DÍEZ, José A. y MOULINES, C. Ulises (1999): Fundamentos de Filosofía de la Ciencia. Barcelona: Ariel. Pp. 267-391.

Materiales del Seminario de Epistemología II. **La Universidad del Zulia**, Maracaibo, Venezuela.
Doctorado en Ciencias Humanas. Profesor: Dr. José Padrón Guillén, http://padron.entretemas.com.

CAPÍTULO 8. Análisis sincrónico de teorías I. Concepción axiomática: las teorías como cálculos interpretados

1. Teorías axiomáticas. 1.1. Cálculos y teorías axiomáticas: términos primitivos, axiomas y teoremas; definiciones y términos derivados. 1.2. Ejemplo: teorías del parentesco. Reducción y equivalencia. 1.3. Aritmética, teoría de conjuntos y lógica proposicional.

2. Teorías y modelos (Estructura. Interpretación. Modelo, realización, Realización posible; realización efectiva).

3. Caracterización general de las teorías empíricas como cálculos interpretados. 3.1. Teorías formales y teorías empíricas (Concepción Heredada. Términos teóricos. Definición implícita. Convencionalismo. Formalismo. Observación; observación directa). 3.2. Cálculos interpretados: vocabulario; axiomas y reglas de correspondencia (Términos teóricos. Términos observacionales. Contenido-interpretación empírica. Reglas de correspondencia, definiciones coordinativas, postulados de significación, enunciados interpretativos, definiciones operacionales. Entidades teóricas. Vocabulario: formal, teórico, observacional. Enunciados teóricos. Enunciados observacionales).

4. Las reglas de correspondencia y la cuestión de la eliminabilidad de los términos teóricos. 4.1. Ineliminabilidad de los términos teóricos (Reglas de correspondencia y definiciones explícitas-Enunciados reductivos; eliminativismo. Propiedades categóricas, propiedades disposicionales. Enunciados de reducción parcial. Analítico/sintético). 4.2. Eliminabilidad a lo Ramsey. 4.3. Eliminabilidad a lo Craig.

5. La distinción teórico/observacional y la naturaleza de la base empírica. 5.1. Entidades teóricas y distinción teórico/observacional (Entidades fenoménicas; qualia; fenomenismo. Entidades observables. Observación directa; observación indirecta). 5.2. Neutralidad teórica de los términos observacionales y carga teórica de los hechos (Experiencia-observación neutra, Interpretación teórica. Holismo. Carga teórica de los hechos. Conocimiento de fondo. Nuevos Filósofos de la Ciencia). 5.3. Observación y base empírica (Datos; base empírica; base de contrastación. Teórico/no-teórico. Observacional / no observacional. Vocabulario preteórico; vocabulario teórico. Principios internos; principios puente. Aplicación e interpretación empírica).

6. Consideraciones finales.

CAPÍTULO 9. Análisis sincrónico de teorías II. Concepciones historicistas: las teorías como proyectos de investigación

1. La revuelta historicista y la naturaleza sincrónica de las teorías (Nueva Filosofía de la Ciencia. Cambio y evolución de teorías. Análisis diacrónico y sincrónico de teorías).

2. Los paradigmas-matrices disciplinares de Kuhn. 2.1. Ciencia normal y ciencia revolucionaria (Resolución de enigmas. Anomalías. Crisis científica. Ciencia no-normal, ciencia extraordinaria. Revolución científica). 2.2. Paradigmas qua matrices disciplinares (Paradigmas. Generalizaciones simbólicas; leyes paradigmáticas; principios guía. Modelos; modelos ontológicos y heurísticos. Valores; precisión, simplicidad, fecundidad, compatibi-lidad. Ejemplares; aplicaciones empíricas. Ejemplares paradigmáticos. Significado empírico. Inconmensurabilidad).

3. Los programas de investigación de Lakatos (Conocimiento de fondo. Teoría interpretativa; teoría explicativa. Heurística positiva; heurística negativa. Núcleo del programa de investigación; cinturón protector. Programas progresivos, estancados y regresivos).

4. Las tradiciones de investigación de Laudan (Compromisos metafísicos; normas epistémicas. Articulación teórica. Resolución de problemas: problemas empíricos; problemas conceptuales. Evolución de las tradiciones. Coexistencia de tradiciones).

5. Consideraciones finales.

CAPÍTULO 10. Análisis sincrónico de teorías, III. Concepciones semánticas: las teorías como entidades modelo-teóricas

1. Teorías, enunciados y modelos. 1. l. Axiomas y modelos (Equivalencia e identidad de teorías. Identificación sintáctica. Identificación semántica o modeloteórica). 1.2. El enfoque modeloteórico (Definición de modelos. Aplicaciones empíricas. Afirmaciones empíricas. Complejidad teórica).

2. La noción de teoría de Suppes (Escuela de Stanford. Axiomatización mediante predicado conjuntista. Axiomas impropios; axiomas propios. Realizaciones posibles; realizaciones efectivas, modelos. Ejemplo: mecánica de partículas).

3. Adams y las aplicaciones intencionales (Interpretación pretendida. Medición funda-mental. Modelos pretendidos. Afirmación empírica. Autojustificación. Modelos de datos).

4. La familia semanticista. 4.1. Van Fraassen: espacios de estado; base empírica y observabilidad (Estados, espacios de estados; trayectorias y leyes de sucesión; regiones y leyes de coexistencia. Subestructuras empíricas. Empiricidad y observabilidad. Adecuación empírica. Empirismo constructivo. Equivalencia empírica; incompatibilidad teórica. Infradeterminación de la teoría por la experiencia. Antirrealismo). 4.2. Suppe: sistemas relacionales; fenómenos, datos y teorías (Sistemas relacionales; espacios de estados; leyes. Alcance pretendido; datos, observación. Verdad empírica; verdad teórica. Cuasi-realismo). 4.3. Giere: modelos e hipótesis teóricas (Modelo teórico. Hipótesis teóricas. Similitud. Realismo constructivista). 4.4. Sneed y la concepción estructuralista (Teoricidad relativa. Aplicaciones pretendidas. Datos. Ligaduras. Relaciones interteóricas).

5. La concepción estructuralista de las teorías. 5. l. El núcleo K (Modelos potenciales y modelos actuales. Condiciones de ligadura; ligadura global. T-teoricidad y modelos parciales; procedimiento de determinación). 5.2. Aplicaciones intencionales (Aplicaciones y modelos parciales. Carga teórica de los datos. Determinación intencional y paradigmática del dominio de aplicaciones). 5.3. L,, teorías como elementos teóricos (Contenido y aserción empírica. Contenido teórico y contenido empírico. Aserción empírica. Subsunción o encaje teórico). 5.4. Especialización. Las teorías como redes teóricas (Matrices disciplinares; complejidad de las teor(as. Especialización teórica, redes teóricas, redes conectadas, redes arbóreas; elemento teórico básico). 5.5. Vínculos interteóricos y botones (Leyes mixtas, leyes-puente y vínculos interteóricos. Vínculos y núcleo; vínculo global. Holones teóricos).

6. Consideraciones finales.

CAPÍTULO 11. Relaciones interteóricas

1. Concepto general de relación interteórica (Relaciones interteóricas. Relaciones interteóricas o identidad de teorías. Holismo; tesis Duhem-Quine).

2. Teorización (Términos T-no teóricos. Teorías subyacentes. Base empírica, contrastación y teorización. Teorización total y teorización parcial. Fundacionismo y coherentismo. Subestructuras. Determinación).

3. Reducción (Reduccionismo. Revoluciones científicas, inconmensurabilidad. Reducción exacta; reducción aproximativa. Reducción y cambio teórico. Condiciones de conectabilidad y derivabilidad. Constitución).

4. Equivalencia (Equivalencia y reducción. Equivalencia fuerte o estricta. Equivalencia débil o empírica).

5. Apéndice: Ciencia especial y ciencia básica; reducción, múltiple realizabilidad y superveniencia (*). 5.1. Distinciones previas: términos generales, conceptos expresados y entidades denotadas; acaeciniento-ejemplar y acaecimiento-tipo. 5.2. Identidad conceptual. Reduccionismo semántico. 5.3. Identidad de tipos o propiedades. Reduccionismo ontológico. 5.4. Múltiple realizabilidad. 5.5. Dualismo de propiedades con identidad de ejemplares y superveniencia. 5.6. Dualismo de propiedades con identidad de ejemplares y epifenomenismo. 5.7. Eliminativismo. 5.8. Dualismo de ejemplares.

CAPÍTULO 8

ANÁLISIS SINCRÓNICO DE TEORÍAS I. LA CONCEPCIÓN AXIOMÁTICA: LAS TEORÍAS COMO CÁLCULOS INTERPRETADOS

Con este capítulo iniciamos la presentación de los diversos análisis de la noción de *teoría empírica* con que se pretende elucidar la naturaleza y estructura de las teorías científicas. Como mencionamos brevemente en el capítulo 1, las teorías científicas son entidades que se extienden o perduran en el tiempo, que permanecen a través del cambio. Ello supone que el estudio puramente sincrónico que las considera como entidades estáticas, "congeladas", constituye sólo una primera aproximación que se debe completar con un análisis diacrónico que dé cuenta del carácter persistente de estas entidades, esto es, con una cinemática de teorías. En este capítulo y en los dos siguientes vamos a realizar la primera parte de la tarea, el estudio sincrónico, que se completará en el capítulo 12 con el análisis diacrónico. En las dos primeras secciones de este capítulo se presentan en detalle, y se ilustran con ejemplos puramente formales, las nociones de *teoría axiomática y de modelo (o realización);* en la primera se presenta además una primera idea de otras nociones que se estudiarán en otros capítulos, especialmente las de *contenido, reducción y equivalencia.* En la tercera sección se aplica la primera noción a las teorías empíricas y se introduce el análisis clásico de las teorías empíricas como cálculos axiomáticos empíricamente interpretados. En las dos secciones siguientes se discuten dos cuestiones especialmente importantes de este análisis, (i) la naturaleza y función de las reglas de correspondencia y (ii) la distinción teórico/observacional y el problema de la base empírica.

1. **Teorías axiomáticas**

Según cierta noción de *teoría,* una teoría es un conjunto de afirmaciones sobre un determinado ámbito de la realidad. Así, según esta concepción, la Mecánica Clásica (MC) consiste en una serie de afirmaciones sobre el movimiento de los cuerpos de tamaño medio; la Termodinámica (TM), sobre sistemas que interaccionan intercambiando energía; la Genética Mendeliana (GM), sobre la transmisión de rasgos en la generación de seres vivos; la Economía del Intercambio (El), sobre procesos de transferencia de bienes de consumo; la Aritmética de Peano (AP), sobre los números naturales y sus propiedades; la

Teoría de Conjuntos (TC), sobre las clases, conjuntos o colecciones; o una supuesta Teoría del Parentesco (TP) sobre los hechos que se derivan de las relaciones familiares. Concebidas como conjuntos de afirmaciones sobre un determinado ámbito, las teorías se analizan o reconstruyen como teniendo cierta estructura que expresa las relaciones que mantienen entre sí las diversas afirmaciones y los diversos términos o conceptos con los que se realizan tales afirmaciones. La noción formal que expresa esa estructura es la de *cálculo axiomático o,* simplemente, *teoría axiomática, y* se aplica por igual a teorías empíricas y a teorías puramente formales; la diferencia radica en que esta noción agota el análisis de las segundas pero no el de las primeras, que se debe completar con elementos adicionales. Ahora vamos a contemplar exclusivamente este elemento común.

1.1. CÁLCULOS Y TEORÍAS AXIOMÁTICAS: TÉRMINOS PRIMITIVOS, AXIOMAS Y TEOREMAS; DEFINICIONES Y TÉRMINOS DERIVADOS

La idea básica es que una teoría o conjunto de afirmaciones se puede "resumir" o "concentrar" en algunas de sus afirmaciones, de las que se derivan todas las restantes mediante un proceso de inferencia deductiva. A las afirmaciones que forman parte de ese "conjunto-resumen", consideradas *primitivas,* se las denomina `axiomas', y a las afirmaciones que se deducen de los axiomas, consideradas *derivadas,* se las denomina `teoremas'. Si llamamos *contenido* de una teoría al conjunto de todas sus afirmaciones, entonces tal contenido se encuentra ya *completo,* aunque *implícito,* en los axiomas (cf. cap. 2, §2). El contenido de la teoría, la información que da, es por tanto el conjunto de consecuencias lógicas de los axiomas. Toda afirmación se encuentra ya en los axiomas: explícitamente, si es un axioma, o implícitamente, si no lo es, en cuyo caso se puede hacer explícita deduciéndola lógicamente, esto es, obteniéndola como teorema a partir de los axiomas. Los teoremas, por tanto, no contienen información nueva, sólo hacen explícita información contenida implícitamente en los axiomas. Por supuesto, para que esto sea así es preciso que de los axiomas en cuestión se sigan efectivamente todas las afirmaciones de la teoría o sea, que el conjunto de axiomas sea *suficiente,* o como se dice a veces, *completo.* Al axiomatizar una teoría se pretende dar con un conjunto completo de axiomas para ella. Ésta es pues una condición necesaria para una buena axiomatización.

Es fundamental ver que la anterior condición, aunque necesaria, no es suficiente. Que de los axiomas se obtengan todas las afirmaciones no basta para una buena axiomatización, pues de lo contrario el simple conjunto de todas las afirmaciones sería ya un buen conjunto de axiomas. De tal conjunto se obtienen efectivamente, cómo no, todas las afirmaciones; es, si se quiere, un conjunto de axiomas, pero no es un *buen* conjunto de axiomas pues viola el espíritu que inspira la axiomatización, a saber, dar una versión lo más "resumida" o "concentrada" posible de la teoría. Así pues, es un principio metodológico general que los axiomas han de constituir un conjunto *mínimo* de afirmaciones primitivas, ningún axioma debe ser deducible de los restantes; o, como se dice técnicamente, los axiomas deben ser *independientes* entre sí (para una ejemplificación de ésa y otras nociones, cf. más abajo el ejemplo de las teorías del parentesco). Un buen conjunto de axiomas para una teoría es por tanto un subconjunto de sus afirmaciones que sea completo y cuyos miembros sean independientes entre sí. Es importante señalar que estas condiciones no determinan un único subconjunto de tales afirmaciones. Dada una teoría (en sentido intuitivo), siempre hay más de un subconjunto completo e independiente de afirmaciones, siempre hay axiomatizaciones alternativas.

Hasta aquí hemos hablado sólo de las afirmaciones de la teoría. Debemos hablar ahora de los constituyentes de estas afirmaciones, los términos o conceptos de la teoría.' Como veremos, la referencia a los términos introducirá una complicación adicional en las relaciones entre las afirmaciones, pues además de axiomas y teoremas intervienen entonces las definiciones.

Los términos de una teoría, los constituyentes de sus afirmaciones, expresan el aparato conceptualizador de la teoría, esto es, el aparato con el que se pretenden capturar las entidades de diverso tipo (objetos individuales y sus propiedades y relaciones) que conforman el ámbito de la realidad del que se ocupa la teoría. Así, por ejemplo, MC habla de partículas, masas, velocidades, etc.; AP habla de la propiedad de ser número natural, del cero, de la función siguiente, etc.; TP habla de padres, hermanos, abuelos, etc. Los términos de las teorías, contempladas éstas en su estadio intuitivo, son susceptibles también en general de cierta simplificación. Por ejemplo, si en MC disponemos ya de 'posición' y `tiempo' podemos prescindir como término primitivo de `velocidad', pues se puede introducir o definir a partir de los anteriores; o si en TP disponemos ya de `progenitor' y `hermano', podemos prescindir como término primitivo de `tío', pues se puede introducir o definir a partir de los anteriores. Esta introducción de nuevos términos a partir de otros anteriores supone la entrada en juego de otro tipo de "afirmaciones" o enunciados, las definiciones, pues sólo mediante enunciados (o esquemas de tales) es posible explicitar el modo en que se introduce un término nuevo a partir de otros anteriores. Las definiciones siempre tienen la forma de una equivalencia del tipo:

(1) *"a(t(x,, ..., x,,)) SYSSdef (3(t,, ..., t_k, x,, ..., xJ" (n ? 0, k ?1).*

Aquí *t* es el nuevo término y *t,, ...,* t_k, son términos ya disponibles, esto es, términos primitivos o ya definidos con anterioridad *a t; n* indica el número de variables a las que se aplica el término, esto es, su ariedad; a y 0 son funciones proposicionales. Ésta es la forma estándar que claramente presentan las definiciones de los relatores (n-ádicos), que nombran propiedades o relaciones entre individuos; por ejemplo, `ser impar': "x es impar $syss_{def}$ *x* dividido entre 2 da de resto 1" (análogamente, p.ej., con `ser múltiplo de'). Pero los términos de un lenguaje no siempre son relatores, puede haber también términos singulares (p.ej. `1') y functores (p.ej. `el siguiente de ...', `la intersección de ... y ...') que nombran, respectivamente, a individuos y a funciones-operaciones entre individuos. En

1. Términos o conceptos, respectivamente, según se entienda por `afirmación' el enunciado mismo o su contenido. De acuerdo con la práctica dominante en filosofía de la ciencia, seguiremos ahora en general la primera interpretación, y diremos por tanto que los constituyentes de las afirmaciones son términos.

principio parece que las definiciones de términos singulares y de functores no se ajustan a la forma (1) sino a estas otras:

(2) "$t =_{def} \gamma(t_1, \ldots, t_k)$" para términos singulares, y

(3) "$t(x_1, \ldots, x_n) =_{def} \gamma(t_1, \ldots, t_k, x_1, \ldots, x_n)$" para functores (n-ádicos),

donde en ambos casos la parte derecha, "$\gamma(\ldots)$" es una *descripción* que usa otros términos ya disponibles; por ejemplo: "$1 =_{def}$ el siguiente de 0"; "$x \cap y =_{def}$ el conjunto cuyos elementos son los elementos comunes a x e y". Sin embargo, es fácil ver que estas definiciones se pueden expresar también mediante una equivalencia de la forma (1), esto es, respectivamente, mediante:

(2') "para todo z: $z = t$ *syss*$_{def}$ $z = \gamma(t_1, \ldots, t_k)$",

(3') "para todo z: $z = t(x_1, \ldots, x_n)$ *syss*$_{def}$ $z = \gamma(t_1, \ldots, t_k, x_1, \ldots, x_n)$".

Por tanto, la forma general de la definición queda bien expresada mediante (1) (ligeramente modificada pues, como el lector habrá advertido, (2') y *(3')* contienen cuantificadores y (1) no). Ahora bien, una vez señalado esto, lo usual es introducir los términos singulares y los functores mediante (2) y *(3)* respectivamente, y así lo haremos también aquí.

Las definiciones no son afirmaciones del mismo tipo que los axiomas y los teoremas, no son afirmaciones sustantivas de la teoría sino que expresan meras abreviaturas notacionales. Esto se expresa diciendo que las definiciones deben cumplir dos requisitos: han de ser *a) eliminables y b) no creativas o* inocuas. Lo primero significa que cualquier afirmación que contenga un término definido ha de poder eliminarse usando la definición que introduce dicho signo; esto es, con ayuda de la definición se debe poder probar que tal afirmación es equivalente a otra que no contenga dicho signo, y en última instancia, si eliminamos los otros signos definidos previamente, equivalente a otra afirmación que contenga sólo signos primitivos. Lo segundo significa que si tenemos una afirmación que involucra el término definido *t* cuya prueba recurre, además de a los axiomas y otras definiciones previas, a la definición de *t, su* afirmación equivalente resultante de eliminar *t* ha de poder probarse sin recurrir a la definición de *t, y si* se han eliminado todos los términos definidos, ha de probarse a partir de los axiomas solos. Nótese que en caso contrario la presunta definición contendría subrepticiamente información sustantiva, no sería una mera abreviatura terminológica. En este sentido las definiciones son inocuas, no añaden nada al contenido de la teoría; por ello son teóricamente prescindibles, lo que se dice con su ayuda se puede decir igualmente sin ellas. Por supuesto que si un teorema contiene un término definido, para probarlo no nos bastan los axiomas, necesitamos además al menos una definición, la que ha introducido dicho término (y sólo esa definición si es el único término no primitivo del teorema). Pero si el término se ha introducido correctamente, mediante una definición eliminable y no creativa, ese teorema es equivalente a otro cuya prueba no requiere dicha definición, y en última instancia equivalente a otro que contiene sólo términos primitivos cuya prueba recurre exclusivamente a los axiomas. Las definiciones son pues prescindibles, todo lo que se dice con su ayuda se puede decir sin ella; no se puede decir exactamente de la misma forma, si dicha forma usa términos no primitivos, pero sí de otra equivalente que sólo use términos primitivos. Ahora bien, aunque las definiciones son teóricamente superfluas, no lo son en la práctica de la construcción y aplicación de una teoría; en efecto, para teorías de un mínimo de complejidad conceptual y fuerza expresiva, el prescindir totalmente de definiciones haría a éstas inmanejables y prácticamente incomprensibles. Las definiciones poseen un gran valor de "economía intelectual" en la construcción de las teorías.

En general al axiomatizar una teoría se pretende reducir al mínimo no sólo las afirmaciones primitivas o básicas, los axiomas, sino también los términos primitivos. Es cierto que en esto último la práctica es un poco más laxa y a veces se ofrecen axiomatizaciones en las que algunos términos básicos son definibles a partir de otros, pero ello casi siempre responde a motivaciones pedagógicas (y por los mismos motivos, aunque es menos usual, se dan a veces axiomas no independientes del resto); el ideal estrictamente científico es siempre el mismo: no presentar como primitivo lo que pueda ser derivado, ya sean afirmaciones o conceptos. A veces las simplificaciones de axiomas y términos van de la mano, pues un axioma puede ser eliminado si determinado término, considerado primitivo hasta entonces, se define apropiadamente. Por ejemplo, en TC se puede dar una axiomatización que contenga como término primitivo el signo functor de par ordenado `$< \, , >$' y un axioma que regule su comportamiento, a saber "$<x,y> = <z,v>$ syss x = z y y = v"; pero esta axiomatización se puede simplificar, se puede eliminar dicho término como primitivo mediante la definición "$<x,y> =aef\{x,\{x,y\}\}$", de modo que el mencionado axioma se convierte en un teorema que se prueba a partir de los restantes axiomas y de dicha definición.

Hasta ahora hemos presentado las cosas como si una teoría axiomática fuese el resultado de axiomatizar, reconstruir axiomáticamente, una teoría "en estadio intuitivo" consistente en una serie de afirmaciones sobre un ámbito, afirmaciones que usan una serie de conceptos. Entonces la teoría axiomática resulta de seleccionar apropiadamente algunos de los conceptos como primitivos, seleccionar algunas de las afirmaciones que contienen sólo dichos conceptos como verdades primitivas o axiomas, definir el resto de conceptos que ya usa la teoría, y probar como teoremas el resto de afirmaciones, usando en las pruebas sólo los axiomas cuando las afirmaciones contengan sólo términos primitivos, o usando además las definiciones cuando contengan términos no primitivos. Ésta es la presentación que más naturalmente se corresponde con lo que de hecho ocurre históricamente, pues casi siempre la formulación de una teoría axiomática es el resultado de axiomatizar, en este sentido, una teoría en estadio intuitivo. Según este modo de presentarlo, de los términos no primitivos "ya se dispone" en la teoría.

La distinción entre axiomas y teoremas, y entre términos primitivos y definidos, expresa la "dependencia" de unas afirmaciones respecto de otras y de unos conceptos respecto de otros. Esta dependencia tiene claro sentido en una teoría ya axiomatizada ("regimentada"), pero en una teoría intuitiva (en el estadio intuitivo de una teoría) es dudoso que se puedan identificar relaciones de dependencia o prioridad en este sentido. Desde una perspectiva preaxiomática, todas las afirmaciones y todos los conceptos están, por así decir, "al mismo nivel". Otra cosa son las relaciones de prioridad de otro tipo, por ejemplo epistémico, esto es, qué afirmaciones se consideran mejor fundamentadas, o más fácilmente cognoscibles. De eso no vamos a decir nada aquí, sólo que esta dependencia epistémica es en principio independiente de la que se introduce al axiomatizar y que, por tanto, la axiomatización no tiene por qué seguir criterios de prioridad epistémica.

En la presentación axiomática de una teoría se ofrece simplemente una serie de términos primitivos y una serie de afirmaciones primitivas que usan dichos términos, y los términos definidos se introducen, si se quiere, después para abreviar los teoremas que resultarían de escritura excesivamente larga si se usaran sólo los términos originales. Así vista, una teoría parece "que parte de cero", pero no hemos de olvidar que en la mayoría de los casos lo que se pretende es "poner orden" en un cuerpo de afirmaciones previamente existentes. Es importante señalar, además, que este proceso no siempre consiste en una mera ordenación, pues a veces la versión intuitiva incluye afirmaciones incompatibles entre sí y al axiomatizar se debe tomar partido por alguna de las alternativas. De hecho, algunas de las axiomatizaciones han surgido precisamente como respuesta a determinadas inconsistencias, o en general dificultades, descubiertas en la versión intuitiva; tal es el caso de las teorías axiomáticas de conjuntos, motivadas principalmente por la paradoja de Russell, o de las axiomatizaciones de las geometrías no euclídeas, que surgieron del intento de probar la independencia en la geometría euclídea del axioma de las paralelas respecto de los restantes.

