con la propuesta de Mayo apoyan la tesis de que todo intento de objetivismo metodológico, sea del tipo propuesto por Laudan o Kitcher, o del tipo propuesto por Franklin o Mayo, está destinado al fracaso.

Hoy en día, muchos filósofos están dispuestos a reconocer que la complejidad de las prácticas científicas hace muy implausible que, ya sea en el nivel de las inferencias científicas o en la justificación de las metodologías, pueda encontrarse un fundamento único que permita de manera simple recoger el aporte epistemológico de las prácticas. En la introducción, mencionamos que A.C Crombie (1994) recientemente propuso la existencia de seis estilos de argumentación en la historia de la ciencia occidental. El estilo estadístico de la argumentación es uno de ellos. Estos seis estilos aparecieron y se desarrollaron independientemente en diferentes épocas y contextos, y ninguno de ellos puede reducirse a ningún otro. Además, no hay una jerarquía de importancia en estos seis estilos. Los científicos los usan según las diferentes circunstancias y los diferentes tipos de problemas que tienen que resolver. Según Crombie, la ciencia occidental es objetiva, porque la naturaleza es conocida y el conocimiento del ser humano de la naturaleza es acumulable. Crombie investiga la objetividad de la ciencia occidental desde un punto de vista antropológico y cultural. Sin embargo, eso no puede hacerse asumiendo un fundamento único, o una justificación fija de los métodos disponibles en las investigaciones científicas. Porque a través de la historia de la ciencia, los científicos no sólo cambian la manera de ver los problemas, sino también cambian los estilos de argumentación para resolver sus problemas. Por lo tanto, Crombie considera que hay que estudiar estos seis estilos de argumentación bajo los diferentes contextos históricos. (Crombie, 1994, p,7)

En sus contribuciones a la colección "Disunity of Science" (Galison y Stump (ed),1996), Ian Hacking y Arnold Davidson señalan que la ignorancia de la diversidad de los diferentes estilos de argumentación es una consecuencia de una idea muy popular: la ciencia está o tiende a estar unificada. Pero eso no es cierto. Como todos los contribuidores de este libro arguyen desde diferentes perspectivas, no parece existir ni un método ni un lenguaje unificado común a toda la ciencia. Los lenguajes científicos, las prácticas, las metas y las formas de argumentación son simplemente diversos.

Si consideramos aceptables los siguientes elementos: el rechazo de la idea de la unificación de la ciencia, la existencia de diferentes estilos de argumentación que Crombie muestra, y la diversidad y complejidad del uso de los estilos de argumentación, entonces veremos que tanto los proyectos de Franklin y Mayo como el objetivismo metodológico de Laudan y Kitcher, entre otros, carecen de plausibilidad. Veremos también que se refuerza la idea de que debemos estudiar la filosofía de la experimentación tomando en cuenta los contextos históricos y sociales, y que estos son realmente imprescindibles para entender las prácticas científicas.

Sin embargo, el rechazo del objetivismo metodológico no resuelve el problema de desarrollar una teoría del conocimiento científico que tome en serio a la experimentación. Las preocupaciones de Franklin y Mayo (como: ¿existen metodologías objetivas en las investigaciones científicas? ¿existen las normas epistémicas? etc.) siguen siendo preocupaciones genuinas. Los filósofos tienen que explicar el origen de la normatividad epistémica implícita en las prácticas científicas reconociendo la diversidad de lenguajes científicos, prácticas, metas y formas de argumentación. Hemos visto en la introducción a este trabajo varios intentos recientes de elaborar una segunda aproximación a estos problemas. En esta otra vertiente del nuevo experimentalismo, autores como Hacking y Galison intentan incorporar los razonamientos de los experimentos o de la ciencia de laboratorio en sus contextos históricos y sociales en circunstancias específicas. Pero este tipo de propuesta parece dejar sin responder las inquietudes de Franklin y Mayo y en general de los objetivistas metodológicos. El problema persiste: cómo evitar tomarse en serio este contextualismo sin caer en un relativismo en el que sea muy difícil establecer criterios normativos que permitan evaluar y ayudar dirigir el trabajo experimental, y más en general la determinación de qué es evidencia para qué. Un inicio de respuesta indudablemente tiene que apuntar en la dirección de rechazar la dicotomía implícita entre requerimientos epistémicos y sociales. Reconocer la falsa dicotomía requiere precisamente tomar en serio el tipo de “localismo” o contextualismo propuesto por Hacking y Galison, pero además, requiere elaborar la idea de que los experimentos tienen “vida propia”, y esto a su vez requiere abandonar la idea de que el conocimiento científico es conocimiento de teorías. En este siglo, muchos filósofos, por ejemplo, Neurath, Polanyi y Kuhn entre otros, han reconocido de diferentes maneras que el conocimiento científico no es puramente teórico, sino que también es práctico, y que no existe una línea tajante entre elementos epistémicos y elementos sociales.