Antes de ver algunos ejemplos es conveniente señalar que este análisis de las teorías, o en general cuerpos de conocimiento, plantea inmediatamente cuestiones fundamentales relativas al significado de los términos teóricos y a la justificación de las afirmaciones de la teoría. El significado de los términos definidos se retrotrae al de los términos primitivos a través de las definiciones. La justificación de los teoremas se retrotrae a la de los axiomas a través de las pruebas de aquéllos a partir de éstos. Como aprendemos pronto de niños, no toda afirmación se puede derivar de otras, ni todo término se puede definir eliminativamente a partir de otros. Los términos primitivos y los axiomas son los primitivos en los que nos detenemos; pero entonces el significado de los términos de la teoría pende en última instancia del significado de términos para los que no hay definición explícita, y la justificación de todas las afirmaciones de la teoría pende en última instancia de la justificación de afirmaciones para las que no hay demostración que parta de otras. ¿Cómo adquieren aquéllos significado y éstas justificación? Nótese que, por más que dispongamos de cierta "preconcepción" del significado de los términos primitivos, ello corresponde al estadio intuitivo o preaxiomático de la teoría. Desde una perspectiva axiomática, lo único que especifica explícitamente la teoría al fijar los términos primitivos es su categoría lógica, esto es, si se trata de relatores, functores o términos singulares, categoría que nos permite combinarlos correctamente de acuerdo con la gramática del lenguaje lógico que se utilice. En cuanto a las afirmaciones primitivas, lo único que hace la teoría desde un punto de vista formal es elegir, de entre las infinitas combinaciones bien formadas de términos primitivos y signos complementarios (lógico-matemáticos, según la teoría formal que se presuponga), algunas de ellas para fijarlas como axiomas. La pregunta por el significado de los términos primitivos y por la justificación de los axiomas de la teoría surge pues inmediatamente al contemplar las teorías como sistemas axiomáticos. De esta cuestión nos ocuparemos por extenso más adelante en éste y en los dos próximos capítulos, de momento nos interesa tan sólo la estructura formal de los sistemas axiomáticos.

Para fijar las ideas anteriores vamos a dar el esbozo de algunas teorías axiomáticas, principalmente de las ciencias formales que es donde más claramente, al menos hasta mediados del siglo XX, ha sido aplicada esta noción; lo peculiar de cierto modo de entender las teorías empíricas es que considera que la naturaleza y estructura de las teorías empíricas también se expresa adecuadamente mediante la noción de teoría axiomática, adecuadamente completada para dar cuenta de las peculiaridades del conocimiento empírico frente al lógico-matemático. Antes de ver los ejemplos reales provenientes de las ciencias formales, y puesto que no cabe suponer en general el conocimiento del contenido intuitivo de dichas teorías, comenzaremos con una pseudoteoría cuyo contenido nos es familiar a todos, la "Teoría" del Parentesco. Vamos a ver cómo se puede expresar este cuerpo de afirmaciones como teoría axiomática en el sentido elucidado, de modo que capte "las verdades sobre el parentesco" que conocemos. El esquema es siempre el mismo: n términos primitivos $t_1, ..., t_n$, m axiomas $A_1, ..., A_m$; teoremas $T_1, T_2, ...$ con términos primitivos; p definiciones-abreviaturas $D_1, ..., D_p$ que introducen p términos derivados $t_{n+1}, ..., t_{n+p}$ y finalmente nuevos teoremas que contienen también términos derivados. Toda afirmación de la teoría ha de estar constituida exclusivamente por términos propios de la teoría, primitivos o definidos, más vocabulario lógico ('todo', `y', `no', etc.).

1.2. EJEMPLO: TEORÍAS DEL PARENTESCO. REDUCCIÓN Y EQUIVALENCIA

Vamos a presentar aquí, como ejercicio, algunas "teorías del parentesco" de entre las varias que es posible construir. Estas teorías expresan, mejor o peor, las afirmaciones básicas relativas a ese ámbito de la realidad constituido por las relaciones de parentesco sanguíneo o biológico. Además de fijar las ideas anteriores, estos ejemplos servirán para introducir algunas ideas nuevas concernientes a las posibles relaciones que pueden darse entre diversas teorías. Para una mayor precisión, la presentación de las teorías debería utilizar el lenguaje formal de primer orden, pero de momento, y para facilitar la lectura, escribiremos aquí en general los diversos enunciados de las teorías en lenguaje informal; sólo daremos como muestra la versión formal de los axiomas en el primer ejemplo. El lector debe notar que, como señalamos más arriba, los enunciados contienen exclusivamente términos introducidos explícitamente por la teoría o términos puramente lógicos.

Teoría del Parentesco 1 (TP1):

Términos (o conceptos) primitivos:
C1. Progenitor (relator diádico): P
C2. Varón (relator monádico): V
C3. Hembra (relator monádico): H

Axiomas:

A1. Todo individuo es varón o hembra, y sólo una de las dos cosas.
Vx(VxH-'Hx)

A2. Todo individuo tiene al menos un progenitor varón.
Vx3y (Vy *A Pyx)*

A3. Todo individuo tiene al menos un progenitor hembra.
Vx3y *(Hy A Pyx)*

A4. Todo individuo tiene como máximo dos progenitores.
*Vx,y,z,t(PxtAPytAPzt-*z=xvz=yvx=y)*

A5. Si un individuo es progenitor de otro, éste no lo es de aquél.
Vx,y (Pxy -> - PYx)

Con estos axiomas ya se pueden probar algunos teoremas:

T1. Todo individuo tiene exactamente dos progenitores.
(Prueba: De A2, A3 y Al se sigue que tiene al menos dos progenitores diferentes, de ello y de A4, se sigue que tiene exactamente dos. QED)

T2. Nadie es su propio progenitor.

T3. Para todo individuo existen como mínimo dos hembras que son progenitoras de progenitores de dicho individuo.

Como muestra T3, algunos teoremas pueden ser muy largos, por lo que es conveniente introducir algunas abreviaturas o definiciones (las incompletas las puede completar el lector; nótese que según la noción de hermano que se va a definir, todo individuo es hermano de sí mismo):

DI. Padre: *x* es padre de y syss*br* x es progenitor de *y* y x es varón.
D2. Madre: x es madre de y syss,,*ef* x es progenitor de y y x es hembra.
D3. Her: x es her de y syss,cj x e y tienen los mismos progenitores.
D4. Hermano: x es hermano de *y* syss *J*x es her de *y* y x es varón.
D5. Hermana: x es hermana de y syss, f
D6. Hij: x es hij de y syss, *y* es progenitor de x.
D7. Hijo: x es hijo de y syss *def*
D8. Abu: x es abu de y syss *ef* x es progenitor de algún progenitor de y.
D9. Ti: x es ti de y *syss,,ef*

El lector puede dar las definiciones restantes (sob, sobrino, sobrina, prim, primo, etc.). Con estas definiciones podemos abreviar, p.ej., T3 como T4 y, en general, expresar muchas de las afirmaciones de la teoría intuitiva mediante términos no primitivos:

T4. Todo individuo tiene como mínimo dos abuelas.
T5. Todo individuo tiene exactamente un padre y una madre.
T6. Todo individuo es her de sí mismo.

T7. Los her de her son her entre sí.
T8. Nadie es su propio padre.
T9. Si uno es hij de otro, éste no lo es de aquél.

TP1 es un ejemplo sencillo de teoría axiomática pero que contiene ya *todo lo* esencial; toda teoría axiomática, por muy complicada que sea, tiene *exactamente esta estructura. TPI* sólo contiene relatores y, como hemos visto antes, las teorías pueden contener también como términos, primitivos o derivados, functores y términos singulares (p.ej. podríamos haber dado una teoría "bíblica" del parentesco con los términos singulares primitivos `Adán' y `Eva', en cuyo caso deberíamos modificar algunos axiomas pues A2 y A3 no valen para Adán y Eva). El número y variedad categorial de los términos de una teoría aumenta considerablemente su complejidad, pero la estructura es esencialmente la misma en todos los casos, esto es, como la que TP1 ejemplifica. Por otro lado, como teoría del parentesco TP1 no es muy buena, pues no permite obtener como teoremas afirmaciones de la teoría intuitiva como "nadie es bisabu de sí mismo". La siguiente teoría, que amplía TP1 añadiendo el término nuevo `ancestro' y sus correspondientes axiomas, es un poco mejor.

s

Teoría del Parentesco 2 (TP2):

Términos primitivos:
C1. Ancestro (relator diádico)
C2. Progenitor (relator diádico)
C3. Varón (relator monádico)
C4. Hembra (relator monádico)

Axiomas:
Al. Si uno es ancestro de otro, éste no lo es de aquél.
A2. Si uno es ancestro de otro y éste lo es de un tercero, entonces el primero lo es del último.
A3. Si un individuo es progenitor de otro, también es su ancestro.
A4. Si un individuo es ancestro de otro, entonces le conecta con éste una secuencia finita de progenitores.
A5. Todo individuo es varón o hembra, y sólo una de las dos cosas.
A6. Todo individuo tiene al menos un progenitor varón.
A7. Todo individuo tiene al menos un progenitor hembra.
A8. Todo individuo tiene como máximo dos progenitores.
A9. Si un individuo es progenitor de otro, éste no lo es de aquél.

).

Hemos dado los axiomas de TP2 añadiendo simplemente los cuatro primeros a los de TP1, *extendiendo* o aumentando TP1 con un nuevo término primitivo y nuevos axiomas. El resultado no es muy feliz, pues ahora A9 es redundante, no es independiente del resto pues se sigue de A l y A3. Construyamos ahora una nueva teoría del parentesco simplemente eliminando A9 de TP2:

Teoría del Parentesco 3 (TP3):

Los mismos términos primitivos y derivados que TP2 y los mismos axiomas salvo A9.

TP3 está un poco mejor que TP1 pues se obtienen como teoremas todos los de TP1 más algunas afirmaciones que se le escapaban a TP1, como "nadie es bisabu de sí mismo", y además otras afirmaciones deseables sobre el nuevo concepto, tales como "todo individuo tiene tantos ancestros varones como hembras" y "dos individuos tienen los mismos progenitores si y sólo si tienen los mismos ancestros". TP3 presenta además una característica novedosa respecto de TP1. TP1 contiene, además de sus términos primitivos, sólo términos lógicos, términos de la lógica de primer orden (LPO) que es por tanto la teoría presupuesta por TP1 y "dentro de la cual" realizamos nuestras demostraciones. Pero TP3, en su A4, contiene un término extralógico, el término `secuencia finita'. Este término no pertenece a LPO sino a otra teoría formal específica más rica que LPO, la aritmética o, alternativamente, la teoría de conjuntos. Algunas teorías pueden por tanto utilizar como recursos formales adicionales (términos y principios), no sólo los de la lógica que utilice sino también los de alguna teoría matemático-formal. En TP3 esos recursos teóricos formales adicionales son muy sencillos, pero en teorías físicas altamente matematizadas pueden incluir partes muy elevadas de la matemática (cálculo diferencial, teoría de tensores, etc.).

TP3 ilustra además otro hecho. Diremos que una teoría es *inmediatamente simplificable* si sus axiomas no son independientes. Vimos que *TP2* era una extensión de TP1 inmediatamente simplificable y eliminando el axioma redundante obteníamos TP3, que ya no es inmediatamente simplificable. Pero aunque TP3 es mejor que *TP2,* no es todo lo buena que podría ser pues es simplificable en otro sentido menos inmediato pero igualmente importante, sentido en el que desempeñan un papel fundamental los términos primitivos. El sentido es el siguiente: podemos definir alguno de sus términos primitivos en función de los restantes de modo que alguno de los axiomas pase a ser deducible del resto, esto es, de modo que la nueva teoría sea inmediatamente simplificable. Eso es lo que pasa con TP3, pues puedo definir `progenitor' en función de 'ancestro' (un progenitor es un "ancestro de primera generación") obteniendo una teoría inmediatamente simplificable.

Teoría del Parentesco 4 (TP4):

Los mismos términos primitivos que TP3 salvo `progenitor', que se introduce ahora como término definido del siguiente modo:

x es progenitor de *y* syss$_{def}$ x es ancestro de y y no hay ningún individuo z tal que x es ancestro de z y z es ancestro de y.

Los mismos axiomas que TP3.

Se dirá que ahora los axiomas de TP4 no pueden ser directamente los de TP3 pues contienen un término, `progenitor', que ahora no es primitivo. Pero nada impide que una teoría contenga axiomas con términos definidos; estos términos son meras abreviaturas y, si se prefiere, podemos escribir los axiomas sin ellos, pero también podemos hacerlo con su ayuda si usando sólo términos primitivos resultan muy engorrosos, como ocurriría con los axiomas 3, 6, 7, 8 y, en especial, 4 de TP4 (que ahora diría que no hay secuencias infinitas de "ancestros consecutivos"). Pues bien, los escribamos como los escribamos, en su actual forma o "deshaciendo" la definición de `progenitor', el resultado es que ahora A3 se puede probar como teorema, es decir, TP4 es inmediatamente simplificable.

Esta situación nos permite una primera aproximación intuitiva a dos conceptos de los que nos ocuparemos por extenso más adelante, los de *reducción y equivalencia* (cf. cap. 11). Obtengamos ahora una nueva teoría eliminando de la teoría inmediatamente simplificable TP4 el axioma redundante A3.

Teoría del Parentesco 5 (TP5):

> Los mismos conceptos primitivos y derivados que TP4 y los mismos axiomas menos A3.

Olvidemos por el momento TP2 y TP4, que no son buenas teorías axiomáticas al ser inmediatamente simplificables, y centrémonos en TP1, TP3 y TP5, buenos sistemas axiomáticos con todos sus axiomas independientes. Tal como hemos dado con ellas, deberían estar claros ahora los siguientes hechos: *a)* TP3 y TP5 "dicen lo mismo", tienen el mismo contenido; b) TP3 y TP5 "dicen al menos tanto como" TP1, todo el contenido de TP 1 es también contenido de éstas. En el primer caso decimos que TP3 y TP5 son *equivalentes,* en el segundo que ambas *reducen TP1.* Intuitivamente: una teoría *T* reduce otra *T'* si el contenido de *T'* es parte (quizá no estricta) del contenido de *T;* una teoría *T* es equivalente a otra T' cuando tienen el mismo contenido, es decir, cuando se reducen mutuamente.

Tal como hemos obtenido nuestras teorías, quizá estas relaciones no parezcan muy interesantes. TP3 es una simplificación de TP2, que es una simple ampliación (inmediatamente simplificable) de TP1, por lo que poco tiene de sorprendente que TP2, y con ella TP3, reduzcan TP1; en realidad TP3 tiene incluso todos los conceptos primitivos de TPI, y algunos más, nada muy sorprendente pues que diga lo mismo que ella y quizá algo más. Pero es esencial destacar que la reducción no exige que la teoría reductora contenga como primitivos los términos primitivos de la reducida; TP5 reduce TPI y entre sus términos primitivos ('ancestro', `varón', `hembra') no están todos los primitivos de TP1 ('progenitor', `varón', `hembra'). Y lo mismo ocurre respecto de la equivalencia o reducción recíproca. TP3 y TP5 son equivalentes y sin embargo no tienen exactamente los mismos términos primitivos. En este caso particular, los términos primitivos de una de las teorías, TP3 ('ancestro', `progenitor', `varón', `hembra'), incluyen los primitivos de su equivalente TP5. Pero no es necesario ni que tengan los mismos términos primitivos ni siquiera que los de una lo sean también de la otra. Como ejemplo tómese la siguiente teoría del parentesco.

Teoría del Parentesco 6 (TP6):

Términos primitivos:
C1. Padre (relator diádico)
C2. Madre (relator diádico)
C3. Varón (relator monádico)
C4. Hembra (relator monádico)
Axiomas:
A1. Todo individuo es varón o hembra, y sólo una de las dos cosas.
A2. Todo individuo tiene exactamente un padre.
A3. Todo individuo tiene exactamente una madre.
A4. Los padres de alguien son varones y las madres hembras.
A5. Si un individuo es padre de otro, éste no lo es de aquél.
A6. Si un individuo es madre de otro, éste no lo es de aquél.
Definiciones:
D1. Progenitor: x es progenitor de y syss_{def} x es padre o madre de y.
El resto de definiciones como en TP1.

Pues bien, es fácil ver que TP6 es equivalente a TP1 (y como TP1, por tanto, reducida por TP3 y TP5) y sin embargo ninguna incluye como primitivos todos los conceptos primitivos de la otra. Lo que sí ocurre, y este es el punto esencial, es que los términos primitivos de una pueden ser *definidos* mediante los de la otra (en los términos comunes la definición es inmediata, la gracia está en los no comunes) de modo que los axiomas de una se convierten en afirmaciones (axiomas o teoremas) de la otra. Éste es el concepto un poco más refinado de reducción: T reduce T' si los términos primitivos de T' pueden ser definidos en T de modo que los axiomas de T' se obtienen como axiomas o teoremas de T (sólo como teoremas, si no tienen ningún término en común); T y T' son equivalentes si se reducen mutuamente. Cuando las teorías tienen términos comunes, como en nuestras teorías del parentesco, estas relaciones no suelen ser extremadamente interesantes, después de todo "parecen hablar (al menos parcialmente) de lo mismo". Más interesantes son los casos, que discutiremos por extenso más adelante (cf. en el próximo apartado de esta sección el ejemplo de la teoría de conjuntos y la aritmética, y para ejemplos empíricos el capítulo 11), en los que las teorías involucradas no comparten ni siquiera parcialmente el material conceptual, esto es, cuando las teorías parecen en principio estar hablando de cosas diferentes y se descubre que una teoría reduce otra. Este tipo de situaciones tienen el máximo interés desde el punto de vista metacientífico, como veremos, por su relevancia en los fenómenos de cambio teórico y sus implicaciones epistemológicas y ontológicas (cap. 13).

1.3. ARITMÉTICA, TEORÍA DE CONJUNTOS Y LÓGICA PROPOSICIONAL

Acabaremos dando algunos ejemplos de teorías axiomáticas pertenecientes al campo de las ciencias formales, más interesantes a efectos ilustrativos que nuestra inventada teoría del parentesco pues corresponden a casos reales de teorías axiomatizadas.

Aritmética de Peano (AP):

AP, axiomatizada por Peano (y Dedekind) a finales del siglo xtx, pretende sistematizar axiomáticamente las verdades conocidas y utilizadas informalmente desde antiguo sobre los números naturales y sus propiedades, relaciones y operaciones básicas.

Términos primitivos.
C 1. Número natural (relator monádico)
C2. Cero (término singular)
C3. El siguiente de (functor monádico)
Axiomas:
A1. Si un objeto es número natural, su siguiente también lo es.
A2. El cero es un número natural.
A3. El cero no es el siguiente de ningún número natural.
A4. Dos objetos con el mismo siguiente son el mismo.
A5. Si el cero tiene una propiedad (p y el que un número natural sea cp implica que su siguiente también es (p, entonces todo número natural tiene la propiedad (p.

(Nótese que A5 no es exactamente un único axioma sino lo que se denomina un *esquema axiomático,* un axioma con una (meta)variable libre (p, en este caso una variable para propiedades, que da lugar a axiomas específicos para cada ejemplificación concreta de la variable.)

Teoremas:
TI. El siguiente del siguiente del cero es un número natural.
T2. El siguiente del siguiente del cero no es el siguiente del cero.
T3. Cero no es el siguiente del siguiente de cero.
Definiciones:
DI. Uno =$_{def}$ el siguiente de cero.
D2. Dos =$_{def}$ *el* siguiente de uno.
D3. Tres =$_{def}$
D4. Suma (functor binario: la suma de ... y ...):
a) *x* más cero =$_{def}$ *x*
b) *x* más el siguiente de *y* =$_{def}$ *el* siguiente de (x más y).
D5. Producto (functor binario):
a) *x* por cero =$_{def}$ *cero*
b) *x* por el siguiente de *y* =$_{def}$ *x* más (x por y).
D6. *x* ≤ *y* syss$_{def}$ existe un número natural z tal que *x* más z = y.
D7. x es par syss$_{def}$ existe un natural z tal que x = dos por z.

(D4 y D5 son lo que se denomina *definiciones recursivas;* intuitivamente: el resultado de la operación de un número con otro diferente de cero se da en función del resultado con el anterior del segundo -cláusula b)-; como además se da el resultado de operar cualquier número con el cero -cláusula a)-, queda bien definida la operación para cualesquiera números, pues todos los números surgen del cero mediante la función siguiente.)

Teoremas:
T4. **Dos** es número natural y par.
T5. Dos no es el siguiente del cero.
T6. Cero no es el siguiente de uno.
T7. Uno más dos = tres.
T8. Para todo x, y: *x* más y = y más *x*.
T9. Para todo *x:* x por uno = *x*.
T10. Para todo x, y: x S x más y.

Teoría de Conjuntos (TC):

TC es una teoría desarrollada en su práctica totalidad por el matemático alemán G. Cantor a finales del siglo xix. TC trata de los "agregados", conjuntos, colecciones o clases, de las propiedades, relaciones y operaciones entre estas entidades. TC se axiomatizó a principios del siglo xx como parte de algunas estrategias para resolver los problemas de fundamentos derivados de la inconsistencia de la teoría en su versión intuitiva inicial. Hay varias axiomatizaciones alternativas, y la que damos aquí es parcial pues recoge sólo algunos de los axiomas más comunes; es por tanto insuficiente y no contiene los elementos que hacen propiamente interesantes las diversas axiomatizaciones existentes.

Términos primitivos:
C1. Conjunto (relator monádico).
C2. Pertenencia (relator diádico: ... es elemento de ...).
Axiomas:
Al. **Dos** conjuntos a los que pertenecen los mismos objetos son el mismo.
A2. Dado un conjunto y una propiedad cp, hay un conjunto cuyos elementos son los elementos del primero que tienen la propiedad (p.
A3. Existe algún conjunto que no tiene elementos.
A4. Dados dos conjuntos, existe otro formado por los elementos de los anteriores.
A5. Dados dos objetos, existe un conjunto formado por ambos.

(El lector habrá notado que A2 es un esquema axiomático.)

Teoremas:
T1. Existe un y sólo un conjunto sin elementos.
T2. Dados dos conjuntos, existe un y sólo un conjunto cuyos miembros son los elementos de los anteriores.

T3. Dados dos conjuntos, existe otro formado por los elementos comunes de ambos, y es único.

T4. Dados dos conjuntos, existe otro formado por los elementos del primero que no pertenecen al segundo, y es único.

Definiciones:

DI. $\emptyset =_{def}$ el conjunto sin elementos.

D2. $x \subseteq y$ syss$_{def}$ los elementos de x son también elementos de y.

D3. $x \cup y =_{def}$ el conjunto formado por los elementos de x o de y.

D4. $x \cap y =_{def}$ el conjunto formado por los elementos comunes de x e y.

D5. $x - y =_{def}$ el conjunto formado por los elementos de x que no pertenecen a y.

Teoremas:

T5. Para todo x: $\emptyset \subseteq x$.

T6. Para todo x: $x \cup \emptyset = x$.

T7. Para todo x, y: $x \cap y = y \cap x$.

T8. Para todo x, y, z: $x - (y \cup z) = (x - y) \cap (x - z)$.

T9. Para todo x, y: si $x \subseteq y$ entonces $x \cap y = x$.

AP y TC proporcionan una ilustración interesante de uno de los conceptos que introdujimos más arriba con ocasión de las diversas teorías del parentesco, la relación de reducción. Vimos entonces que TP5 reducía TP1, hecho que no parecía muy interesante dada la inmediata proximidad temática de ambas teorías. Pues bien, uno de los logros más importantes de la historia de las ciencias formales (debido fundamentalmente a Frege) consiste en haber mostrado que AP se reduce a TC (no a esta TC, sino a la teoría de conjuntos en su versión completa). Éste es un hecho en principio sorprendente pues ambas teorías parecen hablar "de cosas diferentes". Pues bien, Frege mostró que hay una manera de definir los números como determinados conjuntos *(o extensiones*, como él decía) de modo que las verdades básicas sobre números se derivan de las verdades básicas sobre conjuntos: es posible definir los términos aritméticos primitivos mediante términos conjuntistas de modo tal que los axiomas de la aritmética se convierten en teoremas que se derivan de los axiomas de la teoría de conjuntos. La reducción se da pues exactamente en el mismo sentido que más arriba vimos respecto de TP1 y TP5, pero ahora es realmente interesante pues las teorías involucradas no comparten, en principio, aparato conceptual.

A modo de ejemplificación de la idea abstracta de reducción, vamos a presentar tan sólo las líneas generales de la reducción de AP a TC (ahora nos expresaremos incorrectamente y mezclaremos definición de términos con identificación de entidades). El *cero* se define identificándolo con el conjunto vacío. Para la función *siguiente* hay varias posibilidades. Una es definir el siguiente de un conjunto x como su unitario $\{x\}$ (cuya existencia unívoca queda garantizada: si existe x, existe por A5 {x, x} que es idéntico a {x} por Al). Otra posibilidad es definir el siguiente de un conjunto x como su unión con su unitario, e.e. como $x \cup \{x\}$ (cuya existencia unívoca también queda garantizada, por la existencia de {x} más A4). La definición de `número natural' requiere algunas complicaciones que no hemos incluido en la versión simplificada de TC que hemos presentado. En especial se requiere el siguiente axioma que habíamos omitido: "existe al menos un conjunto tal que *a)* tiene como elemento 0 y b) si tiene como elemento un conjunto, entonces tiene también el siguiente de dicho conjunto". Este axioma asegura que existe al menos un conjunto con el vacío y con todos los que le siguen (que son infinitos, por eso se le denomina a veces Axioma de Infinitud). A los conjuntos así se les denomina *inductivos*, y es posible que haya varios tales conjuntos "inductivos", si además de tener esos objetos tienen otros. Pues bien, el menor de todos esos conjuntos inductivos, e.e. la intersección de todos ellos, tiene como elementos sólo el vacío y sus siguientes, es decir, los números naturales "definidos" en términos conjuntistas. La propiedad de ser un número natural se define entonces como "pertenecer al menor conjunto inductivo", definición con la que concluye la reducción de la aritmética a la teoría de conjuntos, pues ahora los cinco axiomas de Peano son teoremas de TC (con algunas complicaciones que hemos obviado, sobre todo relacionadas con A5 de AP, pero que no son esenciales para la idea general).

Lógica Proposicional (LO):

Concluimos con la presentación del propio cálculo lógico de la lógica proposicional (lo mismo se podría hacer con la lógica de primer orden). Aunque hoy es usual presentar los cálculos lógicos como cálculos de deducción natural, históricamente se formularon originariamente como cálculos axiomáticos; eso fue (parte de) lo que hicieron, p.ej., Frege en su *Begriffsschrift y* Russell y Whitehead en *Principia Mathematica* para la Lógica de Primer Orden. En lo que sigue referimos sólo los axiomas de una de las diversas versiones posibles para la lógica de enunciados (no referimos las reglas de inferencia, Modus Ponens y Sustitución; las variables están por proposiciones, y los axiomas y teoremas se han de leer, como en las teorías anteriores, clausurados universalmente).

Términos primitivos:
C1. Negador: – (functor monario).
C2. Implicador: -> (functor binario).
Axiomas:
A1. x-->(y->x).
A2. (– x -> – y) --> (y x).
A3. *(x -> (y -> z)) ((x ~ y) --> (x -> z)).*
Teoremas:
T*I*. *x-ax.*
T2. – (x -> – x).
T3. x -a (-i x --> y).
Definiciones:
D1. *xvy=,,ef-,x->y.*
D2. *xAy=af-(x --> -y).*
D3. *x<->y=<1,f(x->y)A(y->x).*
Teoremas:
T4. *– (x A – x).*
T5. xv–x.

T6. *(xA – x)* *y.*
T7. (x- > (y — y)) -~ x.
T8. - (xAy)H(,xv,y).
T9. (x A (y v z)) H ((x A y) v (x A z)).

Con LO como última ilustración concluimos la presentación del concepto de cálculo o teoría axiomática y de otras nociones relacionadas, como las de reducción y equivalencia. Para fijar estos conceptos hemos utilizado intencionadamente como ejemplos (además de nuestras inventadas teorías del parentesco) teorías pertenecientes al campo de las ciencias formales. El motivo es que el análisis de las teorías de las ciencias empíricas como cálculos axiomáticos presenta problemas específicos que no tienen que ver con el concepto abstracto de teoría axiomática. Una vez fijado el instrumental conceptual, en las secciones subsiguientes nos ocuparemos por extenso de su eventual aplicación a las ciencias empíricas y los problemas específicos que tal aplicación comporta. Finalizaremos esta introducción del aparato conceptual presentando el concepto formal de *modelo o* realización de una teoría axiomática.

2. **Teorías** y **modelos**

En el lenguaje común el término `modelo' es un término extremadamente polisémico, y dentro mismo de la filosofía de la ciencia se usa con toda una variedad de significados diferentes (para un análisis de los mismos, cf. Falguera, 1993). Una familia de tales significados tiene que ver con la idea de *caso o realización* de una afirmación o conjunto de ellas. Así, por ejemplo, podemos decir que Romeo y Julieta, o mejor ellos "junto con su amor", son un modelo, caso o realización de la afirmación "los amantes prefieren la muerte a la separación"; o que España y Bélgica, "con todo lo que llevan dentro", son en la actualidad modelos o casos de monarquía constitucional, esto es, de una serie de principios o reglas políticas, y que Francia, Italia y Portugal lo son de estados republicanos (aunque España y Bélgica no realizan exactamente los mismos principios monárquicos, sólo comparten parte de ellos, y lo mismo Francia, Italia y Portugal respecto de los principios republicanos). Parte del sentido de este uso es explicitado y precisado por una teoría lógico-matemática altamente abstracta, la Teoría de Modelos. No vamos a ver aquí siquiera los rudimentos de tal teoría, nos limitaremos a presentar informalmente el concepto de modelo del que ella se ocupa, pues desempeña un papel importante en algunos análisis metateóricos que veremos en éste y próximos capítulos.

Un modelo en el sentido de la Teoría de Modelos (en adelante escribiremos simplemente `modelo') es un sistema o estructura, un "trozo de la realidad" constituido por entidades de diverso tipo, que *realiza* una teoría o conjunto de axiomas en el sentido de que en dicho sistema "pasa lo que la teoría dice" o, más precisamente, la teoría es verdadera en dicho sistema. Si tomamos los principios monárquicos generales comunes a las constituciones española y belga, y los bautizamos como Teoría Mínima de la Monarquía Constitucional, entonces España y Bélgica, y p.ej. también Suecia, como sistemas o "partes de la realidad", son modelos de dicha teoría, y Francia, Italia y Portugal no lo son. Esta idea intuitiva se puede hacer precisa mediante la noción formal de sistema o estructura que presentamos en el Apéndice. Recordemos que un sistema es simplemente una tupla o secuencia de entidades conjuntistas construidas a partir de un universo o dominio básico de objetos. A veces puede haber varios dominios básicos, pero eso no es una diferencia esencial, pues siempre se puede tomar su unión como el universo y después destacar de él los subconjuntos principales. Lo importante es que un sistema es o representa "un pedazo de la realidad", no es por tanto una entidad lingüística, salvo quizá, en algunas ocasiones, en un sentido derivado cuando los objetos del universo son ellos mismos entidades lingüísticas. El dominio básico puede constar de personas, números, proposiciones, partículas o cualesquiera otras entidades, por ejemplo enunciados. En el primer caso tenemos un sistema "humano", en el segundo otro "numérico", y en el último caso tenemos un sistema "lingüístico", constituido por entidades lingüísticas. Pero incluso si en este caso queremos decir que el sistema es una entidad lingüística en el sentido de estar construido por entidades lingüísticas, ello sólo es un modo de hablar, pues estos sistemas "lingüísticos" no son ellos mismos entidades lingüísticas. Por tanto los sistemas son simplemente *partes estructuradas* de la realidad, y como tales no son entidades lingüísticas susceptibles de ser verdaderas o falsas o de tener significado. Son más bien la realidad respecto de la cual ciertas entidades lingüísticas, enunciados o conjuntos de ellos, las teorías entendidas en el sentido axiomático visto, son verdaderas o falsas. O si se quiere, contemplada la relación en la dirección opuesta, son partes de la realidad que se comportan o no como afirma la teoría, que satisfacen o no las afirmaciones de la teoría. Puesto que en una teoría axiomática todas sus afirmaciones, su contenido, está expresado plenamente, aunque implícitamente, por los axiomas, para ver si un sistema es o no modelo de una teoría, basta ver si satisface o no sus axiomas.

Para que un sistema pueda siquiera ser modelo de una teoría es necesario que tenga el tipo lógico apropiado, es decir, que esté constituido por entidades del mismo tipo lógico que los términos primitivos de la teoría, pues las entidades del sistema son "el significado en el sistema", esto es la *interpretación*, de los términos de la teoría. Si la teoría contiene relatores diádicos y en el sistema no hay relaciones binarias es obvio que ni siquiera podemos ponernos a ver si la teoría es verdadera o falsa en dicho sistema. Sea una teoría T cuyos términos primitivos son ***j*** relatores *R,, ..., R;* (cada uno con su ariedad especificada), *k* functores *f,, ...,* f_k (con sus ariedades especificadas) y *m* términos singulares o constantes individuales c,, ..., c,,,. Diremos entonces que un sistema S es una *realización posible* de *T* si tiene el tipo lógico apropiado. Vamos a utilizar las mismas letras en negrita para denotar las entidades que interpretan en el sistema los términos de la teoría. S es una realización posible de T si S consta de un universo U y, construidas sobre U, ***j*** relaciones **R**, ..., R;,, *k* funciones f,, ..., f_k *y m* individuos destacados c,, ..., c,,,, **S** = < U, R,, ..., **R;, f,, ...,** f_k, **c,, ..., c,,,**>, tales que cada relación y función es de la misma ariedad que el relator o functor que interpreta. Estos sistemas son las entidades de las que tiene sentido preguntarse si son o no modelos de la teoría, si en ellas la teoría es verdadera o falsa; por ello se denominan `posibles realizaciones'. Por supuesto que (salvo que sea una teoría tautológica) no todas las posibles realizaciones serán realizaciones efectivas o modelos de la teoría. Las realizaciones efectivas, *o modelos de la teoría,* son aquellas realizaciones posibles en las que ocurre lo que la teoría afirma, las que de hecho se comportan como la teoría dice, o técnicamente, en las que los axiomas (y con ellos todo el resto de afirmaciones) de la teoría son verdaderos.

Para fijar ideas concluiremos dando una serie de axiomas y diversos sistemas, algunos de los cuales satisfacen todos los axiomas y otros sólo parte de ellos.

Términos primitivos:

C1. M (relator diádico).
C2. * (functor diádico).
C3. $\circ$ (functor diádico).
C4. ~ (functor monádico).
C5. e (constante individual).

Axiomas (léanse clausurados universalmente):

A1. $x\mathrm{M}x$
A2. $x\mathrm{M}y \wedge y\mathrm{M}z \rightarrow x\mathrm{M}z$
A3. $x*y = y*x$
A4. $x\circ y = y\circ x$
A5. $(x*y)*z = x*(y*z)$
A6. $x\circ x = x$
A7. $\sim\sim x = x$
A8. $x\mathrm{M}x*y$
A9. $x\circ y\mathrm{M}x$
A10. $x\circ(y*z) = (x\circ y)*(x\circ z)$
A11. $\sim(x\circ y) = \sim x * \sim y$
A12. $x*e = x$
A13. $x\circ e = e$
A14. $x * \sim x = e$

Contemplemos ahora los siguientes sistemas, en los que el primer constituyente, después del universo, interpreta `M', el segundo '*', el tercero `o', el cuarto '-' y el quinto `e'.

S, = <N, S, +, •, s, 0> (los naturales con "menor o igual que", la suma, el producto, la función "siguiente de" y el cero).
S_2 = <Z, <-, +, •, –, 0> (los enteros con "menor o igual que", la suma, el producto, el opuesto y el cero).
S_3 = <I, <-, +, •, –, 0> (los enteros impares con "menor o igual que", la suma, el producto, el opuesto y el cero).
S_4 = <Q, <, •/ :,-', 1> (los racionales con "menor o igual que", el producto, el cociente, el inverso y el uno).
S_5 = <C, c, u, n, -, O> (los conjuntos con la inclusión, la unión, la intersección, el complemento y el conjunto vacío).
S_6 = <P, i--, v, A, ¯¯i, c> (las proposiciones con la relación de consecuencia lógica, la disyunción, la conyunción, la negación y la (o una) contradicción).

El lector puede comprobar que todos ellos son realizaciones posibles, tienen el tipo lógico apropiado para *poder* satisfacer los axiomas, pero que sólo S_5 y S_6 son modelos de todos los axiomas, satisfacen de hecho todos los axiomas a la vez. En S, fallan los axiomas 6, 7, 9, 11 y 14; S_2 y S_3 satisfacen los mismos, a saber, todos menos 6, 8, 9 y 11; por último, S_4 sólo satisface 1, 2, 3, 5, 7, 12 y 14.

Como se ve, una misma teoría, un mismo conjunto de axiomas, puede tener modelos muy diferentes. De hecho no hay ninguna teoría que tenga un único modelo o realización, al menos si estamos dispuestos a aceptar siempre modelos matemáticos. Ahora bien, aunque las teorías no determinan unívocamente sus modelos en este sentido tan estricto, lo pueden hacer en otro sentido todavía interesante, un poco más débil que el anterior y de hecho más razonable. En la interpretación que venimos usando, una teoría pretende "describir (un trozo de) la realidad". Pues bien, si los diversos modelos son "extremadamente semejantes" entre sí, aunque no se describa un único modelo se tratará de una buena descripción en el sentido de *suficientemente unívoca*. Dicho técnicamente: si los diversos modelos son isomorfos entre sí (para la noción de isomorfía, cf. Apéndice), la teoría determina la realidad del modo más fuerte que es razonable exigir; a una teoría así se le denomina *categórica*. Dar con una teoría categórica no es fácil. Por ejemplo, la teoría consistente en todos los axiomas que acabamos de presentar tiene modelos no isomorfos, S_5 y S_6. La teoría consistente en los axiomas 1, 2, 3, 4, 5, 7, 10, 12, 13 y 14, tiene modelos isomorfos, S_2 y S_3, pero otros no isomorfos, pues S_5 y S_6 son también modelos de esa teoría y no son isomorfos entre sí ni respecto de S_2 y S_3. En realidad no siempre es bueno pretender que la teoría sea categórica. Quizá eso sea deseable en las teorías formales, pero desde luego no lo es en las teorías empíricas. En las teorías empíricas es natural pretender que una teoría tenga modelos que sean partes de otros modelos o, como se dice técnicamente, que entre sus modelos algunos puedan ser extensiones de otros, esto es, partes más grandes de la realidad, por así decir. Eso no siempre impide la isomorfía, por ejemplo S_2 es una extensión de S_3 y son isomorfos; pero la impide cuando los modelos, como parece razonable no descartar en las teorías empíricas, tienen un universo finito. Sobre este tipo de cuestiones trataremos más adelante. De momento es conveniente insistir ahora tan sólo en el siguiente hecho, trivial pero interesante: todos los modelos de una teoría, sean o no isomorfos, se parecen mucho en cierto sentido, a saber, todos se comportan como la teoría dice. Y eso por supuesto es un parecido digno de tener en cuenta, de hecho es *el* tipo de parecido a tener en cuenta cuando se trata de teorías empíricas.

3. Caracterización general de las teorías empíricas como cálculos interpretados

Examinemos ahora los primeros análisis que se hicieron del concepto de *teoría empírica*. Según la concepción a que dichos análisis dieron lugar, una teoría empírica es un *cálculo interpretado,* donde por `cálculo' se entiende un cálculo o teoría axiomática en el sentido presentado en la sec. 1. Allí vimos unos ejemplos puramente formales, y la ilustración era intencionada, pues de hecho los primeros filósofos de la ciencia tomaron la idea de las axiomatizaciones que entonces se hacían de algunas teorías lógicas y matemáticas

El lector puede comprobar que todos ellos son realizaciones posibles, tienen el tipo lógico apropiado para *poder* satisfacer los axiomas, pero que sólo S_5 y S_6 son modelos de todos los axiomas, satisfacen de hecho todos los axiomas a la vez. En S, fallan los axiomas 6, 7, 9, 11 y 14; S_2 y S_3 satisfacen los mismos, a saber, todos menos 6, 8, 9 y 11; por último, S_4 sólo satisface 1, 2, 3, 5, 7, 12 y 14.

Como se ve, una misma teoría, un mismo conjunto de axiomas, puede tener modelos muy diferentes. De hecho no hay ninguna teoría que tenga un único modelo o realización, al menos si estamos dispuestos a aceptar siempre modelos matemáticos. Ahora bien, aunque las teorías no determinan unívocamente sus modelos en este sentido tan estricto, lo pueden hacer en otro sentido todavía interesante, un poco más débil que el anterior y de hecho más razonable. En la interpretación que venimos usando, una teoría pretende "describir (un trozo de) la realidad". Pues bien, si los diversos modelos son "extremadamente semejantes" entre sí, aunque no se describa un único modelo se tratará de una buena descripción en el sentido de *suficientemente unívoca*. Dicho técnicamente: si los diversos modelos son isomorfos entre sí (para la noción de isomorfía, cf. Apéndice), la teoría determina la realidad del modo más fuerte que es razonable exigir; a una teoría así se le denomina *categórica*. Dar con una teoría categórica no es fácil. Por ejemplo, la teoría consistente en todos los axiomas que acabamos de presentar tiene modelos no isomorfos, S_5 y S_6. La teoría consistente en los axiomas 1, 2, 3, 4, 5, 7, 10, 12, 13 y 14, tiene modelos isomorfos, S_2 y S_3, pero otros no isomorfos, pues S_5 y S_6 son también modelos de esa teoría y no son isomorfos entre sí ni respecto de S_2 y S,. En realidad no siempre es bueno pretender que la teoría sea categórica. Quizá eso sea deseable en las teorías formales, pero desde luego no lo es en las teorías empíricas. En las teorías empíricas es natural pretender que una teoría tenga modelos que sean partes de otros modelos o, como se dice técnicamente, que entre sus modelos algunos puedan ser extensiones de otros, esto es, partes más grandes de la realidad, por así decir. Eso no siempre impide la isomorfía, por ejemplo S_2 es una extensión de S_3 y son isomorfos; pero la impide cuando los modelos, como parece razonable no descartar en las teorías empíricas, tienen un universo finito. Sobre este tipo de cuestiones trataremos más adelante. De momento es conveniente insistir ahora tan sólo en el siguiente hecho, trivial pero interesante: todos los modelos de una teoría, sean o no isomorfos, se parecen mucho en cierto sentido, a saber, todos se comportan como la teoría dice. Y eso por supuesto es un parecido digno de tener en cuenta, de hecho es *el* tipo de parecido a tener en cuenta cuando se trata de teorías empíricas.

3. Caracterización general de las teorías empíricas como cálculos interpretados

Examinemos ahora los primeros análisis que se hicieron del concepto de *teoría empírica*. Según la concepción a que dichos análisis dieron lugar, una teoría empírica es un *cálculo interpretado*, donde por `cálculo' se entiende un cálculo o teoría axiomática en el sentido presentado en la sec. 1. Allí vimos unos ejemplos puramente formales, y la ilustración era intencionada, pues de hecho los primeros filósofos de la ciencia tomaron la idea de las axiomatizaciones que entonces se hacían de algunas teorías lógicas y matemáticas (axiomatizaciones en matemáticas que hasta mediados de siglo seguirían invariablemente el esquema de la sección 1). Es más, el primer análisis detallado de las teorías empíricas como cálculos interpretados se presenta explícitamente en relación con una teoría axiomática puramente matemática. Se trata del estudio que hace Reichenbach en los años veinte de las semejanzas y diferencias en naturaleza y estructura de la Geometría Pura (GP) y la Geometría Física (GF) (cf. 1928). Las semejanzas consistían básicamente en la estructura axiomática de ambas; las diferencias se derivaban de la naturaleza empírica de la segunda. A diferencia de GP, cuyo análisis se agota al dar su estructura axiomática puramente formal, el carácter empírico de GF obliga a completar la parte puramente axiomático-formal con elementos adicionales que den cuenta de su carácter "físico'; estos elementos deben hacer explícitos los modos en que el formalismo abstracto se pone en contacto con la experiencia, esto es, el modo en que recibe una *interpretación física* determinada.

Éste es el origen y núcleo del análisis de las teorías empíricas como cálculos interpretados. La idea básica es desarrollada, en los años veinte y treinta del siglo xx, de modo parcialmente coincidente por Reichenbach, Ramsey, Bridgman, Campbell y Carnap, que sería su principal impulsor. Esta idea se conforma como el núcleo central de lo que se denominará más tarde *Concepción Heredada* y encuentra su expresión más elaborada en las principales monografías que en los años cincuenta y sesenta escriben sus principales representantes (cf. especialmente Braithwaite, 1959; Carnap, 1966; Nagel, 1961 y Hempel; 1965 y 1966) y será prácticamente dominante en filosofía de la ciencia hasta casi los años setenta. Veremos primero en esta sección cuáles son sus aspectos más generales y discutiremos en las siguientes con algo de detalle algunos de sus elementos, la evolución que sufrieron y las últimas revisiones críticas, principalmente por parte de Hempel.

3.1. TEORÍAS FORMALES Y TEORÍAS EMPÍRICAS

Según la posición dominante en filosofía de las ciencias formales, al menos durante la primera mitad del siglo xx, los axiomas del formalismo abstracto son lo único que interviene en la caracterización de las entidades "de las que habla" una teoría matemática; qué cosas son esas de las que pretendemos hablar al usar los términos de la teoría es algo que depende únicamente de los axiomas, las entidades en cuestión son cualesquiera de las que los axiomas sean verdaderos. Así, por ejemplo, los números naturales, es decir, la propiedad de ser número natural, el número cero y la función siguiente, serán cualesquiera entidades de las que resulten ser verdaderos los axiomas de Peano. A veces se expresa esto diciendo que los axiomas caracterizan las entidades de la teoría o, también, que *definen implícitamente los* términos primitivos. Debe quedar claro que no se trata de una definición en el sentido introducido en la sección 1. De los términos primitivos no puede haber *definición explícita,* pues entonces no serían primitivos (las definiciones explícitas son justamente el modo en que se introducen los términos derivados a partir de los primitivos). Los axiomas "definen" implícitamente los términos primitivos en el sentido apuntado, a saber, ellos son los únicos elementos constitutivos del significado de los términos; cualquier estructura que sea modelo de los axiomas es una interpretación admisible de los mismos; esto es, los constituyentes de cualquiera de tales estructuras son interpretaciones admisibles de los términos con que se formulan los axiomas.

Algunos autores, como Carnap, han dado una interpretación *convencionalista* de las definiciones implícitas, según la cual los axiomas *estipulan* el significado de los términos y por tanto son "verdades por convención". Otros, como Hilbert inicialmente, han defendido una interpretación *formalista,* de acuerdo con la cual los axiomas son simplemente reglas para el manejo de los signos involucrados, y por tanto en sentido estricto ni siquiera es razonable considerarlos susceptibles de ser verdaderos o falsos. En tal caso es discutible que tenga incluso sentido hablar propiamente del significado o interpretación de los términos involucrados. Convencionalismo y formalismo son dos modos específicos de entender la idea general de que en un sistema axiomático los axiomas definen implícitamente los términos, que los axiomas solos caracterizan *plenamente* el "uso correcto" de los términos. Pero se puede defender esta idea sin defender ninguna de esas interpretaciones específicas de la misma, defendiendo que existen "realmente" las entidades matemáticas significadas por los términos primitivos; los números naturales son cualquier cosa (tomando ahora `cosa' en serio en un sentido no lingüístico) que satisfaga los axiomas de Peano.

Independientemente de cuál sea su interpretación específica (salvo en el caso quizá del formalismo radical) hay algo de plausible en la idea general de que las cosas de las que una teoría matemática habla son cualesquiera entidades que satisfagan los axiomas. Ese elemento de plausibilidad es el que justifica que aceptemos que Frege redujera la aritmética a teoría de conjuntos. Como hemos visto en la sección 1, Frege demostró que los axiomas de Peano que usan los términos `cero', `el siguiente de' y `número natural' son verdaderos de "el conjunto vacío", "la unión con el propio unitario" y "pertenecer al menor conjunto inductivo". Entonces *eso* son (los) naturales, pues de esas cosas son verdaderos los axiomas de la aritmética. Quien proteste y diga que a pesar de ello no se trata de la aritmética, pues está claro que "esas cosas" son conjuntos y no números, está rechazando de plano dicha idea general; y en la medida en que tal protesta se considere injustificada se considerará plausible la idea en cuestión. Y efectivamente la protesta parece algo injustificada, lo que son las cosas depende de sus propiedades, de su "comportamiento", que es lo que establecen los axiomas, no depende de las expresiones lingüísticas con las que la teoría las nombre. De todas formas las intuiciones no son del todo claras pues, ¿qué diríamos si de hecho AP fuese verdadera también de (supongamos por un momento que efectivamente existen tales entidades) el individuo Lucifer, la función "el corruptor de" y la propiedad "ser demonio"?, ¿aceptaríamos también sin protestar que estamos ante números naturales?