No obstante la falsa dicotomía sigue influyendo la manera de pensar de muchos filósofos y sociólogos de la ciencia contemporáneos. Pickering señala que la escuela de sociología del conocimiento científico intenta reducir todas las prácticas científicas a una red conceptual sociológica, que se explica causalmente en términos de intereses y negociaciones sociales. Según

Pickering este tipo de red conceptual es demasiado idealizada y simplificada para estudiar los conocimientos científicos, porque no captura el complejo procedimiento por el que los científicos consiguen los conocimientos. (Pickering, 1992, pp.4-6) Joseph Rouse identifica y critica un presupuesto básico compartido por positivistas, realistas, racionalistas históricos y constructivistas sociales: todos ellos asumen que se requiere una legitimación global del conocimiento científico, ya sea que esta legitimación se obtenga por medio de algunos presupuestos epistémicos universales, o bien por medio de un número determinado de factores sociales. Rouse arguye que los proyectos de legitimación global no son plausibles, por que o bien buscan una norma universal o bien se comprometen con un descriptivismo total, de manera que no pueden captar qué es lo que sucede realmente en las investigaciones científicas concretas.(Rouse, 1996, pp.21-25)

El concepto de práctica ha sido introducido por otros autores contemporáneos como parte de un intento por resolver el problema. Sin embargo, ellos no comparten una manera fija para caracterizar las prácticas científicas. (Pickering, 1992; Buchwald, 1995; Galison y Stump, 1996; Rouse,1996; Martínez,1997, entre otros)

Como hemos visto en la introducción, tanto el sistema holista de autojustificación que Hacking propone como el concepto de restricciones que Galison plantea, son dos diferentes maneras de caracterizar las prácticas científicas, especialmente las prácticas experimentales. Desde diferentes perspectivas, tanto Hacking como Galison defienden la idea de que la objetividad de la evidencia sólo se da localmente por medio de las relaciones dialécticas entre teorías, instrumentos y manipulaciones.

Como parte de su rechazo a un proyecto global de legitimación, Rouse introduce el concepto de práctica científica. Para Rouse, se entiende a las prácticas científicas como las situaciones en que las actividades científicas pueden tener significación, por medio de la interacción con el mundo. Por lo tanto, las investigaciones científicas no son epistémicamente arbitrarias, porque los científicos comparten situaciones a través de instrumentos, técnicas o conceptos. Sin embargo, las prácticas científicas son dinámicas, porque por un lado, están siempre en proceso de reproducirse y cambiar según las diferentes situaciones que los científicos encuentran, y por otro lado, los diferentes científicos pueden tener diferentes interpretaciones de ellas. Es decir, no es necesario que los científicos compartan de todo una situación, ni que la compartan de una manera fija. Eso significa que las prácticas científicas son temporales, y que debemos estudiarlas histórica y socialmente. (Rouse, 1996, pp.25-34 y pp.133-157)

Martínez introduce el concepto de estructura heurística para caracterizar las prácticas científicas. Según Martínez, las normas epistémicas sí existen en la ciencia y también son muy importantes para los filósofos. Pero estas normas son contextuales e implícitas en las prácticas. Estas

normas se articulan en lo que Martínez llama una "estructura heurística". Una estructura heurística esta constituida por procedimientos heurísticos. Un procedimiento heurístico "es un proceso de transformación de un sistema material que tiene como consecuencia la implementación de una regla heurística." (Martínez,1997,p,24) Por ejemplo, el experimento para investigar la relación entre fumar y el temblor de las manos es un procedimiento heurístico que implementa las reglas heurísticas caracterizadas en la estadística no bayesiana.

Una estructura heurística es "una colección de procedimientos heurísticos jerárquicamente organizados alrededor de la tarea de resolver un cierto tipo de problema, o de construir o diseñar un cierto tipo de objeto, y que tienen una base material común." (Martínez,1997,p,25) El tipo de la prueba de Neyman-Pearson es una estructura heurística, también lo es el tipo de experimentos en que se usan las cámaras de niebla. Una estructura heurística lleva con ella sus propias normas epistémicas que son efectivas para resolver ciertos problemas usando ciertos instrumentos específicos. Además, una estructura heurística no tiene una forma fija, sino está siempre cambiando, dependiendo del cambio teórico y tecnológico.

La relación de apoyo mutuo que sin duda existe entre la nueva dirección en la filosofía de la ciencia que Martínez propone y esta segunda corriente del nuevo experimentalismo no es casual. El proyecto de Martínez trata de estudiar la filosofía de la ciencia através del trazo de un mapa (que está cambiando históricamente) de las razones epistémicas en sus contextos locales de prácticas científicas. Y esta segunda corriente del nuevo experimentalismo trata de hacer filosofía de la experimentación investigando históricamente la relación entre teorías, instrumentos y tecnologías de los experimentos. Por ello puede describirse como una parte imprescindible de la nueva dirección que Martínez propone para hacer filosofía de la ciencia.

La segunda corriente del nuevo experimentalismo no requiere que los filósofos compartan necesariamente una misma manera de enfocar los problemas, por el contrario, es recomendable explorar las diferentes perspectivas, para que puedan revelarse los diferentes aspectos de la ciencia. Por ejemplo, como hemos visto, Hacking intenta entender la ciencia de laboratorio estudiando la objetividad local en cada uno de los sistemas de autojustificación; y Galison intenta entender las complejas prácticas de la generación de los datos experimentales en la física moderna de partículas, analizando los diferentes niveles de las restricciones en diferentes pasos específicos. Usar el concepto de estructura heurística para estudiar la experimentación, también nos servirá como una nueva perspectiva para ver cómo las normas epistémicas (o reglas heurísticas) de los experimentos específicos se desarrollan históricamente bajo diferentes contextos sociales. Bajo este modo de hacer filosofía de la experimentación, los estudios de Mayo podría servirnos como una buena manera de

ver cómo los científicos usan las pruebas de Neyman-Pearson para hacer objetivamente las inferencias científicas en ciertos experimentos. Las pruebas de Neyman-Pearson (o la estadística del error, como Mayo llama) en este caso no es un fundamento para explicar toda la experimentación, sino que podría ser un sistema autojustificado en la perspectiva de Hacking, o un sistema de restricciones a la de Galison, o una estructura heurística al estilo de Martínez.

Las diferentes maneras de caracterizar las prácticas científicas (Hacking, Galison, Rouse, Martínez), muestran diferentes intentos prometedores para resolver los problemas que Mayo no puede resolver. La segunda corriente del nuevo experimentalismo es uno de esos intentos. Ellos comparten con Mayo el mismo propósito de explicar porqué la ciencia es objetiva, pero a partir de rechazar la falsa dicotomía mencionada y de introducir el concepto de prácticas, lo que permite evitar el dilema que Mayo tiene que afrontar: o bien sostiene la estadística del error como un fundamento universal para explicar la racionalidad científica, a costa de idealizar, simplificar o aún dejar sin explicar muchos elementos importantes y cruciales de las actividades científicas; o bien reconocer la limitación y la localización de éstas, pero abandonar inevitablemente toda norma epistémica y caer en el relativismo extremo.