Esta cuestión es extremadamente compleja y no vamos a ocuparnos aquí de ella, pues eso es tarea de la filosofía de la matemática. Si nos hemos extendido algo sobre este punto es para presentar la intuición central del análisis de las teorías *empíricas.* La idea que queremos destacar es la siguiente: mientras que en las ciencias formales parece razonable, o al menos defendible, la tesis de que las entidades a las que la teoría se refiere son cualesquiera de las que sean verdaderas los axiomas, ella es totalmente inaceptable aplicada a las ciencias empíricas. Por ejemplo, si los principios de la mecánica newtoniana, formulados con términos como `partícula', `masa' y `fuerza', fuesen por casualidad verdaderos de los ángeles, su "cantidad de espiritualidad" y sus "afinidades", no por ello dinamos que ésas son cosas de las que habla la teoría mecánica, no diríamos que son sistemas mecánicos. O, con un ejemplo menos absurdo, tales principios son de hecho verdaderos de ciertos dominios puramente numéricos y de ciertas funciones puramente numéricas entre los números de los dominios, pero eso no hace que la mecánica hable (quizá entre otras cosas) de puros números, no por ello los números son partículas mecánicas. Eso es indiscutible en este caso; la idea de que los términos de la mecánica refieren a cualesquiera entidades que satisfagan el formalismo abstracto es claramente inaceptable. El motivo es que, a diferencia de las ciencias formales donde esa idea es cuando menos discutible, las teorías empíricas tienen, además de las constricciones derivadas del sistema axiomático abstracto, otras constricciones derivadas de su vinculación con *el mundo físico-natural,* o mejor dicho, con algún aspecto específico del mismo del que pretenden dar cuenta. Por supuesto que esta diferencia depende de que se considere que hay a su vez una diferencia entre teorías formales y teorías empíricas, y efectivamente la mayoría de los autores de este período (y también posteriormente) consideran que dicha diferencia es un dato firme de nuestras intuiciones.

Aceptando esta peculiaridad de las teorías empíricas, ¿cómo se debe recoger este hecho específico en el análisis de las mismas? La respuesta parece inmediata: incluyendo, junto con el sistema axiomático abstracto, otro elemento que exprese la conexión de dicho formalismo con "situaciones de la experiencia" en las que interactuamos o "contactamos" con el mundo físico. La articulación específica de esta respuesta que se impondrá en la Concepción Heredada es que esas situaciones de experiencia en las que se da el contacto básico con el mundo físico son situaciones de *observación directa* de fenómenos físicos. En la formulación y expansión de esta idea influyó sin duda el neoempirismo (a veces radical) dominante en los círculos filosóficos donde se plantearon estas cuestiones, pero independientemente de ello la idea tiene cierta plausibilidad en sí misma. La interacción básica con nuestro entorno físico inmediato la realizamos a través de los órganos sensoriales, interacción que se plasma en forma de *observación,* en un sentido amplio, no estrictamente visual, del término. Pero hay que distinguir esta idea general difícilmente recusable de otras dos más fuertes: *a)* El "estar basado en la observación" característico de nuestro sistema cognitivo global, caracteriza también sus diversas partes, y en particular es característico en exactamente el mismo sentido de cada teoría científica concreta. *b)* Por observación directa se ha de entender observación independiente de esquemas cognitivos elaborados. Aunque hemos formulado estas tesis muy vagamente, debe notarse que no son ya indiscutibles, que son mucho más fuertes que la idea general claramente defendible. En la penúltima sección volveremos sobre ellas para discutirlas en detalle; por el momento basta saber que eran defendidas por muchos de los iniciadores de la Concepción Heredada y que se acabarían imponiendo en la versión oficial de la misma (aunque, como veremos, la particular versión que cada autor ofrece de ellas incluye importantes matizaciones y cualificaciones para evitar algunas de las consecuencias más patentemente inaceptables).

3.2. CÁLCULOS INTERPRETADOS: VOCABULARIO; AXIOMAS Y REGLAS DE CORRESPONDENCIA

Según el análisis que estamos presentando, por tanto, cada teoría científica está conformada por un cálculo axiomático abstracto y otro componente que conecta las expresiones de dicho cálculo abstracto con situaciones de la experiencia entendidas como situaciones de observación directa. Este segundo elemento está conformado por enunciados que vinculan los términos del sistema axiomático con términos *observacionales* que refieren a objetos, propiedades o relaciones directamente observables. A esos "enunciados conectores" se les ha denominado de varios modos: *reglas de correspondencia* (Carnap, Nagel, Margenau), *definiciones coordinativas* (Reichenbach), *enunciados interpretativos* (Hempel), *postulados de significación* (Carnap, Hempel), *diccionario* (Campbell, Ramsey) o *definiciones operacionales* (Bridgman). Pero, a pesar de que hay algunas diferencias de concepción (especialmente notables en el último caso), su función es en términos generales la misma, proporcionar interpretación empírica al cálculo axiomático que por sí mismo está vacío de contenido empírico. Las teorías empíricas son pues cálculos axiomáticos *interpretados* empíricamente a través de esos enunciados que conectan los términos del formalismo con situaciones de observación directa. En este punto, y antes de esquematizar los elementos centrales de esta concepción, es necesario hacer una observación.

Hemos dicho que el cálculo axiomático por sí solo carece de contenido empírico y que, para dar cuenta de su interpretación empírica, el análisis añade junto a los axiomas del cálculo un conjunto de reglas de correspondencia. Pero podría pensarse, quizá, que las cosas no tienen por qué ser así. Aunque lo distintivo de las teorías empíricas sea que tienen contenido empírico, e incluso si tal contenido se adquiere a partir de ciertas situaciones de observación, no es necesario introducir las reglas de correspondencia, puede bastar con el cálculo axiomático si algunos de sus términos fuesen términos de observación, en cuyo caso algunos de los axiomas estarían ya cargados de contenido empírico. Bien, esta cuestión es en parte puramente nominal, derivada de cómo hemos presentado las cosas. Tomemos todas las afirmaciones (primitivas) de la teoría y seleccionemos entre ellas las que no contienen términos observacionales: diremos que estas afirmaciones constituyen el cálculo axiomático abstracto. Tomemos las afirmaciones que incluyen términos tanto observacionales como no observacionales, éstas son las reglas de correspondencia. Se trata, si se quiere ver así, de que el análisis "divide" el sistema de afirmaciones completo en dos partes, o en realidad en tres partes, pues hay que añadir las afirmaciones que sólo contienen términos observacionales, esto es, las afirmaciones puramente observacionales.

Esta estrategia de análisis sería insatisfactoria si toda afirmación contuviera términos observacionales, pues en tal caso lo que hemos venido llamando cálculo axiomático abstracto no existiría. Pero ello no es así, toda teoría mínimamente compleja y sistematizada, que no sea un mero informe de afirmaciones observacionales, contiene afirmaciones sin términos observacionales. Y no sólo eso, sino que la mayoría de sus afirmaciones son de ese tipo, al menos tal y como aparecen las teorías formuladas en los libros de texto avanzados. Esas afirmaciones, aisladas, están desconectadas de la experiencia, de modo que lo que nos presenta una formulación estándar de una teoría empírica compleja altamente elaborada se parece mucho a lo que hemos caracterizado como un cálculo axiomático abstracto no interpretado (o al menos así lo pensaban los autores mencionados). Eso representa un serio reto para los filósofos de orientación empirista que, como parte de un proyecto que arranca de la Ilustración, quieren descartar como carentes de sentido las afirmaciones (no puramente analíticas) desconectadas de la experiencia, típicamente las de la metafísica especulativa y las de pseudociencias como la astrología. Es un reto, pues las afirmaciones que aparecen en los textos avanzados de muchas teorías científicas parecen en principio de ese tipo. La solución radica en que, aparezcan o no en las exposiciones usuales de la teoría, forman parte esencial de la teoría otras afirmaciones que ponen en conexión expresiones que aparecen en las primeras con situaciones observacionales; y si no existen afirmaciones de ese tipo, no se trata de una teoría empírica. Ésa es la diferencia entre la ciencia empírica y la metafísica; la diferencia con las pseudociencias consistiría en que, o carecen (como la metafísica) de reglas de correspondencia, o si tienen tales conexiones con la experiencia, entonces son patentemente falsas.

Pasemos ahora a esquematizar sumariamente los principales elementos del análisis presentado. Las teorías empíricas dan cuenta de fenómenos empíricos postulando ciertas entidades o procesos gobernados por ciertas leyes; esas entidades postuladas no están directamente dadas en la observación, están "alejadas" de la experiencia observable, contrariamente a los fenómenos de los que pretenden dar cuenta, directamente accesibles a la observación. La teoría introduce nuevos términos para referirse a esas entidades y procesos no observables. Diremos de esas entidades que son *entidades teóricas* y de los términos introducidos para referirnos a ellas que son *términos teóricos.*

Vocabulario. Podemos dividir el conjunto de expresiones o vocabulario V de una teoría en tres partes (nos limitamos aquí a los términos primitivos, los términos derivados se introducen mediante definiciones explícitas del modo indicado en la sección 1):

(1) Términos puramente lógico-matemáticos. Éste es el vocabulario formal V_F de la teoría, esto es, el vocabulario de apoyo que proporciona el lenguaje o instrumental formal y que en algunos casos puede incluir partes muy elevadas de la matemática.

(2) Términos observacionales. Éste es el vocabulario observacional V_O de la teoría, esto es, el vocabulario que se refiere a entidades directamente observables y a propiedades y relaciones entre ellas directamente observables. Son términos observables, por ejemplo, `rojo', `caliente', `vara', `más largo que', `más voluminoso que', `más liviano que', etc.

(3) Términos teóricos. Éste es el vocabulario teórico V_T de la teoría, esto es, el vocabulario que se refiere a entidades, propiedades y relaciones no directamente observables postuladas para dar cuenta de los fenómenos. Son términos teóricos, por ejemplo, `electrón', `masa', `campo eléctrico', `gen', 'entropía', etc.

Si llamamos vocabulario descriptivo V_D al vocabulario no meramente formal de apoyo, tenemos $V = V_F \cup V_D$, $V_D = V_O \cup V_T$, $V_F \cap V_D = \emptyset$ y $V_T \cap V_O = \emptyset$.

Afirmaciones. La anterior partición del vocabulario de una teoría genera otra a nivel de los enunciados o afirmaciones. Toda afirmación de la teoría contiene vocabulario formal, pero no sólo vocabulario formal, también contiene términos descriptivos. Eso nos deja las siguientes tres posibilidades:

(1) Enunciados (puramente) teóricos. Contienen como vocabulario descriptivo únicamente términos teóricos. De entre ellos se seleccionan algunos como axiomas o postulados primitivos: A,, ..., A,; el resto se deriva de ellos como teoremas. Son los enunciados que expresan el comportamiento de las entidades teóricas. Son enunciados teóricos, por ejemplo, `la fuerza eléctrica es directamente proporcional al producto de las cargas', `los genes tienen dos pares de alelos', etc.

(2) Enunciados (puramente) observacionales. Contienen como vocabulario descriptivo únicamente términos observacionales. Algunos describen situaciones observables particulares y otros son afirmaciones generales, esto es, expresan generalizaciones o leyes puramente empíricas-observacionales. Son enunciados observacionales, por ejemplo, `Juan Pérez tiene los ojos verdes', `esta porción de agua se ha solidificado', `los líquidos se solidifican al enfriarse', etc.

(3) Reglas de correspondencia. Contienen tanto términos teóricos como términos observacionales. En la medida en que unas se puedan derivar de otras, también se pueden escoger de entre ellas unas que hagan de primitivas: *R*,, ..., *R*,,,. Son los enunciados que conectan los términos teóricos con la experiencia observable cargando así de interpretación empírica los axiomas puramente teóricos. Son ejemplos de reglas de correspondencia, por ejemplo, `a presión constante, el volumen aumenta con la temperatura', `diferencias en el color de los ojos van acompañadas de diferencias en los genes', `al solidificarse un líquido disminuye su entropía', etc. Estos enunciados son el puente que permite pasar de lo observacional a lo teórico y viceversa.

Esta clasificación de los términos y los enunciados permite expresar de un modo simple la estructura de las teorías en tanto que cálculos interpretados: una teoría T es un par T = <A, R>, donde A es el conjunto (o la conjunción) de todos los axiomas y R es el conjunto (o la conjunción) de todas las reglas de correspondencia. Las teorías empíricas son cálculos interpretados: A es el cálculo axiomático, *R* proporciona la interpretación empírica. Se dirá que faltan los enunciados puramente observacionales, pero no es así, pues se derivan de A y *R y,* por tanto, ya están incluidos en T. Se comienza por una serie de observaciones particulares, que quizá den lugar a generalizaciones empíricas de las que queremos dar cuenta. Para ello se postulan una serie de entidades teóricas regidas por ciertas leyes expresadas por A. Una vez determinadas las reglas *R* que expresan los efectos observacionales de las entidades teóricas, se derivan, si se tiene éxito, de A y *R* las generalizaciones o fenómenos de los que queríamos dar cuenta (explicación), u otros nuevos que servirán para contrastar la teoría (predicción).[2]

Esta estructura es plásticamente expresada por Hempel mediante su famosa metáfora de la red: "Una teoría científica puede entonces ser comparada con una red espacial compleja: sus términos están representados por los nudos, mientras que los hilos que unen éstos corresponden en parte a las definiciones y en parte a las hipótesis fundamentales y derivadas incluidas en la teoría. Todo el sistema flota, por así decir, sobre el plano de la observación y está anclado en él mediante reglas de interpretación. A éstas se las puede considerar como cuerdas que no forman parte de la red pero que conectan ciertas partes de ella con lugares específicos de la observación. En virtud de estas conexiones interpretativas, la red puede funcionar como teoría científica: a partir de ciertos datos observacionales podemos ascender a través de una cuerda interpretativa hasta algún punto de la red teórica, de aquí pasar a través de definiciones e hipótesis a otros puntos, desde los cuales otras cuerdas interpretativas permiten descender al plano de la observación" (1952, §7).

Hasta aquí las líneas generales del análisis de las teorías empíricas, de sus constituyentes, naturaleza y funcionamiento, propio de la Concepción Heredada. Para concluir con este enfoque comentaremos en las próximas secciones dos cuestiones que fueron objeto de especial atención: a) la naturaleza de las reglas de correspondencia y sus consecuencias para la supuesta eliminabilidad de los términos teóricos; y b) la distinción teórico/observacional y la naturaleza de la base empírica. En esta última cuestión presentaremos las críticas internas de Popper y el último Hempel, que sugieren modificaciones fundamentales.

4. Las reglas de correspondencia y la cuestión de la eliminabilidad de los términos teóricos

Hasta ahora hemos presentado las reglas de correspondencia de una forma extremadamente general, sin especificar su estructura, tan sólo hemos dicho de ellas que contienen términos tanto teóricos como observacionales. Ésta es una caracterización muy débil compatible con prácticamente cualquier forma sintáctica.[3] Pero eso no fue siempre así y en los inicios se pretendió imponer constricciones más fuertes sobre la forma de las reglas, constricciones derivadas de la finalidad que se les atribuía.

2. Algunos representantes de esta concepción, como Nagel (cf. 1961, cap. 5, §11.3, también Hesse, 1966) incluyen, como "elemento adicional", además de axiomas y reglas, *modelos*. Pero la referencia a modelos en esta concepción es excepcional y, cuando se hace, está poco desarrollada, mal estructurada con el resto de elementos y, en general, es muy confusa. Al no pasar a formar parte de la versión oficial, no vamos a detenernos aquí en ella.

3. Aunque no con cualquiera, al menos no si se entiende la inclusión de términos de ambos tipos en sentido esencial, e.e., que las reglas contienen *esencialmente* términos tanto teóricos como observacionales. Si eso es así y, supuesto que *t* es el único término de uno de los tipos, la regla no puede tener la forma, P.C., "a A (y(t) *v -' y(t)) -> (3"*, pues ahí *t* no ocurre esencialmente (a no ser que ocurra esencialmente en (x o en (3); es sólo este tipo de formas el que queda excluido por la caracterización general anterior.

4.1. INELIMINABILIDAD DE LOS TÉRMINOS TEÓRICOS

Las reglas expresan la conexión de los términos teóricos con la experiencia observable, cargan de contenido o significación empírica tales términos. Como hemos indicado más arriba, una de las preocupaciones de estos filósofos, de orientación general neoempirista, era dejar clara la legitimidad semántica de las expresiones científicas, por contraposición a otras según ellos carentes de sentido. Esa legitimidad la da el anclaje en la experiencia observable, y es tanto mayor cuanto más fuerte sea dicho anclaje observacional. Si eso es así, entonces la alternativa más fuerte a considerar es que los términos teóricos sean completamente definibles mediante términos observacionales, esto es, que haya *definiciones explícitas* de los términos teóricos mediante vocabulario observacional. Más arriba hemos indicado que no puede haber definiciones explícitas de los términos primitivos del cálculo axiomático, pero en ese contexto estaba claro que teníamos en cuenta únicamente la intervención de términos teóricos. La opción ahora es definir explícitamente los términos primitivos teóricos del formalismo abstracto, no mediante otros términos teóricos sino mediante términos observacionales; puesto que con los términos teóricos primitivos se definen los restantes términos teóricos, la alternativa implica la eliminabilidad total de los términos teóricos, los convierte en meras abreviaturas de expresiones más complejas cuyos componentes se refieran sólo a entidades observables.

Esta alternativa determina la forma que deben tener las reglas de correspondencia, a saber, la de las definiciones explícitas: para cada término t de V_T hay una regla de correspondencia que tiene la forma "y(t) t-> cp(o,, ..., o_k)", donde t es el único término teórico que ocurre en y y cp sólo contiene como términos descriptivos k términos observacionales o,, ..., *ok.* Si esta propuesta fuese viable, entonces la teoría estaría utilizando siempre un vocabulario en realidad exclusivamente observacional, sólo que usando a menudo abreviaturas notacionales. No es difícil ver que ello conferiría la máxima legitimidad observacional al lenguaje de la teoría: las entidades teóricas "desaparecen" o, más suavemente, se reducen a, o se construyen como, complejos de observables.

El principal defensor de dicha propuesta, Carnap, reconoció pronto su inviabilidad. El problema principal lo ejemplificaban los términos de propiedades *disposicionales,* como `frágil', `elástico' o `soluble'. Estos términos se refieren a propiedades que se caracterizan por cierta reacción ante ciertas circunstancias; por ejemplo, un cuerpo es soluble si, al sumergirse en agua, se disuelve (las propiedades no disposicionales se denominan *categóricas,* p.ej. ser molécula de ácido sulfúrico o ser vertebrado). Aunque no lo parezcan en primera instancia, muchas de las propiedades teóricas de la ciencia son disposicionales, lo que representa un serio problema para el programa eliminativista. Las únicas definiciones explícitas para estas propiedades deben tener la forma

(1) *Dx H (Cx -* Rx),*

donde *D* es la propiedad disposicional que queremos definir, C son las condiciones observables en las que se actualiza la disposición y R es la respuesta observable que la disposición produce en las condiciones C; por ejemplo, "x es soluble syss, si x se sumerge en agua, entonces x se disuelve". El problema es que, por la lógica del condicional material, estas definiciones atribuyen la propiedad disposicional a todo individuo que no sea sometido a las condiciones C, por ejemplo a toda sustancia que no se sumerja nunca en agua, lo cual es inaceptable. La solución de Carnap (cf. 1936-1937, §7) es abandonar la propuesta reduccionista radical, optar por otra que no tenga consecuencias inaceptables aunque no permita la eliminación vía definición explícita de los términos teóricos.

La nueva propuesta consiste en modificar la forma de las reglas para términos disposicionales del siguiente modo:

(2) $$Cx \to (Dx\ \mathrm{H}\ Rx).$$

Es claro que (2) no tiene los problemas de (1), pero también que no permite eliminar el término disposicional al no ser una definición explícita, sino una definición "parcial" o, en expresión de Carnap, un *enunciado de reducción parcial.* Una reducción (definición, eliminación) parcial es, simplemente, una no reducción (definición, eliminación), pues ahora, cuando las condiciones de prueba C no se satisfacen, la posesión o no de la propiedad disposicional D queda simplemente indeterminada; por ejemplo, de una sustancia que nunca se sumerja en agua queda indeterminado, según (2), si es o no soluble. Eso sería una consecuencia inaceptable si pretendiéramos que (2) es una definición, esto es, si pretendiéramos que *determina las condiciones necesarias y suficientes* de que algo sea o no soluble. Pero ahora ya no se pretende tal cosa.'

Los términos disposicionales no son los únicos que sugieren estas modificaciones, aunque son los que mejor las ilustran, al menos en primera instancia. Se reconoce que lo mismo ocurre con términos en principio no disposicionales, como `temperatura'. También en estos casos las reglas sólo proporcionan interpretaciones empíricas parciales. Por ejemplo, la regla "si al introducir un tubo de vidrio con mercurio en una sustancia y después introducirlo en otra, la columna de mercurio asciende, entonces la segunda sustancia está a mayor temperatura que la primera" interpreta sólo parcialmente el término `temperatura', pues no se aplica a sólidos, o a temperaturas muy altas, o muy bajas, etc. (cf. p.ej. Carnap 1966, cap. XXVIII). Y lo mismo ocurre con las demás reglas para el término. Podría pensarse que la situación se resuelve conyuntando todas las reglas de correspondencia para cada término, pero, y esto es verdaderamente importante, no es así. La conyunción proporciona la total interpretación *empírica,* pero no constituye una definición o eliminación del término, pues no incluye situaciones en las que, según los axiomas teóricos, también se aplica; por ejemplo, no hay ninguna regla de correspondencia directa para la situación consistente en que el centro del Sol está a mayor temperatura que su superficie.

Una vez abandonada la propuesta eliminativista radical y abierta la puerta a reglas de correspondencia no definicionales, no hay especial razón para imponer constricciones

4. Es cierto que ésta no es la única alternativa al problema, hay otras que mantienen la vocación definicional. La más inmediata es sustituir en (1) el condicional material por un condicional contrafáctico o de necesidad física (cf. cap. 5), pero para la mayoría de nuestros filósofos neoempiristas (especialmente Carnap, pero no sólo él) las soluciones en esta línea son inaceptables por apelar a conceptos modales, como el de *necesidad,* que prefieren evitar en una reconstrucción lógica de la ciencia.

muy específicas a la forma de las reglas. De este modo se acaba admitiendo como regla cualquier tipo de enunciado mientras contenga esencialmente términos teóricos y observacionales. O para ser más precisos, de los tres tipos de enunciados que puede contener una teoría científica, a saber, enunciados sólo con términos teóricos, enunciados sólo con términos observacionales y enunciados con términos tanto teóricos como observacionales, se seleccionan estos últimos (o una subclase-representante de los mismos) como las reglas de correspondencia de la teoría sin importar la forma sintáctica que tengan (incluso a veces, como señalaron Ramsey, Carnap y Braithwaite, las reglas pueden tener la forma de definiciones explícitas de términos observacionales mediante términos teóricos). El propio Carnap acaba poniendo como ejemplo de regla de correspondencia enunciados que simplemente conectan mediante un condicional material un término teórico con otro observacional, por ejemplo "si *u* es más caliente que v, entonces la temperatura de *u* es mayor que la de v" (cf. Carnap, 1956, §V). Esta liberalización en la forma lógica de las reglas va acompañada de otra en apariencia más radical, a saber, ni siquiera es necesario que todo término teórico intervenga esencialmente en al menos una regla de correspondencia. Pero esta liberalización es más radical sólo en apariencia. En efecto, si no se trata de *definir* observacionalmente los términos teóricos, si basta con que estén conectados con términos observacionales mediante las reglas, entonces no es necesario que esa conexión deba ser *directa* para todos y cada uno de los términos teóricos; esto es, puede que algunos se conecten sólo *indirectamente* con la base observacional a través de su conexión axiomática con términos teóricos conectados directamente con la base observacional. Algunos términos teóricos tendrán varias reglas (p.ej. varios enunciados de la forma (2) con diferentes Cs y Rs), pero otros pueden no tener ninguna y no por eso carecen de contenido empírico (y con ello de legitimidad semántica) pues adquieren tal contenido (legitimidad) indirectamente por su conexión a través de los axiomas con otros términos para los que sí hay reglas de correspondencia.

Resumiendo, los términos teóricos (primitivos), por tanto, no son eliminables mediante definiciones explícitas a partir de términos observacionales. Son términos con "vida propia" que fijan su contenido o significado por dos vías, cada una de las cuales los "define" sólo parcialmente: *a)* su conexión con otros términos teóricos a través del cálculo axiomático, y b) su conexión, directa o indirecta, con términos observacionales a través de las reglas de correspondencia. Así pues, el significado de los términos teóricos no es puramente observacional, las conexiones axiomáticas contribuyen esencial e ineliminablemente al mismo.

Nótese que tampoco es viable la alternativa opuesta, a saber, que el significado fuese puramente teórico, que los axiomas diesen el significado (implícito) completo de los términos teóricos y que las reglas fuesen hipótesis empíricas que no contribuyeran al significado de tales términos. Si eso fuera así, los axiomas teóricos, las leyes, serían (como en las ciencias formales) verdades analíticas, verdades en virtud del significado de los términos que involucran, carentes por tanto de todo contenido empírico; sólo tendrían contenido empírico las reglas de correspondencia, la mayoría de las afirmaciones de las teorías consideradas empíricas serían, contra toda apariencia, analíticas. Puesto que ésta parece una conclusión claramente rechazable, el significado de los términos teóricos no puede depender de los axiomas solos, como tampoco depende de las reglas solas, sino de ambos a la vez. Aquí, sin embargo, se abre uno de los problemas más profundos de la filosofía de la ciencia, relativo al significado de los términos teóricos y al estatuto epistémico de las afirmaciones científicas. El lector avisado habrá advertido que, si el significado de los términos teóricos no es constituido por los axiomas solos, ni por las reglas solas, sino por ambos a la vez, entonces parece que se puede decir de "los axiomas más las reglas" lo mismo que se dice en las ciencias formales de los axiomas, a saber, que puesto que constituyen el significado de los términos, entonces axiomas y reglas son analíticamente verdaderos, verdaderos en virtud de definiciones. Ésta es en parte la brecha por la que Quine ataca la distinción analítico/sintético (cf. 1951) al poner de manifiesto toda una serie de problemas en la distinción tradicional que obligarán a revisar la relación entre *analítico y empíricamente revisable. No* podemos detenernos aquí en esta cuestión. Para concluir con las reglas de correspondencia mencionaremos brevemente dos modos en los que, en este enfoque, se acepta que los términos teóricos son eliminables en cierto sentido en favor de los observacionales, aunque no mediante definiciones explícitas.