Referencias Bibliográficas

- Ackermann, R** (1985) *Data, Instruments, and Theory* Princeton Univ Press.
- (1989) "The New Experimentalism" en *Brit. J. Phil. Sci.* Vol,40; pp.185-190.
- Braithwaite, R.B** (1953) *Scientific Explanation---A Study of the Function of Theory, Probability and Law in Science* Univ of Cambridge.
- Buchwald, J.D** (1993) "Design for Experimenting" en Horwich, P (edi.)
- Buchwald, J.D** (edi.) (1995) *Scientific Practice---Theories and Stories of Doing Physics* The Univ of Chicago Press.
- Cartwright, N** (1983) *How the Laws of Physics Lie* Clarendon Press. Oxford.
- (1989) *Nature's Capacities and Their Measurement* Clarendon Press. Oxford.
- Coffa, J.A** (1974) "Hempel's Ambiguity" en *Synthesis* 28, 141-164.
- Collins, H** (1975) "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics" en *Sociology* 9: pp.205-224.
- (1981) "Son of the Seven Sexes: The Social Destruction of a Physical Phenomenon" en *Social Studies of Science*, Vol. 11; pp.33-62.
- (1985) *Changing Order: Replication and Induction in Scientific Practice* London and Beverly Hills: Sage.
- (1994) "A Strong Confirmation of the Experimenter's Regress" en *Stud. Hist. Phil. Sci.* Vol.25; No,3; pp.493-503.
- Crombie, A.C** (1994) *Styles of Scientific Thinking in the European Tradition---The History of Argument and Explanation Especially in the Mathematical and Biomedical Science and Arts* Duckworth.
- Earman, J** (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* Cambridge, Mass.; MIT Press.
- Elston, R C y Johnson, W.D** (1983), *Essentials of Biostatistics* Philadelphia: F.A. Davis Company.
- Franklin, A** (1986) *The Neglect of Experiment* Cambridge Univ Press.
- (1990) *Experiment, Right and Wrong* Cambridge Univ Press.
- Feldman, T.S** (1990) "Late Enlightenment Meteorology" en T. Frängsmyr et al.(edi.) *The Quantifying Spirit in the Eighteenth* Unvi of California Press.
- Galileo, G** (1975) *Diálogo sobre los Sistemas Máximos*, editado por José Manuel Revuelta, ed. Aguilar. Buenos Aires.
- Galison, P** (1983) *How Experiments End* Univ of Chicago Press.
- (1988) "Multiple Constrains, Simultaneous Solutions" en PSA 1988, Vol 2; pp.157-163.
- (1995) "Context and Constraints" en J.D Buchwald(edi.); pp.13-41.
- (1997) *Image and Logic---A Material Culture of Microphysics* Univ of Chicago Press.
- Galison, P y Stump, D.J.** (edi.) (1996) *The Disunity of Science---Boundaries, Contexts, and Power* Stanford Univ Press.
- Giere, R.G** (1984) *Understanding Scientific Reasoning*, Holt, Rinehart and Winston.
- Good, I.J** (1967) "The White Shoe is a Red Herring" *The British Journal for the Philosophy of Science*, Vol 17, No.4, p,322.
- Hacking, I** (1965) *Logic of Statistical Inference* Cambridge Univ Press.
- (1983) *Representing and Intervening* Cambridge University Press.(la edición castellana, *Representar e Intervenir*, traducción por Sergio Martínez, Paidós, 1996. Todas las citas son de esta edición)
- (1988) "Philosophy of Experiment" en PSA Vol 2; pp.147-156.
- (1992) "The Self-Vindication of the Laboratory Sciences" en A. Pickering (edi). pp.29-64.
- Harré, R** (1981) *Great Scientific Experiments* Phaidon Press Limited.(la edición castellana, *Grandes Experimentos Científicos*, traducción por Luis Bou García, Labor, S.A; 1986. Todas las citas son de esta edición)

- Hempel, C.G (1965) *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, The Free Press, N.Y.
- Horwich, P (edi.) *World Changes---Thomas Kuhn and the Nature of Science* A Bradford Book, The MIT Press.
- Howson, C. (1997) "A Logic of Induction" en *Philosophy of Science*, Vol.64; pp.268-290.
- Howson, C. and Urbach, P (1989, primera edición) y (1996, segunda edición) *Scientific Reasoning: The Bayesian Approach* 1989, La Salle, Ill.: Open Court.
- Hull, C.D (1924) "The Influence of Tobacco Smoking on Mental and Motor Efficiency: An Experimental Investigation" *Psychological Monographs*, 33,no.3, pp. 131-142.
- Kahneman, D; Slovic, P y Tversky, A (1982) *Judgement under Uncertainty: Heuristics and Biases* Cambridge Univ Press, Cambridge.
- Kitcher, P (1993) *The Advancement of Science---Science without Legend, Objectivity without Illusions*, Oxford Univ Press.
- Kitcher, P y Salmon, W(ed.) (1989) *Minnesota Studies in the Philosophy of Science, Vol XII, Scientific Explanation*, Unvi of Minnesota Press.
- Kendall, P (1970) "Note on Significance Tests" en Morrison y Henkel, pp, 87-90;
- Karin D. Knorr, Roger Krohn y Richard Whitley (eds.)(1980) *The Social Process of Scientific Investigation. Sociology of the Science* Vol. IV; Reidel Publishing Company, Dordrecht-Holland.
- Labovitz, S (1970) "Criteria for Selecting a Significance Level: A Note on the Sacredness of .05" en Morrison, D y Henkel, R, pp.166-171.
- Latour, B y Woolgar, S (1986) *Laboratory Life: The Social Construction of Scientific Facts*, segunda edición, Priceton Univ Press.
- Laudan, L (1996) *Beyond Positivism and Relativism---Theory, Method and Evidence*, Westview Press.
- Lenoir, T (1998) "Practice, Reason, Context: The Dialogue Between Theory and Experiment" en *Science in Context* Vol 2. pp. 3-22.
- Lipton,P (1991) *Inference to the Best Exliation* Routledge.
- MacKenzie, D (1981) *Statistics in Britain 1865-1930: The Social Construction of Scientific Knowledge*, Edinburgh Univ Press.
- Martínez, S (1997) "La geografía de la razón científica: dependencia epistémica y estructura social de la cognición" manuscrito.
- Mayo, D. G (1994) "The New Experimentalism, Topical Hypotheses, and Learning from Error" en *PSA*, vol. 1; pp.270-279.
- (1996) *Error and the Growth of Experimental Knowledge* The Univ of Chicago Press.(todas las citas de Mayo sin otras alaraciones son de este libro)
- (1997) "Response to Howson and Laudan" en *Philosophy of Science*, Vol.64; pp.323-333
- Morrison, D.E y Henkel, R.E(edi.)(1970) *The Significance Test Controversy--A Reader* Aldine Publishing Company, Chicago.
- Peterson, O (1993) *Early Physics and Astronomy* Cambridge Univ Press.
- Pickering, A (1984) *Constructing Quarks: A Sociological History of Particle Physics* 1984, Edinburgh Univ Press.
- (1992) (edi) *Science as Practice and Culture* Univ of Chicago Press.
- Popper, K (1979) *Objective Knowledge: An Eolutionary Approach*, Oxford Univ Press.
- Porter, T.M (1986) *The Rise of Statistical Thinking 1820-1900* Princeton Univ Press.
- Rouse, J (1996) *Engaging Science---How to Understand Its Practices Philosophically* Cornell Univ Press.
- Salmon, W.C (1984) *Scientific Explanation and the Causal Structure of the World* Princeton Unvi Press.
- (1989) "Four Decades of Scientific Explanation" en Kitcher y Salmon(ed.)
- Selltiz, H (1968) *Research Methods in Social Relations*, New York: Holt, Rinehart and Winston.

- Selvin, H.C** (1970) "A Critique of Tests of Significance in Survey Research" en **Morrison, D.E y Henkel, R.E**, pp, 94-106.
- Skipper, J.K., Jr; Guenther, A.L y Nass, G** (1970) "The Sacredness of .05: A Note Concerning the Uses of Statistical Levels of Significance in Social Science", en **Morrison, D.E y Henkel, R.E**, pp,155-160.
- Suppes, P** (1969) "Models of Data", en *Studies in the Methodology and Foundations of Science*, pp,24-35 Dordrecht, The Netherlands: D.Reidel.
- Pinch.T** (1980) "Theoreticians and the Production of Experimental Anomaly: The Case of solar-neutrinos" In **Karin D. Knorr, Roger Krohn Y Richard Whitley** (eds.); PP.77-106.
- (1981) "The Sun-Set: The Presentation of Certainty in Scientific Life" en *Social Studies of Science* 11; pp.131-158.
- (1985a) "Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics" en *Social Studies of Science*;15; pp.3-36.
- (1985b) "Theory Testing in Science--The Case of Solar Neutrinos: Do Crucial Experiments Test Theories or Theorists?" en *Phil.Soc.Sci*; 15; pp.167-187.
- van Frassen, B** (1980)*The Scientific Image*, Oxford: Clarendon Press.
- Wilson.E.B** (1952) *An Introduction to Scientific Research*, McGraw-Hill Book Company, Inc.
- Wise. N** (edi.) (1995) *The Values of Precision* Princeton Univ Press.
- Winer, B.J** (1962) *Statistical Principles in Experimental Desigh*, Now York: McGraw-Hill.