4.2. ELIMINABILIDAD *A LO* RAMSEY

El primero de los procedimientos se debe a F. P. Ramsey. Ramsey mostró (cf. 1929) que, dada una teoría T = <A, *R*>, siempre es posible dar con otra que tenga el mismo contenido empírico, es decir las mismas consecuencias observacionales, y que no use términos teóricos. El expediente es sencillo: sustituimos cada enunciado "y(t)" de *A o* de R que contenga un término teórico *t* por otro de la forma "3x y(x)"; por ejemplo, sustituimos "si *u* es más caliente que v entonces Temp(u) > Temp(v)" por "3P (si *u* es más caliente que v entonces *P(u) > P(v)".* En realidad no se realiza la existencialización en cada enunciado suelto, pues cuando un mismo término teórico aparece en varios enunciados, la variable para su existencialización debe ser la misma. Una teoría T = <A, *R*> (con p términos teóricos y q términos observacionales) se puede identificar con la conyunción *"Ax, A Ax2 A ... Rc, A Rc2 A ...",* de los axiomas de *A y* las reglas de *R,* que abreviaremos mediante *"AR (t,, ..., t,,, o,, ..., o4)".* Si *T* es una teoría, la versión-Ramsey de *T* es:

$$TR = 3x,, ...,3x,, AR(x,, ..., x,,, o,, ..., o,,).$$

Pues bien, se puede demostrar entonces que todo enunciado (puramente) observacional que se sigue de T se sigue también de 7[1], enunciado éste que, como hemos visto, no contiene *términos teóricos.*

En este sentido los términos teóricos son ciertamente eliminables. Sin embargo, este resultado tiene poca trascendencia filosófica si lo que se pretende es prescindir de las entidades teóricas (lo que no era la pretensión de Ramsey). En primer lugar, la versión-Ramsey de la teoría requiere lógica de segundo orden, pues algunas constantes descriptivas teóricas serán predicados, con lo que la versión-Ramsey de enunciados con predicados cuantificará sobre variables predicativas (como en nuestro ejemplo, que cuantifica sobre una variable de función, un tipo de variable predicativa). Segundo, relacionado con lo anterior, y verdaderamente importante, la versión-Ramsey *P* prescinde de, o elimina, los *términos* teóricos, cierto, pero no las *entidades* teóricas. T presupone la existencia de entidades teóricas tanto como T, pues las variables introducidas en TR deben tener algún valor. En T, las entidades teóricas son los referentes de las constantes descriptivas teóricas, en TI son los valores de las nuevas variables introducidas. Por tanto, mediante este expediente, al desaparecer los *términos* teóricos, nos libramos quizá nominalmente de la formulación del problema semántico acerca de la legitimidad de estos términos bajo sospecha para el empirista, pero no nos libramos en absoluto (ni siquiera nominalmente) de la cuestión ontológica relativa a las entidades teóricas pues la nueva versión sigue apelando a ellas, aunque mediante otro recurso expresivo, las variables. Lewis (1970) utiliza el método de Ramsey para mostrar cómo se puede dar una definición "funcional" de los términos teóricos, esto es, cómo se puede denotar una entidad teórica mediante una expresión que no contenga términos teóricos, a saber, mediante una descripción que describa su función en la teoría; por ejemplo, la masa es la denotación de la descripción "la función x; tal que ...", donde los puntos suspensivos contienen la versión-Ramsey de la Mecánica Clásica (y j es un subíndice concreto). Queda claro por tanto que el método de Ramsey no permite eliminar las entidades teóricas sino tan sólo el modo usual de referirnos a ellas mediante constantes predicativas (o funcionales).

4.3. ELIMINABILIDAD *A LO* CRAIG

La alternativa de Ramsey no depende de que los dos grupos de términos sobre los que se realiza la eliminabilidad relativa sean los teóricos y los observacionales en el sentido pretendido, se aplica a cualquier teoría en la que dividamos el vocabulario en dos conjuntos disjuntos. Lo mismo sucede con el segundo expediente de eliminación, debido a Craig y que es consecuencia de un teorema de lógica formal del mismo autor. Craig mostró (cf. 1953 y 1956) que si el vocabulario *V* de una teoría *T* se divide en dos conjuntos disjuntos de términos V, y V_2, y la teoría satisface ciertos requisitos formales (no especialmente estrictos), entonces siempre existe otra teoría T* que usa términos sólo de un tipo, digamos V, y de la cual se derivan los mismos V,-enunciados (e.e. enunciados que involucran sólo términos de *V,)* que se derivaban de *T; T y T** son por tanto V,-equivalentes. Además T* no contiene, contrariamente a la versión de Ramsey, recursos expresivos nuevos, nuevas variables. Aplicado a la distinción entre los vocabularios teórico y observacional, este resultado implica que las mismas consecuencias observacionales que se derivan de una teoría con términos teóricos, se derivan también de otra teoría que no contiene términos teóricos ni variables que los sustituyan. En este sentido parece que los términos teóricos son eliminables o prescindibles, y ahora no se trata sólo de los *términos* sino también de las *entidades* teóricas mismas.

Pero, como antes, esta vía no es tan prometedora para el eliminativista como a primera vista parece. Aunque ahora parecen ser eliminables las entidades teóricas mismas, ello es sólo "en principio". En primer lugar, la eliminabilidad es sólo *a posteriori,* esto es, una vez tenemos previamente la teoría original con sus términos teóricos, por lo que la teoría puramente observacional "sustituta" no puede desempeñar ninguna función heurística o metodológica efectiva. Pero además el expediente es tal que la teoría puramente observacional T* consiste siempre en un conjunto infinito de axiomas no simplificable de manera significativa (ni siquiera mediante esquemas axiomáticos). Las consecuencias filosóficas de la eliminabilidad a lo Craig son prácticamente nulas, a lo sumo satisfacer la mala conciencia de las mentes empiristas radicales con una eliminabilidad en principio completamente irrelevante para la práctica científica. Pero si nos contentamos con eso, ni siquiera se precisa de complejos resultados formales, pues es trivial construir una teoría T' puramente observacional y observacionalmente equivalente a otra T que use sólo términos observacionales: simplemente seleccionamos como axiomas para T' todas las (infinitas) consecuencias puramente observacionales de T (dado $T = AR$, $T' =$ {(X la es consecuencia de AR y contiene sólo términos observacionales).

5. La distinción teórico/observacional y la naturaleza de la base empírica

Hasta aquí hemos procedido como si estuviera clara la naturaleza de los términos, y las entidades, observacionales. Pero eso dista mucho de ser así y en la Concepción Heredada se plantearon, casi desde los inicios, diversos problemas relativos a la naturaleza de estos términos. Comentaremos aquí muy brevemente tres que están íntimamente conectados, dos de los cuales hemos mencionado anteriormente: a) el problema ontológico de la naturaleza de las entidades teóricas, la fundamentación a partir de ella de la distinción teórico/observacional y el carácter rígido o fluido de tal distinción; b) el problema semántico de la supuesta neutralidad teórica de los términos observacionales; c) el problema metodológico de la supuesta naturaleza observacional de la base empírica de contrastación, no sólo del conjunto de nuestro conocimiento, sino para cada teoría científica particular. Estas cuestiones motivaron multitud de debates, han sido tratadas por casi todos los filósofos de la ciencia y en relación con ellas surgieron algunas de las posiciones que dieron lugar a concepciones alternativas a la Concepción Heredada. En el primer parágrafo nos limitaremos a los aspectos más generales, y en los dos siguientes desarrollaremos algunos problemas específicos.

5.1. ENTIDADES OBSERVABLES Y DISTINCIÓN TEÓRICO/OBSERVACIONAL

Para muchos empiristas y positivistas lógicos del período de entreguerras, y especialmente para aquéllos en tomo a los cuales se gestan las primeras versiones de la concepción estándar, la fundamentación del conocimiento en la experiencia se entendía en términos fenomenalistas: los primeros datos sobre los que se construye todo conocimiento, que justifican nuestras creencias, son datos de la experiencia fenoménica. Esta posición extrema plantea múltiples dificultades en las que no podemos detenemos aquí, y el fenomenalismo termina por ser abandonado, al menos como base de experiencia para las teorías científicas. Las entidades fenoménicas *(qualia,* datos sensoriales) son entonces sustituidas por entidades que se caracterizan simplemente como "directamente presentes a la observación". Sin embargo, esta nueva versión, que se convertirá en estándar, tiene sus propios problemas, el principal de ellos su vaguedad. Las entidades fenoménicas son claramente distinguibles de las no fenoménicas, pero por su "privacidad" o subjetividad son poco plausibles como constituyentes de la base de experiencia para la ciencia. Las entidades observables, públicas, parecen en primera instancia poder desempeñar más plausiblemente tal función, pero ahora el problema es la dificultad para distinguir nítidamente entre entidades observables y no observables (teóricas).

Inicialmente, Carnap intentó una caracterización precisa de los términos observacionales como aquellas expresiones del lenguaje tales que, en condiciones normales, un observador puede determinar a través de una serie de observaciones, y con un alto grado de confirmación, si el término se aplica o no en una situación dada (cf. Carnap, 1936-1937). Esta caracterización es inadecuada, pues, sin más precisiones, se aplica también a predicados pretendidamente no observacionales. En escritos posteriores, Carnap se limitó a caracterizar el vocabulario observacional como aquel que se refiere *a entidades observables* (cf. 1956, §11): los términos observacionales son predicados que denotan propiedades observables de acontecimientos o cosas, o relaciones observables entre ellos. Pero es claro que si no se especifica lo que caracteriza las *entidades* observables, simplemente se desplaza el problema. Hempel presentó las cosas de modo parecido al hablar de entidades o fenómenos "que podemos observar directamente" (1958, §11). La cuestión es: ¿qué cuenta como observación directa? Aunque no se da una respuesta a esta cuestión, parece que en este primer momento se sigue pretendiendo que la distinción que hay tras ella es relativamente rígida y no dependiente del contexto.

Después de una serie de críticas, especialmente de Putnam (cf. 1962, y también Hanson, 1958), el primer exponente de la doctrina oficial en reconocer el carácter fluido de la distinción fue Nagel, quien en su monografía de 1961 afirma: "es dudoso que haya un sentido riguroso que pueda ser asignado con utilidad a la palabra `observable'; y en la medida en que la distinción [entre leyes empíricas y axiomas teóricos] se base en el contraste entre lo que es observable y lo que no, la distinción patentemente no es nítida" (cap. 5, §1). Carnap, en su monografía de 1966, acabó también aceptando explícitamente que la distinción es gradual. Por ejemplo, si la percepción visual directa cuenta como observación, ¿qué ocurre con la asistida de lentes?, ¿y de prismáticos o catalejos?, ¿y de telescopio óptico?, ¿y de telescopio de radio? O, para ir en la dirección contraria, ¿cuenta como observación la realizada con lupa?, ¿y con microscopio óptico?, ¿y con microscopio electrónico? ¿Observa directamente el físico la trayectoria de una partícula cuando ve el rastro en una cámara de niebla?; ¿se observa la corriente eléctrica al ver moverse la aguja de un amperímetro? Preguntas como éstas son las que le hacen concluir que "hay un continuo que comienza con observaciones sensoriales directas y pasa a métodos de observación enormemente complejos e indirectos, [...] el físico habla de observables en un sentido muy amplio, comparado con el estricto sentido que da el filósofo a la palabra, pero en ambos casos la línea de separación entre lo observable y lo inobservable es muy arbitraria" (cap. XXIII).

A pesar de la fluidez o vaguedad de la distinción, tanto Nagel como Carnap insisten en su utilidad para la caracterización de la naturaleza y estructura de las teorías. Así, por ejemplo, Carnap insiste en que las leyes empíricas son las que contienen términos que refieren a entidades "directamente observables por los sentidos o medibles mediante técnicas relativamente simples" *(ibid.).* Pero sorprendentemente menciona ahora como ejemplos, además de regularidades cualitativas simples (como la típica "todos los cuervos son negros") también leyes cuantitativas (como las de los gases, que relacionan presión, volumen y temperatura para los gases, o la ley de Ohm, que relaciona potencial, resistencia e intensidad de corriente) que involucran términos que había considerado tradicionalmente teóricos (como `temperatura' o `intensidad de corriente eléctrica'). El cambio se debe sin duda a la aceptación de la fluidez de la distinción. La cuestión que surge ahora es si en estos nuevos términos la distinción teórico/observacional puede desempeñar la función para la que fue originalmente introducida. Numerosos críticos, como Putnam (1962), Shapere (1965), Maxwell (1961), Achinstein (1968) o el propio Hempel posteriormente (1973), argumentaron en contra de ello. Veamos algunas de las principales dificultades.

5.2. NEUTRALIDAD TEÓRICA DE LOS TÉRMINOS OBSERVACIONALES Y CARGA TEÓRICA DE LOS HECHOS

El principal motivo de la introducción de la distinción teórico/observacional era proporcionar legitimidad semántica, según los criterios empiristas, a los términos "sin conexión empírica inmediata" que las teorías científicas introducen a través de sus leyes para dar cuenta de los fenómenos. Esta finalidad semántica va acompañada de otra metodológica, pues se pretende que la base observacional es la que proporciona la experiencia "neutra" con la cual contrastar las afirmaciones de la teoría. Esta neutralidad teórica de la base de contrastación parece en primera instancia fundamental, pues de lo contrario parecería que la teoría resulta autojustificativa. Si la experiencia observacional que se usa para contrastar la validez de una teoría fuese dependiente de la teoría en cuestión, esto es, si la elaboración de los informes observacionales que sirven de base de contrastación presupusiera la validez de la teoría, entonces tendríamos un círculo autojustificativo. Por tanto, la base observacional, si ha de servir para la contrastación, debe ser teóricamente neutral. Esta cuestión está íntimamente ligada a la anterior, pues la distinción T/O parece problemática en la medida que lo que consideramos usualmente observaciones requieran adiestramiento o conceptualización teórica.

Ya antes de la formulación explícita de la Concepción Heredada, Duhem (1914) objetó a lo que iba a ser este elemento de la misma. Duhem rechazó que la observación esté libre de conceptualización teórica, aunque usualmente sí lo está respecto de algunas teorías, esto es, puede ser que las observaciones no presupongan una teoría que usa de ellas en su contrastación. Debe recordarse que originalmente la observabilidad no se pretende relativizada a una teoría, los estados de cosas son observables o no sin más, y los que supuestamente lo son se usan para contrastar unas teorías u otras. Lo que constató Duhem es que toda observación, o mejor dicho todo informe observacional, supone una interpretación de los datos de los sentidos, y una interpretación no es más que una conceptualización teórica, sea explícita o implícita. Quizá el aparato conceptual interpretador que genera la base observacional no corresponde a cierta teoría que usa dicha base en la contrastación, pero en cualquier caso corresponderá a otro "constructo teórico"; este constructo presupondrá a su vez otro en la descripción de sus propios fenómenos empíricos y así sucesivamente. No hay (en general) una autojustificación inmediata de cada teoría, pero sí un círculo global autojustificativo en el conjunto de la ciencia. Duhem ejemplifica esta tesis con múltiples casos históricos y con referencias a la práctica experimental usual en laboratorios (cf. p.ej. su ejemplo de la oscilación de una barra de hierro en cierto mecanismo y la medición de la resistencia, 1914, p. 218). Ésta es la base del conocido *holismo* de Duhem, de gran influencia en el siglo xx, y sobre el que volveremos brevemente más adelante (cap. 11).

En el Círculo de Viena fue Neurath quien más radicalmente se distanció de la tesis oficial inicial de la neutralidad de los "informes protocolares de experiencia" y a él, y a Duhem, apelará después Quine como inspiradores de sus propias tesis holistas. Pero en el campo específico de la filosofía de la ciencia, en el contexto neopositivista de entreguerras, fue Popper quien primero expresó de forma explícita el componente teórico de la base empírica de contrastación, lo que después se denominará *carga teórica de los hechos.* Popper es uno de los mayores críticos de las tesis centrales del Círculo de Viena (al que, como insiste en declarar, no pertenecía), pero comparte en general la caracterización de las teorías como cálculos interpretados. El principal punto de desacuerdo tiene que ver con la epistemología de la contrastación; como veremos en detalle en el capítulo 12, frente al confirmacionismo y la lógica inductiva de Carnap, de los que Popper fue el primer y más severo crítico, él defiende una lógica de la falsación. Pero otro de los puntos de disensión tiene que ver con nuestra actual cuestión. Aunque no sacara todas las consecuencias (consecuencias que acaban cuestionando sus tesis falsacionistas más radicales, cf. cap. 12, §4 y §5), declaró abiertamente que en la determinación de la base de contrastación, "en la determinación de los hechos", interviene un conocimiento de fondo necesitado de aceptación previa. Al someter a prueba una teoría, señala, no sólo intervienen en ella las condiciones iniciales y los supuestos auxiliares (según el esquema comúnmente admitido) sino también cierto *conocimiento de fondo* sobre los hechos singulares. Este conocimiento de fondo, que "contiene" lo que se acepta como hechos, se puede considerar constituido por teorías de bajo nivel que se aceptan como altamente corroboradas y que no entran en el juego de la contrastación. Y no entran en el juego por *decisión* (no necesariamente consciente): "Siempre que una teoría se somete a contrastación [...] hay que detenerse en algún enunciado básico que decidimos aceptar: si no llegamos a decisión alguna a este respecto, [...] la contrastación no lleva a ninguna parte" (1935-58, §29).

Esta idea pone de manifiesto lo que se denomina, siguiendo a Hanson, la *carga teórica de los hechos.* Hanson fue el primero en hacer de este fenómeno algo esencial para el análisis de la ciencia y en defender la opinión de que ello modifica dramáticamente la visión tradicional de la misma. Apoyándose en los casos de ambigüedad perceptiva estudiados por la psicología de la Gestalt, destacó la importancia del contexto y los elementos organizativos ya en la percepción. IIustró esta tesis con el siguiente ejemplo (cf. 1958, cap. 1).

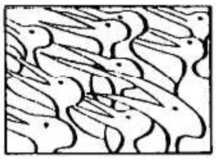


FIG. 1

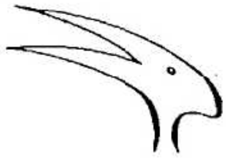


FIG. 2

FIG. 3

Al contemplar las figuras 1 y 3, se ven en los extremos inferiores derechos dos animales diferentes a pesar de que son "la misma cosa" (figura 2); además, cuando contemplamos el dibujo aislado podemos ver una cosa u otra, pero no las dos a la vez. En parte se ve lo mismo (hay una excitación similar del córtex) y en parte no, y el sentido interesante de `ver' relevante para la ciencia es el segundo. No se trata de interpretaciones diferentes *a partir de* una misma visión; eso, afirma, no tiene sentido, pues "interpretar", si se quiere llamar así, es parte constitutiva de "ver". Además, el contexto puede no darse explícitamente, no es esencial al hecho que el ejemplo pretende mostrar que en él el contexto esté manifiesto; piénsese, afirma Hanson, en lo que ven un físico y un profano ante los rastros de una cámara de niebla. Este fenómeno, que salvo radicales diferencias culturales tiene en la vida cotidiana escasa trascendencia, es determinante en la ciencia, donde la dependencia del contexto es altamente teórica y, en momentos de cambio conceptual en los que se contraponen diferentes contextos de fondo, deviene crucial. Cuando Tycho y Kepler ven el Sol al amanecer, dice Hanson, en parte ven lo mismo y en parte no: Tycho ve un astro móvil, Kepler uno estático, "y es el sentido en que no ven la misma cosa el que debe tomarse en cuenta para entender los desacuerdos que existen dentro de la física" *(ibid. B).*

Consideraciones parecidas a éstas se encuentran en otros autores. Toulmin afirma que los fenómenos no sólo son seleccionados por la actividad teórica sino que incluso están definidos por la misma: hay una "continua interacción entre teoría y hecho [...], las teorías se construyen sobre la base de hechos, a la vez que les dan significación y aun determinan lo que son "hechos" para nosotros" (1961, p. 95). Feyerabend, defendiendo su pluralismo metodológico (cf. 1964, 1965, 1981), sostiene que la descripción de los hechos depende siempre de una teoría (aunque en general no de la que se contrasta) y que hay hechos que sólo pueden salir a la luz con ayuda de teorías alternativas incompatibles. Rechaza por tanto la tesis de que "los hechos existen y están disponibles independientemente de la consideración de alternativas a la teoría que se contrasta" (1981, §5). La consecuencia de ello es lo que él caracteriza como la inversión de la relación tradicional entre teoría y observación. El significado de las oraciones de observación está determinado por las teorías con las que están relacionadas, no son significativas a menos que se hayan relacionado con las teorías: "la interpretación de un lenguaje de observación está determinada por las teorías que usamos para explicar lo que observamos, y cambia tan pronto como estas teorías cambian" (1981, §6). En su monografía de 1975 ilustró esta tesis con innumerables ejemplos extraídos de la historia de la ciencia que, en su opinión, la confirman; muestra, por ejemplo (cap. 10), cómo, en la controversia entre Galileo y los aristotélicos sobre las consecuencias de la observación telescópica de diversos fenómenos astronómicos, el primero se apoyaba, entre otros supuestos que los aristotélicos tenían buenas razones para rechazar, en una teoría óptica inaceptable. Kuhn, como veremos, sostuvo por su parte que las teorías contienen elementos que determinan el contenido de la experiencia y que defensores de teorías diferentes viven en mundos experienciales diferentes. También Lakatos apuntaba en la misma dirección cuando, siguiendo a su maestro Popper, afirmaba que en la contrastación no comparamos la teoría con hechos neutros sino con otras teorías más básicas presupuestas por los hechos.

El fenómeno de la carga teórica de los hechos, y el ofrecer una imagen de las teorías y de la actividad científica adecuada a este fenómeno y a los casos históricos del mismo, es una de las principales motivaciones de las nuevas concepciones que surgen en torno a estos *nuevos filósofos de la ciencia* (así denominados, en su día, por contraposición a la Concepción Heredada). Nosotros dejamos provisionalmente la cuestión aquí sólo apuntada y volveremos sobre ella en los dos próximos capítulos. Es esencial darse cuenta de que toda esta discusión presupone una identificación casi siempre aceptada implícitamente. Se comienza cuestionando la neutralidad teórica de los informes observacionales y se concluye que los datos, fenómenos o hechos que constituyen la base de experiencia en la contrastación están teóricamente cargados. Ello supone la identificación entre *a) informes de experiencia o datos de contrastación y b) informes observacionales.* Parte de las críticas expuestas se deben entender como cuestionando esta identificación. Para concluir nos detendremos brevemente en este aspecto de la cuestión.

5.3. OBSERVACIÓN Y BASE EMPÍRICA

Las teorías empíricas se generan a partir de una serie de fenómenos de los que, tras la elaboración teórica, se pretende dar cuenta; esos mismos fenómenos, u otros nuevos del mismo tipo, constituyen el ámbito de experiencia sobre el que la teoría hace predicciones y se somete a contrastación. Llamemos a esos datos, fenómenos o hechos que constituyen el ámbito de experiencia y contrastación de una teoría, la *base empírica o base de contrastación* de la teoría en cuestión. Hemos visto que en la versión oficial de la Concepción Heredada se entiende la base empírica en términos observacionales.

Por otro lado, aceptemos, como demuestran múltiples estudios tanto empíricos como teóricos, que la observación "directa" incluye conceptualización. A pesar de ello, cabe suponer que algunos aspectos de esa conceptualización, los cognitivamente más básicos, serán generales, comunes a todo sistema cognitivo (o al menos, en su dimensión biológica-evolutiva, comunes a todos los seres humanos). Si eso es así, del hecho de que la observación presuponga cierta conceptualización no se sigue que dicha conceptualización dependa siempre esencialmente de las teorías científicas. Por tanto, si la base de contrastación *fuese* observacional, ello no implicaría que lo que cuenta como base empírica depende esencialmente de las *teorías científicas.* En realidad, pues, lo que hay implícitamente detrás de las consideraciones críticas sobre la carga *teórica (científicamente* teórica) de *todo* dato de contrastación es una puesta en cuestión del supuesto de la Concepción Heredada de que la base de contrastación es en general de naturaleza observacional. Tras muchas de las críticas a la supuesta neutralidad de las observaciones, lo que hay en realidad es un rechazo a la identificación entre base empírica de contrastación y experiencia directamente observable. El principal motivo para identificar la base empírica con la experiencia observable directa es el viejo anhelo empirista de fundamentar y justificar todo nuestro conocimiento en la experiencia sensorial. Todo conocimiento (empírico) empieza con las afecciones de nuestro entorno sobre nuestro aparato sensorial y toda justificación del mismo debe apelar en última instancia a esa "observación directa" del entorno. Pero, como dijimos más arriba (sección 3), de este supuesto razonable no se sigue que la justificación de cada pieza de nuestro conocimiento deba proceder del mismo modo, que esta tesis global sea también válida localmente. Puede ocurrir que, como organismos vivos, la interacción más básica con nuestro entorno la realicemos en términos globales perceptualmente mediante observación directa, pero que en algunas partes de nuestro sistema cognitivo, especialmente en las muy complejas que dan lugar a las teorías científicas (muy escasas y raras en términos evolutivos globales), la base de experiencia no se dé a través de observación directa inmediata. Puede que todo empiece por la observación pero, si el sistema cognitivo es modular y jerárquico, no en todas partes. Si eso es así, la base de contrastación de muchas, o (casi) todas, las teorías científicas puede estar constituida por datos o fenómenos que no sean de observación directa; y por tanto, alternativamente, lo distintivo de los términos teóricos no será que denotan entidades inobservables.

Ya en 1962, Putnam se opuso a identificar la distinción "inobservable/observable" con "teórico/no teórico". Afirmaba, por un lado, que hay teorías cuyo aparato teórico se refiere a entidades observables, y, por otro, que casi nunca los fenómenos a explicar son accesibles mediante observación directa. Se trata de dos dicotomías diferentes. Un término teórico es un término que proviene de una teoría científica y "el problema apenas tocado en treinta años que se lleva escribiendo acerca de `términos teóricos' es qué es lo *realmente* distintivo de dichos términos" (1962, §1). Poco antes, Ryle había distinguido entre expresiones de una teoría que están cargadas con el peso de esa teoría particular y expresiones que no lo están; así, p.ej., "los términos técnicos de la genética están impregnados de teoría, [...] no sólo con equipaje teórico de alguna clase, sino con el de la teoría genética" (1956, p. 90). Estas consideraciones apuntan a la idea de que un término es teórico o no en relación con una teoría en función de si depende o no, de un modo que hay que especificar, de la teoría en cuestión. Achinstein (1968, cap. 6) hace explícita esta caracterización y discute varios sentidos en que un término puede depender de una teoría. Como veremos en el próximo capítulo, durante los años sesenta, Kuhn y Lakatos hicieron también consideraciones que apuntan en la misma dirección.

El primero en dar una caracterización mínimamente articulada y elaborada de la nueva distinción que se está gestando fue Hempel en una serie de trabajos de finales de los años sesenta y principios de los setenta (1966, cap. 6, 1970, 1973). En estos trabajos, Hempel divide ahora el vocabulario básico de cada teoría en dos clases que se pretenden nítidamente separadas y relativizadas a una teoría específica. Una clase está formada por los términos con los que se describen los fenómenos a explicar, la base empírica. Estos términos constituyen el *vocabulario preteórico* o, como también dice, "previamente disponible". Estos términos preteóricos no corresponden en general a situaciones observables en sentido estricto, sino que a menudo se introducen en la ciencia en el contexto de una teoría anterior. Los otros términos descriptivos usados en la teoría son los que ella introduce para llevar a cabo la elaboración teórica que da cuenta de los fenómenos preteóricamente descritos; ellos constituyen el *vocabulario teórico* de dicha teoría. Es importante enfatizar dos puntos de esta nueva distinción: *a)* es una distinción relativizada a las teorías, un término no es teórico o preteórico sin más, sino respecto de una teoría específica, y por tanto un término puede ser preteórico en una teoría y teórico en otra; aunque no lo afirma explícitamente, de su caracterización informal parece seguirse que un término puede ser preteórico en varias teorías, aunque normalmente será teórico sólo en una; *b)* el criterio para la distinción es el uso o no del término en la descripción de los fenómenos empíricos a explicar; por tanto, la distinción será precisa en la medida en que se dé un criterio preciso para determinar qué enunciados son los que describen los fenómenos a explicar, pero Hempel no lo da.

Junto con esta nueva caracterización del vocabulario básico de una teoría, Hempel introduce otra para los enunciados. Además de enunciados puramente empíricos, la teoría contiene: *(¡)principios internos,* que son los que especifican "el escenario teórico", los que sistematizan el nuevo aparato conceptual introducido por la teoría; (ii) *principios-puente,* que indican la forma en que "se relaciona lo que ocurre a nivel del escenario teórico con los fenómenos que la teoría debe explicar" (1973, §1). Esta clasificación de los enunciados parece una nueva versión de la anterior, axiomas teóricos y reglas de correspondencia, pero no es así. Aunque hay enunciados cuyos únicos términos descriptivos son preteóricos (a saber, los informes empíricos particulares y sus generalizaciones), no hay ahora enunciados que contengan sólo términos teóricos; tanto los principios internos como los principios-puente contienen esencialmente tanto términos teóricos como preteóricos.

En cuanto a la presunta función de los enunciados en la fijación del significado de los términos, Hempel sostiene ahora que el significado de los términos teóricos no está totalmente *determinado* por los principios internos más los principios-puente. Ambos tipos de enunciados ofrecen al aprendiz de la teoría el acceso principal a la compresión de las expresiones, pero no determinan completamente su significado. La idea clásica de que el significado de los términos se fija completamente *mediante enunciados* que los conectan con otros términos es errónea; y, como ya había sugerido Putnam (1962), el problema del significado de los términos teóricos planteado en ese esquema no existe, es un pseudoproblema. El motivo es que los términos científicos adquieren su significado por vías diversas, quizá en algunos casos (parcialmente) mediante enunciados, pero usualmente de otros modos; especialmente, como los términos del lenguaje ordinario, vinculándolos a aplicaciones específicas, "mediante instancias de uso en contextos particulares" (Hempel, 1973, §7, donde menciona además a Kuhn como referencia explícita para estas ideas; cf. también al respecto las teorías causales de la referencia, p.ej. Putnam, 1975).

Por último, Hempel considera ahora que la pretensión de la Concepción Heredada de caracterizar una teoría empírica a través de su reconstrucción axiomática es inadecuada, pues siempre hay varias axiomatizaciones posibles, ninguna de las cuales expresa mejor que las otras la naturaleza de la teoría; una teoría no se puede identificar pues con un sistema *específico* de enunciados dotados de cierta estructura o sistematización.

Todas estas innovaciones del último Hempel son importantes y apuntan a elementos esenciales en la caracterización de las teorías que se desarrollarán en otras concepciones, pero en esta particular versión son sumamente insatisfactorios. Sus principales contribuciones son: a) la relativización de la distinción teórico/preteórico a las teorías y b) la caracterización de la base empírica en términos preteóricos, con lo cual los datos se consideran cargados de teoría pero no de la misma teoría para la que constituyen su base empírica. Las principales dificultades radican en: (i) la imposibilidad de distinguir entre principios internos y principios-puente, y (ii) la inexistencia de un criterio preciso para poner en obra la distinción entre términos teóricos y preteóricos.

6. Consideraciones finales

En este capítulo hemos examinado los primeros análisis que, en el contexto de la llamada Concepción Heredada, se dieron de la naturaleza y estructura sincrónica de las teorías empíricas. La idea básica que inspira este análisis es que una teoría empírica es un conjunto de afirmaciones que *a)* son susceptibles de ser estructuradas mediante relaciones de dependencia lógica y b) versan sobre la realidad física, algunas directamente y otras indirectamente a través de las primeras. El núcleo de este análisis lo constituye la noción de *cálculo axiomático empíricamente interpretado.* La articulación de esta noción supone la distinción entre vocabulario teórico y observacional, y entre afirmaciones puramente teóricas (axiomas del cálculo abstracto), afirmaciones puramente observacionales (enunciados fácticos particulares y generalizaciones empíricas) y afirmaciones "mixtas" (reglas de correspondencia).

La intuición que hay tras la idea básica es esencialmente correcta, pero el modo específico en que la Concepción Heredada la desarrolla presenta diversas dificultades. Estas dificultades son básicamente de tres tipos. De ellas, sólo la última ha recibido atención en las secciones anteriores, pues sólo ella fue tematizada por los representantes de esta concepción, las otras serán examinadas en detalle en los dos próximos capítulos.

1. La primera dificultad tiene que ver con la excesiva "rigidez" del uso que se hace de la noción de cálculo axiomático. Tal como se presenta aquí, todos los axiomas (y reglas) de una teoría están al mismo nivel, no hay unos más esenciales, básicos, y otros menos, complementarios. Nótese que estamos hablando sólo de los axiomas, no se piense por tanto que la distinción entre axiomas y teoremas tiene que ver con ésta que ahora estamos apuntando. Por lo que a los axiomas se refiere, si todos están al mismo nivel, si todos son igualmente esenciales, entonces es difícil que esta noción tan rígida, "monolítica", de teoría sincrónica permita una elucidación adecuada de las teorías en sentido diacrónico. Si todo está al mismo nivel, si no se distingue entre afirmaciones esenciales y otras sólo complementarias ("no esenciales"), entonces el más mínimo cambio implica un cambio de toda la teoría, la sustitución de una teoría por otra teoría diferente. Esto es intuitivamente insatisfactorio. Una de las principales contribuciones de los nuevos filósofos de la ciencia que estudiaremos en el próximo capítulo consiste

precisamente en llamar la atención sobre esta inadecuación y proponer un concepto de teoría mucho más rico y dúctil.

2. La segunda dificultad tiene que ver con la noción misma de sistema axiomático. Tal como se presenta la identidad de una teoría, ésta parece depender de los axiomas que se elijan en su axiomatización, lo cual es intuitivamente insatisfactorio. Parece que una teoría puede decir "lo mismo" mediante recursos expresivos diferentes. Esto no lo niegan los representantes de la Concepción Heredada pero, simplemente, la articulación "enunciativista" que hacen de la idea básica no les permite recoger plenamente las intuiciones. Como veremos en el capítulo 10, para ello es necesario hacer jugar un papel más central en la caracterización de las teorías a la noción de modelo que presentamos en la sección 2. Ésta será la principal contribución de las concepciones semanticistas que estudiaremos en ese capítulo.

3. Por último, en la discusión sobre la naturaleza de la base de contrastación, su supuesta carga teórica y su eventual naturaleza observacional, se deben distinguir dos niveles, el local y el global. Por lo que se refiere al supuesto carácter observable de la base de contrastación, una cosa es a) que las teorías científicas, globalmente consideradas como partes del sistema total de nuestro conocimiento, descansen en última instancia, por lo que a su justificación se refiere, en los modos más básicos de experiencia "observable" ("observabilismo" global), y otra *b)* que *cada* teoría científica sea tal que los enunciados que expresan los hechos con los que se contrasta involucren sólo expresiones que se refieren a situaciones observacionales básicas (` observabilismo" local). Lo primero es seguramente cierto, lo segundo es, a la luz de las teorías reconstruidas, muy poco plausible. En cuanto a la carga teórica de la base de contrastación, una cosa es c) que la determinación de los datos de contrastación presuponga "directamente" la teoría que se quiere contrastar mediante dichos datos (autojustificacionismo local), y otra d) que tal determinación presuponga otra u otras teorías vinculadas a nivel global de una disciplina, o incluso la ciencia entera, con la teoría original (holismo de contrastación). Lo primero es claramente inaceptable, lo segundo merece un juicio filosófico más detenido. Sobre estas cuestiones volveremos en los próximos capítulos.

CAPÍTULO 9

ANÁLISIS SINCRÓNICO DE TEORÍAS II. LAS CONCEPCIONES HISTORICISTAS: LAS TEORÍAS COMO PROYECTOS DE INVESTIGACIÓN

1. La ***revuelta historicista y*** la naturaleza sincrónica de las teorías

Durante los años sesenta, y en parte como consecuencia de los debates sobre algunas de las cuestiones que hemos expuesto en el capítulo anterior, se gestan y desarrollan concepciones alternativas a la Concepción Heredada que cuestionan sus supuestos fundamentales. De ellas, la que más pronto cristaliza como alternativa es la que entonces se denominó *nueva filosofía de la ciencia,* vinculada a autores como Hanson, Toulmin, Kuhn, Feyerabend y Lakatos, y mucho más tarde y sin pertenecer oficialmente al grupo, pero con orientaciones parecidas, Laudan. Una de las características de estos pensadores es su mayor preocupación por, y su mejor conocimiento de, la historia de la ciencia (el más representativo e influyente de ellos, T. Kuhn, se había dado a conocer años antes como un extraordinario y renovador historiador de la ciencia). En su opinión, la atención a la ciencia real que la historia nos presenta obliga a modificar la práctica totalidad de la imagen de la misma que se ofrece en la Concepción Heredada. Esta *revuelta historicista* propicia una revisión drástica en prácticamente todos los ámbitos metacientíficos. Aunque, como siempre, también en esta concepción las tesis centrales en los diversos ámbitos están extremadamente interrelacionadas, vamos a ocuparnos aquí, en la medida de lo posible, exclusivamente de las tesis relativas a la naturaleza y estructura de las teorías científicas en su dimensión estática o sincrónica (en el capítulo 13 nos detendremos en los aspectos diacrónicos).

Conviene advertir que, contrariamente a la Concepción Heredada, ésta no es una cuestión que reconozcan como central los nuevos filósofos, ni siquiera hacen de ella un tema de estudio explícitamente declarado (salvo quizá Kuhn en una segunda etapa). En la medida en que se ocupan de las teorías o constructos teóricos, lo hacen siempre, consecuentemente con su orientación general historicista, desde una perspectiva diacrónica, centrándose en los aspectos dinámicos de las teorías como entidades que se extienden en el tiempo, esto es, que nacen, se desarrollan y "mueren" (se desalojan mutuamente). Sin embargo, debe quedar claro que, independientemente de que se reconozca o no explícitamente, el estudio diacrónico presupone una concepción de la naturaleza sincrónica de las teorías. Cualquier análisis de la dimensión diacrónica de las teorías científicas debe partir de que las teorías diacrónicamente consideradas, en tanto que entidades que perduran en el cambio a través del tiempo, consisten en determinadas secuencias de "teorías en sentido sincrónico". La "historia" de una teoría consiste en la sucesión de las diversas "etapas" o versiones por las que pasa. Estas etapas, en tanto que imágenes "congeladas" de la teoría en cierto momento, se deben considerar aproximadamente estables o "estáticas". La cinemática de la teoría, su "historia", consiste en la sucesión de sus diversas versiones estáticas, en la sucesión de "etapas" por las que la teoría pasa.

Esta intuición básica, en la que descansa cualquier análisis de la estructura diacrónica de la ciencia, implica entonces que los análisis diacrónicos presuponen alguna noción de la estructura sincrónica de las teorías, de las etapas cuya sucesión constituye la teoría-en-el-tiempo. El análisis y discusión de la evolución de los constructos teóricos contiene pues, cuando menos implícitamente, cierta preconcepción de la naturaleza de los diversos estadios por los que atraviesa ese constructo teórico, de sus elementos constituyentes y su estructuración. Esta preconcepción de la naturaleza sincrónica de las teorías que subyace a los estudios diacrónicos puede estar mejor o peor articulada y en algunos de estos autores está, aunque no siempre explícita, altamente estructurada y elaborada. Éste es el caso particularmente de Kuhn y también, aunque en menor medida, de Lakatos (su temprana muerte le impidió concluir la reelaboración de sus ideas que estaba preparando). Revisaremos aquí las contribuciones de estos dos autores, con especial detenimiento en Kuhn, y concluiremos comentando brevemente las de Laudan quien, aunque posterior, desarrolla una posición parcialmente parecida en abierta polémica con los anteriores.

Como veremos al final, estos autores realizaron contribuciones fundamentales a la caracterización sincrónica de las teorías. Ahora bien, en su opinión, esos nuevos elementos que señalan, al estar esencialmente vinculados a la actividad científica como actividad práctica con componentes históricos y sociales ineliminables, son inaccesibles al análisis formal. Todo el proyecto original de desarrollar una lógica de la ciencia, incluida en ella la reconstrucción formal de las teorías, está según ellos abocado al fracaso. Uno de los mayores retos de la filosofía de la ciencia posterior será dar cuenta en términos formales, o semiformales, de las principales contribuciones de estos autores. En general, estos nuevos elementos van a conformar una noción de teoría mucho más dúctil que la extremadamente rígida de la Concepción Heredada. Ahora, el análisis de las teorías ha de ser tal que éstas resulten entidades susceptibles de evolución, que puedan sufrir modificaciones extendiéndose en el tiempo sin perder su identidad. Para ello es imprescindible que sus estadios, las teorías en su dimensión estática o sincrónica, sean dúctiles, tengan partes más accidentales que puedan cambiar manteniendo su identidad, esto es, preservando sus componentes más esenciales. Veamos cómo se concreta esta idea básica en las nociones de *paradigma o matriz disciplinar,* de Kuhn, de *programa de investigación,* de Lakatos, y de *tradición de investigación,* de Laudan.

2. Los paradigmas-matrices disciplinares de Kuhn

En 1962, Kuhn presenta en *La Estructura de las Revoluciones Científicas* una visión de la ciencia, de los constructos teóricos, de las comunidades científicas y de su actividad, radicalmente novedosa y contraria a la dominante hasta entonces. Se ha señalado que esa nueva perspectiva tiene muchos puntos en común con la que esbozara el científico y filósofo polaco L. Fleck treinta años antes (cf. Fleck, 1935). El propio Kuhn reconoce en la introducción a su obra no sólo la semejanza, sino la influencia de las ideas de Fleck. Pero, reconocidos sus méritos como precursor adelantado, es indudable que por la articulación y desarrollo de las tesis, por la elaboración y precisión posterior de las mismas y, sobre todo, por la enorme influencia que ejercieron, corresponde sin duda a Kuhn el mayor protagonismo en el surgimiento de esta nueva concepción. En esta obra se tratan prácticamente todos los temas fundamentales de la filosofía de la ciencia y todos ellos bajo una perspectiva nueva. Nos ceñiremos aquí a los relativos a la estructura de los constructos teóricos o, como Kuhn los denomina inicialmente, *paradigmas.*

2.1. CIENCIA NORMAL Y CIENCIA REVOLUCIONARIA

En las ciencias maduras, Kuhn distingue dos modos de "hacer ciencia" que además, en su opinión, se suceden históricamente. Al primero lo llama *normal* pues es el modo usual en que opera la ciencia, la manera en que ésta se desarrolla la mayor parte del tiempo. Al segundo lo denomina, por oposición, *no-normal o extraordinario y,* a veces, *revolucionario.* Es importante insistir en que éste es el modo en que, según Kuhn, procede la ciencia madura, pues el panorama que vamos a trazar no se aplica a los períodos de formación o asentamiento de una disciplina.

Los períodos de ciencia normal se caracterizan por el hecho de que la comunidad de científicos que trabaja en un determinado ámbito comparten ciertos presupuestos de muy diverso tipo (teóricos, experimentales, metodológicos y otros) que son los que les permiten *ir haciendo ciencia.* Estos elementos compartidos se encuentran, implícitamente unos, explícitamente otros, en los canales usuales de enseñanza y transmisión de una disciplina (principalmente los libros de texto) y el futuro científico los adquiere por regla general en su período de aprendizaje. En ciencia normal la tarea casi exclusiva consiste en lo que Kuhn llama *trabajo de resolución de enigmas o rompecabezas.* Esta tarea consiste, *grosso modo,* en ir ampliando y perfeccionando la aplicación del aparato teórico-conceptual a la experiencia, y a la vez y como consecuencia de ello, en ir ajustando y puliendo la base teórico-conceptual. Algunas de las tareas típicas de la investigación normal son la precisión de constantes ya conocidas, la determinación de otras nuevas, encontrar formas específicas de leyes generales y aplicar las ya disponibles a nuevos fenómenos. Para llevar a cabo este trabajo es esencial que el científico no cuestione los supuestos compartidos, pues son precisamente ellos los que guían su investigación y le permiten abrigar esperanzas de éxito. La ciencia normal no discute sobre fundamentos ni "tiende hacia novedades fácticas o teóricas y, cuando tiene éxito, no descubre ninguna" (1962-1970, p. 43).

Ahora bien, la ciencia normal es sólo *un* modo en que se desarrolla la empresa científica. La ciencia (madura) no discurre siempre de este modo. Un tipo importante de enigmas tiene que ver con la presencia de *anomalías,* experiencias que "no encajan" en el aparato teórico. Aunque a menudo se resuelven con éxito, a veces algunas anomalías (o, más raramente, algún otro tipo de enigma) se muestran recalcitrantes. Si ello ocurre con varias, o con alguna considerada especialmente importante, puede ocurrir que, tras cierto tiempo, algunos miembros de la comunidad desesperen de encontrar una solución, o que, aunque la encuentren, consideren excesivas las modificaciones *normales* a que obliga. Cuando este sentimiento se generaliza en la comunidad científica sobreviene lo que Kuhn llama una *crisis:* se comienzan a cuestionar los supuestos que guiaban la investigación, se pierde la confianza en ellos y se empiezan a revisar y a discutir los fundamentos. En estos períodos de crisis se suceden propuestas alternativas hasta que en torno a alguna de ellas se comienza a organizar un nuevo cuerpo de supuestos desde los que mirar las viejas cosas de un modo nuevo y más prometedor. Con el tiempo, y si el trabajo basado en los nuevos supuestos permite abrigar esperanzas de éxito, reciben la confianza de los especialistas de la comunidad y acaban suplantando a los antiguos como guía para la investigación. Los viejos supuestos son desplazados por los nuevos consumándose lo que Kuhn llama una *revolución científica,* tras la cual se inicia un nuevo período de ciencia normal.

Éste es el tipo de actividad que caracteriza la ciencia *no-normal o extraordinaria,* la que se desarrolla en los períodos revolucionarios. A ella están asociados los "grandes nombres" de la historia de la ciencia, como Copérnico, Newton, Darwin o Einstein. Pero es importante señalar que no sólo ellos pues, si bien la ciencia extraordinaria es un fenómeno mucho más extraño que la ciencia normal, según Kuhn se da con más frecuencia de la que la referencia a estas grandes revoluciones puede sugerir (cf. 1969, p. 149; el "tamaño" de las revoluciones es una de las cuestiones que Kuhn nunca ha precisado suficientemente). Es importante señalar que el paso de un período normal a otro no viene obligado por necesidad lógica. Se trata de un desplazamiento de confianza y, en ausencia de un nuevo programa, el antiguo puede mantenerse largo tiempo aunque haya entrado en crisis. Recuérdese que ésta es la situación en disciplinas científicas ya asentadas. En los períodos de gestación no hay un paradigma dominante y lo que sucede es algo muy parecido a lo que sucede en los períodos de crisis en la ciencia madura, a saber, una extraordinaria proliferación de alternativas rivales que compiten por imponerse en la comunidad.

2.2. PARADIGMAS *QUA* MATRICES DISCIPLINARES

El concepto básico que articula esta nueva concepción de la ciencia es el de *paradigma.* Un napaSEtyµa (del griego `et(xpa': cercano, aproximado; y `Ssty ta': muestra, mostración) es un ejemplo o caso de algo que hace de modelo para otros casos de lo mismo, es un ejemplo-tipo o típico. Así decimos, por ejemplo, que María Callas es un paradigma de cantante de ópera, que es una cantante de ópera paradigmática; o que Romeo y Julieta son un paradigma de amantes apasionados; o, el ejemplo preferido del propio

Kuhn, que *amo-amas-amat-amamus-amatis-amant* es un paradigma de la conjugación del verbo latino. Este significado original del término `paradigma' se desplaza en *La Estructura* en varias direcciones hasta llegar a tener sentidos muy diferentes (cf. p.ej. Masterman, 1970, para un análisis de los mismos). De entre ellos, el dominante, el que sustenta esta concepción, es desafortunadamente el más impreciso de la obra. En este sentido un paradigma es el conjunto de supuestos compartidos por una comunidad que guían su investigación normal. La ciencia normal es *ciencia-basada-en-(un-)paradigma y* la ciencia extraordinaria o revolucionaria es el paso de un paradigma a otro. En esta última, al igual que en la fase inmadura o preparadigmática de una disciplina, se trabaja sin (el dominio de *un)* paradigma, hay una proliferación de hipótesis diferentes. Las disciplinas maduras, aquellas en que ha surgido ya un primer paradigma, se desarrollan *de paradigma en paradigma a través de revoluciones.*

En este primer trabajo, Kuhn no es lo suficientemente explícito acerca de estos supuestos compartidos como para extraer una idea clara de los mismos. Muchas de las críticas que se le dirigieron inicialmente no sólo se dirigían contra la equivocidad del término sino también, y principalmente, contra la vaguedad de éste su sentido preponderante. Parecía que hablar de paradigmas no era sino otro modo, poco afortunado, de referirse en términos muy generales a teorías. En trabajos posteriores (cf. especialmente 1962-1970, "Postscriptum", y 1970c), Kuhn intenta distinguir y precisar los diferentes sentidos con que introdujo el término. Los diversos usos que de él hacía en su primera obra los reagrupa ahora en dos sentidos principales. El primero es global y comprende todos los compromisos compartidos por un grupo científico, la completa constelación de creencias, valores, técnicas y demás elementos compartidos por los miembros de una comunidad científica dada. Como veremos, el segundo es concreto y denota un componente específico de lo anterior, un tipo especialmente importante de tales compromisos. Aunque entre los estudiosos de la ciencia el término ha acabado por usarse en el sentido global, fue el segundo el que motivó originalmente su introducción y el que se adecúa a su significado etimológico. Para no confundirlos, Kuhn denomina en estos trabajos `matriz disciplinar' *('disciplinary matrix')* a lo primero y `ejemplar' a lo segundo.

Un paradigma *qua* matriz disciplinar es por tanto lo compartido por una comunidad científica, lo que guía en un momento dado su investigación normal. Sus principales componentes son los siguientes:

Generalizaciones simbólicas

Éste es el componente formal o fácilmente formalizable de la matriz disciplinar y comprende, aproximadamente, las tradicionales *leyes. A* menudo se encuentran ya en forma simbólica, como $f = ma$, $I = V/R$ o $\nabla^2\psi + 8\pi^2 m/h^2(E - V)\psi = 0$; pero también pueden venir expresadas en palabras, como "la acción es igual a la reacción" o "la combinación química se produce según proporciones constantes de peso". Estas generalizaciones simbólicas, consideradas aisladamente, funcionan como expresiones de un sistema matemático puro de uso compartido por los miembros de una comunidad científica; como mero formalismo abstracto, son expresiones vacías de significado o aplicación empírica.

No todas las generalizaciones simbólicas son paradigmáticas, esto es, no todas se consideran incuestionables. Es típico que así ocurra con las que tienen cierto carácter fundacional o programático. De entre ellas, son especialmente importantes las "más generales", cuasi-vacías o cuasi-tautológicas, como $f = ma$ o la ecuación de onda de Schródinger, que más que generalizaciones son esquemas de tales: "no son tanto generalizaciones como esquemas de generalizaciones, formas esquemáticas cuya expresión simbólica detallada cambia de una aplicación a la siguiente" (1970c, §I1I). Una de las tareas de la ciencia normal consiste precisamente en intentar aplicarlas a situaciones empíricas concretas encontrando formas especiales de las mismas: "En el problema de la caída libre, $f = ma$ pasa a ser $mg = md^2s/dt^2$. Para el péndulo simple se convierte en mg seno $= - md^2s/dt^2$. Para osciladores armónicos acoplados, la mencionada fórmula se convierte en dos ecuaciones, la primera de las cuales puede escribirse m,d^{2}s,/dt^2 + $k,s, = k2(d +S2 - s,)$. Problemas mecánicos de mayor interés, por ejemplo el movimiento de un giroscopio, mostrarían aún mayor disparidad entre $f = ma$ y la generalización simbólica a la que efectivamente se aplica la lógica y la matemática" (¡bid.). Es en este sentido que $f = ma$ no es tanto una generalización específica cuanto un esquema que va adquiriendo formas específicas para casos de aplicación específicos. Y por eso es, *considerada en sí misma,* cuasi-vacía o cuasi-tautológica y, por tanto, difícilmente refutable, difícilmente puede entrar en conflicto con la experiencia. Por sí sola apenas tiene "contenido", son sus versiones específicas las que lo tienen y las que entran en conflicto con la experiencia. Pero si ello ocurre, siempre es posible mantener la ley más general y retocar sólo sus desarrollos específicos. La idea es que tales leyes generales son "programáticas", son algo así como "guías para la investigación": si tienes un fenómeno cinemático a explicar, busca fuerzas responsables del mismo de modo que la suma de todas ellas sea igual al producto de la masa por la aceleración; si la suma de fuerzas no coincide con dicho valor, la conclusión no es que la segunda ley es falsa, sino que debes seguir buscando nuevas fuerzas o precisar mejor la naturaleza y magnitud de las ya detectadas. En este sentido, este tipo de generalizaciones son "irrefutables" y sólo sus versiones específicas entran en conflicto con la experiencia. Su abandono es un fenómeno *revolucionario.* Durante los períodos de ciencia normal no se cuestionan, sólo se cuestionan en los momentos de crisis y si se terminan abandonando es porque han perdido la confianza de la comunidad como principios que guían la investigación. Las revoluciones entrañan, entre otras cosas, el abandono de estos principios, de estas *leyes paradigmáticas,* pero como parte del proceso general de pérdida de confianza en el paradigma-matriz en crisis (sobre la forma lógica de estos *principios guía,* cf. Moulines, 1982, cap. 2.3).

Modelos

Kuhn usa aquí `modelo' en el sentido de imagen, algo a lo que se puede asimilar otra cosa, por ejemplo cuando decimos que un computador es un modelo de la mente. Los modelos proporcionan al grupo las analogías preferidas o, si se las sostiene a fondo, una ontología. En un primer sentido, los modelos son simples analogías, son sólo *heurísticos,* por ejemplo la asimilación del comportamiento de un gas con el de un conjunto de pequeñas bolas en movimiento, o del funcionamiento de la mente con el de un computador. En un segundo sentido, más fuerte, los modelos son objeto de compromiso metafísico, son *ontológicos,* por ejemplo la creencia de que todo fenómeno perceptible es debido al movimiento e interacción de átomos en el vacío. Kuhn admite que ambos tipos de modelos son conceptualmente diferentes, pero los subsume en un mismo grupo de compromisos porque su función metodológica y epistémica es muy parecida (cf. 1962-1970, "Postscriptum", n. 9). Además de proporcionar a la comunidad científica sus analogías preferidas, muchas veces determinan qué puede ser aceptado como solución a un problema. Por otro lado, a veces modelos heurísticos pueden pasar a convertirse en ontológicos, como ocurrió en la reducción de la termodinámica a la mecánica estadística con la asimilación del calor con la energía cinética media de las moléculas. Kuhn enfatiza que, aunque usualmente los miembros de una comunidad comparten los modelos, ello no es esencial, pueden no compartirlos, ni siquiera los heurísticos. En este punto, sin embargo, no está claro qué grado de "esencialidad" tienen estos componentes de las matrices (cf. 1962-1970, pp. 151-152).

Valores

Los valores son el conjunto de criterios axiológicos que emplea la comunidad al evaluar su propia actividad. Los más destacados son los relativos a la no vaguedad de las predicciones, el margen de error admisible de las observaciones respecto de las predicciones, la fecundidad, coherencia y simplicidad del aparato teórico y la compatibilidad con otras teorías aceptadas. A veces también se contemplan otros más externos relacionados, por ejemplo, con la utilidad de la ciencia o su función social. Los valores operan también en la ciencia normal, pero juegan su principal papel en el surgimiento de las crisis y en su resolución, en la elección de paradigmas alternativos. Generalmente estos valores son compartidos por varias comunidades dentro de una misma disciplina, pero no por ello tienen siempre el mismo efecto. Puesto que su aplicación conjunta produce conflictos al no ser plenamente compatibles entre sí, es forzoso a veces conceder más importancia a unos que a otros, y diferentes comunidades, o la misma en momentos diferentes, pueden hacerlo de diferente modo. Ésta es una de las razones por las que no hay un procedimiento mecánico que nos diga cuándo un paradigma debe ser abandonado, se le debe retirar la confianza, o qué elección hacer entre paradigmas alternativos. Nótese que, en las revoluciones, una opción, el viejo paradigma, está muy desarrollada y otra, el nuevo, poco, por lo que no está claro cómo contrapesar los resultados de las exigencias que impone el mismo grupo de valores. Si concedemos mucha importancia a la fecundidad, el nuevo paradigma incipiente sale de momento perdiendo, pero si se la concedemos a la resolución de problemas recalcitrantes, saldrá ganando. El proceso de decisión de acuerdo con esos valores no es pues automático o mecanizable, y a veces puede depender esencialmente de otros elementos externos a la actividad científica, elementos de carácter social, o económico, o incluso político o ideológico.

Ejemplares

Éste es el componente más importante, junto con las generalizaciones simbólicas, de la matriz. A él se refiere el otro sentido de `paradigma' anunciado, sentido que motiva originalmente la introducción del término. Los ahora llamados `ejemplares' son paradigmas en sentido etimológico: casos que hacen de modelo, ejemplos modélicos. Los ejemplares son aplicaciones empíricas específicas del aparato formal que sirven de modelo o guía para el trabajo de resolución de rompecabezas, para otras aplicaciones; son las "partes de la realidad" a las que típicamente se aplica el formalismo. Pueden ser logros especialmente importantes de la teoría, como la aplicación al sistema solar de la mecánica newtoniana, o la aplicación al cometa Halley de esa misma teoría, o la aplicación al perihelio de Mercurio de la mecánica relativista, etc. Otros son ejemplos-tipo, típicos, para su aplicación, como una experiencia de laboratorio con un plano inclinado, o un problema-resuelto de un libro de texto. A ellos se refiere Kuhn en *La Estructura* cuando afirma que los paradigmas son "realizaciones científicas universalmente reconocidas que, durante cierto tiempo, proporcionan modelos de problemas y soluciones a una comunidad científica" (p. ix).

Mediante los ejemplares se ven situaciones nuevas como semejantes a otras anteriores bien establecidas. Se ven fenómenos diferentes de modo similar, como casos de aplicación de una misma ley; por ejemplo se puede ver el movimiento de un péndulo como semejante, con ciertas idealizaciones, al de una bola moviéndose en un doble plano inclinado, o puede verse, como propusieron los copernicanos tras el descubrimiento de las lunas de Júpiter, el sistema Sol-planetas como semejante al sistema Júpiter-lunas. También aquí no todos estos elementos son considerados igualmente esenciales, no todos los casos de aplicación son igualmente importantes, de entre ellos sólo algunos son considerados *paradigmáticos*. Al hacer de modelos-paradigmas para la resolución de enigmas, los ejemplares guían, junto con las leyes paradigmáticas, la investigación normal, el desarrollo de la matriz disciplinar. En gran medida, la ciencia normal consiste en ir ampliando con éxito el ámbito de situaciones semejantes a los ejemplares, intento que obliga generalmente a alguna modificación de las leyes más específicas (no paradigmáticas).

Al presentar el primer elemento de las matrices, las generalizaciones simbólicas, vimos que Kuhn enfatiza que por sí solas son simples componentes de un formalismo abstracto vacío de contenido empírico. Pues bien, según Kuhn es justamente a través de los ejemplares como, al menos en parte, se cargan de contenido empírico los términos de las generalizaciones que constituyen el formalismo abstracto. Con los ejemplares se aprende cómo el aparato conceptual se aplica a la naturaleza y, en consecuencia, parte de su significado. Los ejemplares desempeñan ahora (al menos parte de) el papel de las antiguas reglas de correspondencia. En la sección en que se ocupa de la conexión de la teoría con la experiencia, Kuhn afirma explícitamente que la "habilidad adquirida para ver semejanzas entre problemas aparentemente dispares desempeña en las ciencias una parte importante del papel atribuido corrientemente a las reglas de correspondencia" (1970c, §IV); a esta idea se refería Hempel cuando, como mencionamos en el capítulo anterior, señala a Kuhn como una de las fuentes de sus nuevas tesis sobre el modo en que los términos teóricos se cargan de contenido empírico. Por otro lado, los ejemplares no son *experiencias puras,* descripciones neutras de la naturaleza. Son ejemplares *dentro* de un paradigma y están en parte ya conceptualizados. Ello hace que, al cambiar el paradigma, con *todos* sus componentes, cambie, según Kuhn, al menos parte del significado de los términos y a su vez el modo-de-ver-guiado-por-ejemplares las cosas. Las "mismas" situaciones se ven de modo diferente y quienes mantienen paradigmas diferentes viven, en cierto sentido, en mundos diferentes. Éste es el origen del fenómeno de la *inconmensurabilidad* que en opinión de Kuhn acompaña los cambios revolucionarios.

Esta nueva concepción de los constructos teóricos tiene importantes consecuencias para las más importantes cuestiones epistemológicas y semánticas de la filosofía de la ciencia, como el ya mencionado de la inconmesurabilidad, o los del relativismo, la racionalidad, el significado de los términos científicos, la confirmación y la falsación, etc. Por ahora nos limitamos a los elementos fundamentales de las matrices disciplinares y más adelante, cuando tratemos algunas de estas cuestiones, volveremos sobre el resto de las tesis kuhnianas (cf. especialmente cap. 12, §5, sobre sus consecuencias para el problema de la inmunidad y cap. 13, §4, §5 para las relativas a la inconmensurabilidad). Para concluir comentaremos brevemente un problema de ambigüedad relativo al término 'matriz disciplinar'. Kuhn afirma que introduce dicho término en sustitución del equívoco \`paradigma' usado en *La Estructura de las Revoluciones Científicas,* y que mediante él se quiere referir al conjunto de supuestos compartidos por los miembros de una comunidad científica. Estos supuestos son los supuestos de los cuatro tipos que acabamos de examinar, pero lo que no está claro es si la matriz disciplinar incluye *todos* los supuestos de cada tipo o sólo los *paradigmáticos;* o limitándonos a los dos que más centrarán nuestra atención, no está claro si las matrices incluyen todas las leyes y todos los ejemplares o sólo las leyes programáticas generales y los ejemplares paradigmáticos.

Esta cuestión es en parte sólo nominal. Kuhn afirma que su intención era captar lo que tradicionalmente se ha denominado \`teoría', pero que no usa este término porque tal como de hecho lo emplean los científicos "connota estructuras mucho más limitadas en naturaleza y alcance que las requeridas [...]; para los presentes propósitos sugiero \`matriz disciplinar': \`disciplinar' porque se refiere a la posesión común de los practicantes de una disciplina; \`matriz' porque está compuesta de elementos ordenados de varios tipos" (1962-1970, p. 150). No aclara sin embargo si la matriz incluye todos los elementos de cada tipo o sólo algunos, y cuando usa el término unas veces parece referirse a todos y otras sólo a los paradigmáticos. En principio, sus numerosas referencias al "conjunto total de supuestos compartidos" parecen sugerir que incluye todos, pero en realidad ello no es inmediato pues depende de qué se entienda por \`comunidad científica'. Si la comunidad científica es el conjunto de científicos que trabajan en una teoría *en un momento dado,* entonces la matriz incluye *todos los supuestos compartidos en ese estadio de la teoría.* Pero si la comunidad es el conjunto de los científicos que trabajan en la teoría *en toda su historia,* entonces la matriz incluye *todos los supuestos compartidos a lo largo de toda la historia de la teoría,* que serán sólo los paradigmáticos pues, como veremos (cf. cap. 13), éstos son los que se preservan a lo largo de la historia de la teoría. La presente ambigüedad no es, pues, sino consecuencia de la ambigüedad entre los sentidos sincrónico y diacrónico de `teoría' o `matriz disciplinar'. Desde una perspectiva sincrónica, la matriz disciplinar incluye todos los elementos compartidos en un momento dado; desde la perspectiva diacrónica, incluye sólo los que perduran durante la historia de la teoría. Una vez esto queda claro, qué palabras usemos es lo de menos. Cuando en adelante usemos esta expresión, el contexto aclarará el sentido en que se hace, en caso contrario lo aclararemos explícitamente.

3. Los programas de investigación de Lakatos

Lakatos, inicialmente discípulo de Popper, reacciona contra él llegando a puntos de vista similares a los de Kuhn y Feyerabend aunque sin caer, al menos eso pretende, en algunas tesis extremas de éstos, sobre todo las relativas a la racionalidad científica, que considera inaceptables, desde una perspectiva popperiana general que comparte. Lakatos lleva a cabo la revisión del falsacionismo de Popper en una serie de estudios sobre lo que denomina *la metodología de los programas de investigación científica.* Estos trabajos iban a concluir con una extensa monografía titulada, en referencia directa a la obra de Popper, *The Changing Logic of Scientific Discovery,* proyecto que quedó truncado al morir su autor. Una de las finalidades de esta obra era desarrollar la estructura final de los *programas de investigación,* noción con la que Lakatos pretendía recoger lo que consideraba fundamental de los constructos teóricos. Al quedar el proyecto sin conclusión, sólo disponemos, como fuente de sus ideas sobre la estructura de los programas de investigación, de sus primeros trabajos (especialmente 1968b y 1970). El propio Lakatos los considera muy insuficientes y provisionales, pero contienen ya algunos elementos de interés para nuestro actual tema. Veremos aquí muy brevemente las principales contribuciones, que casi siempre están sólo esbozadas.

Lakatos parte de las observaciones de Popper sobre el *conocimiento de fondo y* la contrastación y las lleva a sus últimas consecuencias. Lo que se evalúa en la contrastación, dice, no es *una* teoría comparada con los hechos sino un conjunto de (mini)teorías, de diferente estatus metodológico, comparadas entre sí: "el conflicto no es entre `teorías y hechos', sino entre una *teoría interpretativa* que provee de hechos y una *teoría explicativa* que los explica [...], no se trata de que nosotros propongamos una teoría y la Naturaleza pueda gritar NO, sino que nosotros proponemos una red de teorías y la Naturaleza puede gritar INCONSISTENTES" (1970, p. 130). *Este* conflicto se intenta resolver modificando algunos elementos de la red y se genera así una sucesión de teorías-redes ligadas por "una notable continuidad". Esta serie o sucesión de teorías es lo que Lakatos llama un `programa de investigación' que, como reconoce explícitamente, tiene fuertes reminiscencias de la ciencia normal de Kuhn.

Lakatos presenta los elementos constituyentes de los programas de investigación en el contexto de las heurísticas que caracterizan la metodología de los programas. Todos los programas tienen un *núcleo* que los vertebra y les confiere unidad. Este núcleo lleva asociada una heurística que determina dos tipos de reglas metodológicas: unas nos dicen qué senderos de investigación hemos de evitar, *heurística negativa, y* otras qué senderos hemos de seguir, *heurística positiva.* La heurística negativa prohibe, *por decisión,* aplicar la refutación al núcleo, para lo cual se debe articular un *cinturón protector* de hipótesis auxiliares o complementarias que sí se consideran modificables. La heurística positiva sugiere cómo modificar y desarrollar esta parte "refutable" del programa. Éstas son las líneas maestras de la nueva metodología de Lakatos, contenida sucintamente en el siguiente pasaje:

"Todos los programas de investigación científica se pueden caracterizar por su `núcleo'. La heurística negativa del programa nos prohibe dirigir el *modus tollens a* este `núcleo'. En lugar de ello debemos emplear nuestro ingenio en articular o incluso inventar `hipótesis auxiliares' que conformen un *cinturón protector* en torno a ese núcleo, y es a éstas a quienes debemos dirigir el *modus tollens.* Es ese cinturón protector de hipótesis auxiliares quien tiene que resistir el peso de las contrastaciones e irse ajustando y reajustando, o incluso ser sustituido por completo, para defender al núcleo, que de ese modo se hace más sólido. [...] Este núcleo es `irrefutable' por decisión metodológica de sus protagonistas: las anomalías sólo deben llevar a cambios en el cinturón protector. [...] La heurística negativa especifica el núcleo del programa que es `irrefutable' por decisión metodológica; la heurística positiva consiste en un conjunto parcialmente articulado de sugerencias o indicaciones sobre cómo cambiar y desarrollar las `variantes refutables' del programa de investigación, cómo modificar, sofisticar, el cinturón `refutable' de protección" *(1970, §3a-b).*

El resultado de aplicar esta metodología constituye la evolución de una teoría científica; en términos de Lakatos, se trata de una sucesión de diferentes versiones del mismo programa, esto es, en torno a un mismo núcleo. Como se habrá advertido, esta imagen es similar a la evolución de un paradigma kuhniano. Lakatos ofrece además una tipología de los programas de investigación en función de su mayor o menor "éxito". Un programa es *progresivo* si (entre otras cosas en las que no podemos detenernos ahora) predice hechos que se constatan después, y es, o está, *estancado* si sólo "postdice", esto es, si sólo ofrece explicaciones *ad hoc* de hechos (para él) inesperados. Esto exige dos cualificaciones. En primer lugar, el juicio requiere cierta perspectiva histórica, esto es, a los programas incipientes es racional "concederles cierto tiempo". Por otro lado, e incluso garantizada la perspectiva histórica, las cosas no siempre están tan claras, los casos mencionados son más bien idealizaciones y hay numerosos casos intermedios. Como es usual en los filósofos de esta orientación, Lakatos ilustra su nueva concepción con un detallado estudio histórico, en este caso del programa de investigación de Prout centrado en la idea de que todos los átomos están compuestos de átomos de hidrógeno, y del programa de investigación de Bohr centrado en la tesis de que la emisión de luz se debe al salto de órbita de los electrones en el interior del átomo. Es importante señalar que esta tipología idealizada de programas no debe tomarse como un criterio cuasiformal de sustitución: nada *obliga, y* por supuesto la lógica más "los hechos" tampoco, a abandonar un programa estancado, aunque sólo sea porque siempre es posible su "resurrección", esto es, de todo programa estancado siempre es en principio posible que se convierta de nuevo en uno progresivo.

Esta perspectiva, bastante kuhniana en general, se distancia de la del autor de *La Estructura* en algunos puntos, especialmente los que tienen que ver con los problemas de la inconmensurabilidad y la racionalidad de la ciencia y en los que no podemos detenernos ahora. Por lo que respecta a la estructura de las teorías, la diferencia más señalada, aparte de la extrema imprecisión de las nociones de *núcleo y cinturón protector,* tiene que ver con el "alcance" o "dimensión" de los programas de investigación.

En Kuhn está (aproximadamente) claro que los paradigmas-matrices se identifican, diacrónicamente, con teorías a lo. largo de la historia. La astronomía geocéntrica constituye un paradigma, la heliocéntrica otro diferente; la mecánica desde Newton a Lagrange constituye un paradigma, la mecánica relativista otro diferente; la química del flogisto constituye un paradigma, la química del oxígeno otro diferente; etc. Sin embargo, y quizá como consecuencia de la mencionada imprecisión, no está claro en Lakatos cuáles son los límites de los programas de investigación. En general parece que engloban, como en Kuhn, sólo teorías extendidas en el tiempo; así se refiere, por ejemplo, "al programa de Newton". Pero en ocasiones habla como si varias de esas teorías pudieran constituir una única tradición, que se identificaría entonces con toda una disciplina. En ausencia de mayores precisiones, no es posible entonces distinguir claramente entre *a)* el paso de una versión a otra de la misma teoría, *y b)* el paso de una teoría a otra. Estos dos fenómenos diacrónicos son *prima facie* esencialmente diferentes (como veremos con detenimiento en el capítulo 13) y poder dar cuenta de esta diferencia depende de cómo se caractericen los constructos teóricos en su dimensión sincrónica. Probablemente Lakatos no consideraba que la diferencia de la que se debe dar cuenta sea nítida, pues incluso llega a afirmar que "la ciencia como un todo puede considerarse un inmenso programa de investigación" *(ibid.,* §3). Pero entonces la imprecisión de sus nociones centrales, *programa de investigación, núcleo y cinturón protector,* es deliberada y no se ve claramente cómo es coherente con la aplicación histórica que realiza sobre casos concretos. Seguramente esta cuestión era una de las que quería clarificar en la monografía que preparaba y que no pudo concluir.

4. Las tradiciones de investigación de Laudan

Laudan no pertenece generacionalmente al grupo de los nuevos filósofos de la ciencia, pero su obra, bastante posterior, presenta grandes afinidades, tanto en intereses como en contenidos, con la de aquéllos. Desde finales de los años setenta, Laudan publica una serie de estudios donde revisa las principales tesis acerca de la naturaleza de los constructos teóricos, los valores que rigen la actividad científica y la noción de progreso adecuada a ellos. Como en el caso de Kuhn, el ámbito de estudio es extraordinariamente amplio, somete a revisión crítica prácticamente todas las cuestiones centrales de la filosofía de la ciencia. En relación a nuestro actual tema, el concepto básico que articula su concepción de los constructos teóricos es el de *tradición de investigación,* presentado por primera vez de forma sistemática en su monografía de 1977. Aunque, como los paradigmas de Kuhn, estas entidades se caracterizan diacrónicamente, nos centraremos ahora en los aspectos sincrónicos que tal caracterización diacrónica presupone.

Laudan presenta la noción de *tradición de investigación* en relación explícita con los paradigmas de Kuhn y los programas de investigación de Lakatos, intentando preservar lo que considera adecuado de ellos y modificando los aspectos que considera insatisfactorios (cf. Laudan, 1977, cap. 3). Laudan comienza distinguiendo dos sentidos del término `teoría científica', dos tipos de "redes proposicionales". En primer lugar, el término puede denotar un conjunto relativamente específico de doctrinas, leyes, hipótesis o principios relacionados, que se usan para hacer predicciones experimentales y ofrecer explicaciones de fenómenos naturales. Ejemplos de ello son la teoría newtoniana de la luz, el electromagnetismo de Maxwell, la teoría de la estructura atómica de Bohr, la del efecto fotoeléctrico de Einstein o la de la plusvalía de Marx. Además de este sentido, el término se usa también para referirse a conjuntos de doctrinas o supuestos "mucho más generales y mucho menos fácilmente corroborables empíricamente" (p. 71). Ejemplos de ello son la teoría de la evolución, la teoría atómica o la teoría cinética de los gases. Aunque no siempre es claro al respecto, parece que las teorías en este segundo sentido consisten, al menos, en familias enteras de teorías en el primer sentido vinculadas por principios metodológicos u ontológicos muy generales. Kuhn y Lakatos, afirma, han mostrado que son las teorías en el segundo sentido las unidades en que se debe centrar el estudio de la actividad científica, la "herramienta primaria para la comprensión y valoración del progreso científico [... yo] comparto esa opinión, pero encuentro que las explicaciones dadas hasta ahora de lo que son estas teorías más amplias y de cómo evolucionan, no son completamente satisfactorias" (p. 72). De estas teorías generales, en el segundo sentido del término, es de lo que pretende dar cuenta su noción de *tradición de investigación.* Veamos, sin detenernos en las críticas específicas que hace a Kuhn y Lakatos, cuáles son los principales elementos que caracterizan a estas tradiciones de investigación.

Supuestos compartidos

Las tradiciones constan de dos tipos de supuestos generales, que individualizan una tradición dada y la distingue de otras:

(i) *Compromisos metafísicos.* Conjunto de creencias acerca de qué tipos de entidades y procesos constituyen el dominio de investigación; por ejemplo, en física, las creencias asociadas a las teorías atomistas, o contrariamente, las asociadas a las teorías de campos.

(ii) *Normas epistémicas y metodológicas.* Normas acerca de cómo tiene que investigarse el dominio, cuál es el conocimiento de fondo intocable, cómo han de someterse a prueba las hipótesis, cómo han de recogerse los datos, cómo han de evaluarse la solución a los problemas, etc.

Conjuntamente, los compromisos metafísicos y las normas epistémicas y metodológicas proporcionan a la tradición una *heurística,* orientaciones para la investigación, y una *axiología,* normas de evaluación.

Articulación teórica

Las tradiciones poseen un cierto número de *teorías específicas asociadas* que las ejemplifican y las constituyen parcialmente. Son los elementos empíricamente contrastables de la tradición, el "lugar" donde se contrasta la tradición con la experiencia.

Resolución de problemas

La finalidad principal de las tradiciones, en relación a la cual se evalúan globalmente, es la *resolución de problemas.* Los problemas son de dos tipos:

(i) Problemas empíricos. Derivados de la aplicación de las teorías específicas al dominio empírico de investigación. Estos problemas pueden ser (estar): *resueltos, los* casos de aplicación al dominio empírico exitosos según los estándares de la tradición; *potenciales, los* casos de aplicación que la tradición considera que deben resolverse, pero todavía no resueltos por la tradición en cuestión ni por ninguna otra; *anómalos,* los casos de aplicación que la tradición considera que deben resolverse, que ella todavía no ha resuelto y que han sido resueltos en otra tradición alternativa.

(ii) Problemas conceptuales. Relativos a la estructuración conceptual de alguna teoría específica. Se dan en los siguientes casos: cuando la teoría es inconsistente; cuando contiene supuestos inaceptablemente ambiguos; cuando algunas de sus hipótesis contravienen otras teorías específicas, o los supuestos metafísicos predominantes; cuando sus afirmaciones no proceden según las doctrinas metodológicas y epistemológicas; cuando no acierta a integrar conceptos y principios de teorías más generales a las que está subordinada.

Desarrollo histórico

Las tradiciones discurren en el tiempo a través de un cierto número de formulaciones. Estas formulaciones son la respuesta en un momento específico a la evaluación negativa sobre la solución dada a alguno o varios de los problemas. El modo más usual en que cambia una tradición es modificando sus teorías específicas, pero ocasionalmente puede cambiar alguno de sus elementos nucleares más básicos.

Coexistencia

Las tradiciones no son "dominantes", no se imponen por períodos. En cierto momento dado, en contra de lo que sugiere Kuhn, la coexistencia de tradiciones de investigación rivales es la regla, y no la excepción.

Éstos son los rasgos más generales de las tradiciones de investigación según Laudan, que como se habrá apreciado tienen mucho en común con los paradigmas kuhnianos y los programas de investigación de Lakatos. Como hemos indicado, sin embargo, aunque esté de acuerdo con la idea general, Laudan discrepa en aspectos que considera centrales. Esto se pone de manifiesto en el desarrollo específico de algunos de estos rasgos generales. Concluiremos comentando brevemente algunos de los elementos más problemáticos en el desarrollo de su programa.

En primer lugar, no está muy definido el "tamaño" de las entidades involucradas, tradiciones de investigación y teorías específicas. Los ejemplos que da de tradiciones, como la teoría de la evolución o la teoría cinética de los gases, sugieren entidades como las matrices kuhnianas, está es, teorías científicas a lo largo de su historia. Pero la caracterización general que da parece más bien sugerir grandes orientaciones científicas de una época (quizá coexistiendo y rivalizando), como el mecanicismo o el vitalismo, que pueden incluir varias matrices disciplinares en una época. Esta fluctuación es pareja a la que se da en las teorías específicas. Algunas veces parece que se trata de leyes bastante específicas, del tipo de la ley de caída libre de Galileo, la de gravitación de Newton, o la de gases ideales. Otras veces parece que se refiere a agregados de tales leyes, como el electromagnetismo de Maxwell o la teoría atómica de Bohr. Seguramente la vaguedad es intencionada y se pretende que puede haber diversas dimensiones en ambos tipos de entidades. El problema es cómo caracterizarlas para que la diferencia entre las más grandes de las teorías específicas y las más pequeñas de las tradiciones de investigación, no sea sólo de grado; ello representa un problema en la medida en que Laudan parece requerir que la diferencia entre teorías y tradiciones sea nítida.

En segundo lugar, y relacionado con lo anterior, no está claro si hay alguna relación formal, y si la hay cuál es, entre tradiciones y teorías específicas. Por un lado, no se clarifica si, como decía Kuhn, de los supuestos generales puede formar parte alguna expresión legaliforme, alguna ley especialmente general que funciona cuasi-definicionalmente y de la cual las leyes específicas son concreciones. Quizá se pudieran incluir cosas de este tipo en los compromisos metafísicos generales, pero Laudan no dice si es así ni, caso de que lo sea, cómo se diferenciarían estos compromisos de los otros. Por otro lado, afirma que la relación entre tradiciones y teorías específicas no es de implicación, las tradiciones no implican sus teorías constitutivas. Al argumentar este punto, enfatiza el hecho de que las tradiciones pueden contener, y usualmente contienen, teorías específicas incompatibles (si teorías incompatibles fuesen implicadas por la tradición a la que pertenecen, las tradiciones serían simplemente inconsistentes). El caso más claro es el de teorías específicas sucesivas en la misma tradición, que se contradicen mutuamente; recuérdese que las teorías específicas hacen afirmaciones empíricas específicas, y por tanto si una da cuenta de un problema del que su versión anterior en la tradición no podía dar cuenta (anomalía), ambas versiones *específicas* sucesivas son inconsistentes. Pero también abundan, según Laudan, los casos en que teorías incompatibles coexisten en la tradición *en el mismo período,* como las teorías ópticas rivales de la tradición cartesiana que defendían, respectivamente, el aumento o disminución de la velocidad de la luz en función de la densidad del medio. Se supone que entonces tenemos uno de los problemas conceptuales a que antes nos referíamos, pero una cosa es que se den a veces problemas de este tipo, y otra que se den tan usualmente y tan drásticamente como Laudan parece sugerir.

Por último, como hemos mencionado más arriba, a veces en la evolución de la tradición pueden modificarse algunos de los supuestos básicos. Eso es en principio sorprendente dentro de su propio enfoque, pues Laudan sostiene a la vez que tales supuestos básicos identifican la tradición, por lo que parece que pueden perder (al menos parte de) su identidad y seguir siendo ellas mismas (!). Una vez más, este efecto es querido por Laudan, que discrepa en este punto de Kuhn y Lakatos sobre la persistencia de la esencia o núcleo a lo largo de la tradición. La esencia, afirma, está relativizada respecto al tiempo. El núcleo muy raramente permanece constante a lo largo de toda la evolución de la tradición: "difícilmente hay un conjunto interesante de doctrinas que caracterice las tradiciones de investigación a lo largo de *toda* su historia" *(op. cit.,* cap. 3, §6). Es cierto que existe una continuidad, pero se trata de una "continuidad *relativa* entre etapas *sucesivas* del proceso evolutivo" (¡bid.). Es importante enfatizar que, aun restringida a estadios sucesivos, la continuidad debe ser sólo aproximada, no absoluta, en cada paso se pueden modificar parte de los supuestos básicos; si en cada paso se debieran conservar todos los supuestos básicos, entonces en el conjunto del proceso se seguirían conservando.

Pero si parte de la esencia se puede abandonar, la pregunta que surge inmediatamente es, ¿qué y cuánto se puede cambiar del núcleo sin abandonar la tradición?, ¿qué distingue *a)* el paso de un estadio a otro de la misma tradición de *b)* el paso del estadio terminal de una tradición al inicial de otra tradición nueva diferente?: "Una respuesta parcial a esta pregunta proviene del reconocimiento de que, en *cualquier momento dado,* determinados elementos de una tradición de investigación son más medulares y están más alojados en la tradición de investigación que otros. [...) Ciertos elementos son sagrados y no pueden rechazarse sin repudiar la tradición misma. Pero, a diferencia de Lakatos, quiero insistir en que *el conjunto de elementos de esta clase (los irrechazables) cambia con el tiempo" (ibid.).* Pero esto no parece ser una respuesta ni siquiera parcial, pues si el conjunto de elementos irrechazables cambia con el tiempo, ¿en virtud de qué dos conjuntos diferentes sucesivos se consideran correspondientes a la misma tradición o tradiciones sucesivas diferentes?

Sólo hay dos modos en que, como pretende Laudan, pueda no conservarse nada en el transcurso de una tradición. Uno es que en cada paso no haya un conjunto bien definido de elementos irrechazables, pudiendo ser rechazado cualquiera de ellos mientras se preserve la inmensa mayoría; así, tras una serie de pasos puede que no permanezca ninguno de los originales. Otro es que sí haya un conjunto bien definido de elementos irrechazables, pero que ese conjunto cambie parcialmente en cada estadio; también así puede que tras varios pasos no permanezca ninguno de los elementos originales. Pero si no se dice nada más, en ambos casos no está clara la diferencia entre el cambio *en* una tradición y el cambio *de* tradición. Quizá sea sólo una cuestión de grado, o de convención en la reconstrucción histórica. Pero si no se aceptan estas consecuencias, la única posibilidad es que haya *algo* que se mantenga a lo largo de toda la tradición y cuyo cambio sea precisamente el indicio de un cambio de tradición. Esto es, que las teorías científicas, aun cuando cambien en el tiempo, tengan un núcleo persistente.

5. Consideraciones finales

La incidencia de los *nuevos filósofos* de la ciencia, y otros afines, en nuestra disciplina fue decisiva. La irrupción de la perspectiva historicista que en general les caracteriza marca definitivamente el desarrollo de la reflexión metacientífica posterior. La influencia más determinante afecta quizá a cuestiones como la importancia de los estudios históricos y de los determinantes sociales, la cuestión de la carga teórica de los hechos y el problema de la inconmensurabilidad, los problemas del progreso y la racionalidad en la ciencia, o del relativismo. Sin embargo, a la mayoría de sus tesis subyace, sin implicarlas estrictamente, una nueva visión de la naturaleza y estructura de las teorías científicas, más ajustada a la realidad y más fiel a las teorías tal como la historia nos las presenta. En nuestra opinión, y sin desmerecer sus otras aportaciones, es en esta nueva noción, aunque muy imprecisa, de teoría empírica donde radica su mayor contribución a la disciplina. Desde la perspectiva actual, los principales rasgos de esta nueva noción de teoría que está emergiendo son los siguientes. Las teorías en sentido sincrónico:

1. Son entidades sumamente complejas y dúctiles, susceptibles de evolucionar en el tiempo sin perder su identidad. Aunque la idea de que las teorías son entidades que se extienden en el tiempo a través de diferentes estadios no es un descubrimiento de estos filósofos, sí fueron los primeros en dar a ese hecho todo su valor.
2. No son enunciados o secuencias de enunciados y en un sentido propio no pueden calificarse de verdaderas o falsas, aunque con ellas sí se realizan afirmaciones empíricas verdaderas o falsas.
3. Tienen, al menos, un componente formal, las leyes o hipótesis, y otro empírico o aplicativo, los sistemas a que se pretenden aplicar.
4. Cierta parte de cada uno de estos componentes se considera intocable por decisión metodológica (núcleo). Las teorías tienen pues partes "esenciales" y partes "accidentales", en ello radica su ductilidad. El aparato formal se articula en niveles progresivamente cada vez más específicos, que dan cuenta de situaciones empíricas también específicas. A veces se denomina `teoría', en un sentido más restrictivo, a estos desarrollos concretos del formalismo (p.ej., la teoría de la gravitación).
5. Tienen diversos niveles de empiricidad. Parte de la teoría conceptualiza los hechos y parte explica, y se contrasta con, lo así conceptualizado.
6. Es la parte específica, "accidental", del formalismo la que recibe el peso de la contrastación. Ante una contrastación negativa, el núcleo siempre se puede salvaguardar modificando los elementos no nucleares.
7. Llevan asociadas normas, valores, o simplemente indicaciones metodológicas y evaluativas, algunas de ellas fuertemente dependientes del contexto socio-histórico.

La principal deficiencia de esta nueva caracterización es su imprecisión, en ocasiones tan extrema que termina por difuminar casi totalmente lo que parecen intuiciones correctas. El principal motivo de los positivistas para desarrollar una filosofía formal de la ciencia era justamente eludir el discurso metacientífico vago e impreciso. Y gran parte de las polémicas que surgen tras la irrupción de los nuevos filósofos son generadas por la imprecisión y equivocidad de algunas de sus nociones centrales. La mayoría de los filósofos de la ciencia sensibles a esta nueva perspectiva concluyeron que la complejidad y riqueza de los elementos involucrados en ella escapa a cualquier intento de formalización. No sólo las formalizaciones al estilo de la Concepción Heredada son totalmente inadecuadas para expresar estas entidades en toda su complejidad, sino que no parece razonable esperar que cualquier otro procedimiento de análisis formal pueda capturar los elementos mínimos de esta nueva caracterización. Ésta es la moraleja antiformalista que se extendió en muchos ambientes metacientíficos tras la *revuelta historicista*. Como vamos a ver en el próximo capítulo, no en todos. Tras la digestión de los primeros efectos antiformalistas, algunas de las corrientes más recientes en filosofía de la ciencia muestran que al menos parte de los nuevos elementos señalados, los más estructurales, son susceptibles de un razonable análisis y reconstrucción formales.

CAPÍTULO 10

ANÁLISIS SINCRÓNICO DE TEORÍAS III. LAS CONCEPCIONES SEMÁNTICAS: LAS TEORÍAS COMO ENTIDADES MODELOTEÓRICAS

El efecto de la irrupción historicista durante los años sesenta y principios de los setenta fue doble. Por un lado, a su estela se desarrolla toda una rama de los *science studies* (con importantes, aunque puntuales, antecedentes antes de los años sesenta) que se centra en el estudio de los determinantes sociales de la ciencia apoyándose en una considerable investigación empírica. Esta línea de investigación culmina con el asentamiento durante los años ochenta de la sociología de la ciencia como disciplina. Aunque desde este ámbito se han hecho numerosas incursiones en la filosofía de la ciencia, su importancia para el tema actual, la estructura de las teorías, es escasa, pues sus propuestas son sólo negativas, más bien nihilistas: en la práctica científica no existen en realidad entidades identificables que quepa caracterizar, en ningún sentido del término mínimamente preciso y útil para la comprensión de la actividad científica, como "teorías científicas". No vamos a detenernos aquí en estas tesis.

Por otro lado, asimiladas las contribuciones incuestionables de los historicistas y expurgados sus principales excesos, se recupera durante los años setenta la confianza en la viabilidad de los análisis formales o semiformales de la ciencia, al menos en algunos de sus ámbitos, entre ellos el relativo a la naturaleza de las teorías. A finales de los años setenta y en los ochenta, aunque algunas versiones venían desarrollándose desde bastante antes, se extiende y acaba imponiéndose en general una nueva caracterización de las teorías científicas que se ha denominado *Concepción Semántica de las Teorías.* En realidad no se trata de una única concepción sino de una familia de ellas que comparten algunos elementos generales relativamente unitarios en comparación con las caracterizaciones de la Concepción Heredada. A esta familia pertenecen Suppes, su pionero en los años cincuenta, y su escuela de Stanford; van Fraassen, Giere y Suppe en EEUU; Dalla Chiara y Toraldo di Francia en Italia; Przelecki y Wójcicki en Polonia; y la *concepción estructuralista de las teorías,* iniciada en EEUU por Sneed y desarrollada en Europa, principalmente, por Stegmüller, Moulines y Balzer.

En este capítulo vamos a presentar, en primer lugar, la motivación principal que, en relación con el proyecto de la Concepción Heredada, acompaña a este nuevo enfoque, así como los rasgos más generales comunes a las diferentes versiones del mismo. A continuación veremos los orígenes de la concepción modeloteórica en los trabajos fundacionales de Suppes y la contribución esencial de un miembro de su escuela, E. Adams. Después repasaremos brevemente las principales peculiaridades de cada uno de los enfoques vinculados a la familia semántica. En relación con el tema que nos ocupa, la naturaleza y estructura de las teorías, la concepción estructuralista es la que ha desarrollado un aparato metateórico más rico para el análisis y reconstrucción de las teorías científicas. La última sección está destinada a presentar con cierto detalle los elementos principales del análisis estructuralista.

1. Teorías, enunciados y modelos

El lema de las concepciones semánticas es: "presentar una teoría no es presentar una clase de axiomas, las teorías no se identifican metateóricamente con conjuntos de enunciados; presentar una teoría es presentar una clase de modelos, las teorías se identifican metateóricamente como conjuntos de modelos". Puesto que la noción de modelo es una noción fundamentalmente semántica, se denomina *concepción semántica* a este nuevo enfoque que enfatiza la importancia de los modelos en el análisis de la ciencia; contrariamente, la concepción clásica es calificada de *sintáctica* por su caracterización de las teorías como conjuntos de enunciados y por su énfasis general en los aspectos lingüístico-sintácticos. Este lema expresa por tanto el carácter distintivo frente a la concepción clásica. Pero apreciar en su justa medida cuál es ese carácter distintivo es difícil. Para ello comenzaremos revisando un aspecto de la concepción sintáctica que es claramente insatisfactorio. El enfoque semántico es en parte un intento de mejorar la concepción clásica en ese punto.

1.1. AXIOMAS Y MODELOS

Para apreciar el elemento insatisfactorio más manifiesto de la concepción sintáctico-axiomática es imprescindible tomársela en serio, tomarse en serio la *identificación* de una teoría con una serie de enunciados, los axiomas (ahora no distinguimos entre axiomas y reglas de correspondencia, pues esa distinción no afecta a la cuestión que aquí se trata). Según esta concepción, una teoría *es* una clase de axiomas, y si nos tomamos eso en serio ello implica que *toda* diferencia en axiomas supone una diferencia *de* teorías. Puesto que dos axiomatizaciones diferentes son dos diferentes clases de enunciados, tenemos dos teorías diferentes. Ésta es una consecuencia intuitivamente insatisfactoria, pues podemos tener dos axiomatizaciones diferentes de, intuitivamente, "la misma teoría". Debe quedar claro que nos estamos refiriendo a casos en los que el aparato conceptual en ambos conjuntos de axiomas es el mismo; en caso contrario las intuiciones no están tan claras. Por ejemplo, en casos como el de la equivalencia entre las versiones ondulatoria y matricial de la mecánica cuántica sí cabe decir en un sentido interesante que se trata de teorías diferentes entre las que se da determinada relación interteórica específica, la de *equivalencia* (cf. cap. 11, §4). Pero en los casos a que nos referíamos, cuando ambos conjuntos de axiomas utilizan el mismo aparato conceptual, parece intuitivamente razonable considerar que se trata de axiomatizaciones diferentes *de una misma teoría,* esto es, que no hay ningún sentido interesante en que quepa hablar de dos teorías. Si eso es así entonces una teoría no puede *ser* un conjunto de axiomas, no se representa metateóricamente de forma satisfactoria *identificándola* con un conjunto tal.

Se dirá que eso es ser demasiado rigurosos, poco caritativos con la concepción clásica. Después de todo, ya se reconocía que si dos axiomatizaciones diferentes coinciden en el conjunto de sus teoremas, se trata *en cierto sentido,* no de dos teorías diferentes equivalentes sino de dos axiomatizaciones equivalentes de la misma teoría. El problema es que la caracterización de las teorías que hace esa concepción no es el mejor modo de expresar ese cierto sentido, no puede expresarlo satisfactoriamente. Quizá se piense que sí, pues en muchas presentaciones de la concepción clásica se dice que una teoría es el conjunto de afirmaciones primitivas *más todas sus consecuencias.* Pero, si se mantiene un papel esencial para los axiomas, eso no resuelve el problema. Incluso si incluimos la referencia explícita a las consecuencias, dos conjuntos diferentes de axiomas-junto-con-sus-consecuencias (e.e. *<A, Con(A)> y <A', Con(A')>)* siguen siendo entidades diferentes aunque las consecuencias sean las mismas, pues simplemente los conjuntos de axiomas son diferentes. La única posibilidad es prescindir totalmente, en la individualización de las teorías, de la referencia a los axiomas, identificando la teoría simplemente con el conjunto de las consecuencias. Las teorías serían nombradas por expresiones del tipo `el conjunto de enunciados consecuencias de $A_1, \ldots, A_n$', *y* dos nombres diferentes, que mencionan distintos axiomas, pueden ser nombres de la misma teoría. En este caso la referencia a los axiomas sólo se incluye entonces en el nombre de la teoría, pero en la teoría misma, esto es en la identidad del conjunto infinito de enunciados, no desempeñan ningún papel. Sin embargo, así planteada, esta opción se compadece mal, como veremos, con el *axiomaticismo* que inspiraba a la Concepción Heredada. En parte, la concepción semántica consiste en expresar el núcleo de esta idea *de un modo adecuado,* un modo que no hace desempeñar a los enunciados un papel esencial en la identidad de las teorías. Nótese que el problema con la Concepción Heredada no es que quiera sostener una idea que nos parece inadecuada, no es *que pretenda* que dos teorías con el mismo vocabulario que "digan lo mismo" sean diferentes; el problema es que en su versión sintáctico-axiomática expresa inadecuadamente una intuición correcta, a saber, que en tales casos se trata de una única teoría.

El modo en que la concepción semántica va a expresar las intuiciones contenidas ya en la Concepción Heredada surge de tomarse en serio el hecho de que dos axiomatizaciones diferentes pueden serlo de la misma teoría. ¿Por qué lo son de la misma teoría? Porque el conjunto total de las cosas que dicen de cierta parcela del mundo es el mismo, porque la manera en que según ambas dicha parcela se comporta es la misma. Lo que importa de una teoría, lo que la identifica, es lo que dice sobre el comportamiento de determinada parcela de la realidad, no cómo lo dice. Lo esencial es que caracteriza ciertos trozos de la realidad como comportándose de cierto modo. Esto es, que determina ciertos modelos. Si dos axiomatizaciones lo son de lo mismo, lo son porque ambas determinan la misma clase de modelos o realizaciones. Lo importante es pues qué modelos determina una teoría, no los recursos lingüísticos que emplea para ello. De ahí el lema de la concepción semántica: presentar una teoría es presentar una clase de modelos, no de axiomas.

Se dirá que no es necesario recurrir a los modelos, que apelando sólo al conjunto total de las consecuencias de los axiomas tenemos una vía "sintáctica" equivalente; en lugar de `la clase de modelos que satisfacen A,, ..., A,,' podemos usar igualmente `el conjunto de enunciados consecuencias de A,, ..., A,,' pues nombran entidades biunívocamente relacionadas, a cada conjunto de modelos tales le corresponde un conjunto de enunciados tales, y viceversa. Pero usar la versión de las consecuencias nos mantiene en el plano sintáctico sólo aparentemente; ésta es la razón por la que hemos indicado que esta opción se compadece mal con el espíritu sintacticista propio de la Concepción Heredada. La clave es que apelar a las consecuencias es apelar implícitamente a los modelos, la noción de consecuencia introduce subrepticiamente la de modelo: un enunciado es consecuencia de otros si todos los modelos de éstos son modelos de aquél. Por tanto, si queremos expresar la idea de que mediante axiomas diferentes podemos capturar la misma teoría, debemos hacer necesariamente referencia, explícita o implícitamente, a los modelos. Si es así, lo mejor y más clarificador es hacerlo desde el comienzo: una teoría se caracteriza por determinar una clase de modelos, y su identidad está vinculada a tal clase.

Es importante comprender que esta opción no supone, ni pretende, prescindir de los enunciados o, en general, de las formulaciones lingüísticas; no pretende que los recursos lingüísticos son superfluos para la caracterización metateórica de las teorías. Por supuesto que para determinar o definir una clase de modelos hace falta un lenguaje. Los modelos, en la medida en que en el análisis metateórico se determinen explícita y precisamente, se determinan dando una serie de axiomas, principios o leyes, esto es, mediante enunciados. Nadie pretende negar tal cosa. Lo único que se pretende es que los conceptos relativos a modelos son más provechosos para el análisis filosófico de las teorías científicas, de su naturaleza y funcionamiento, que los relativos a enunciados. Que la naturaleza, función y estructura de las teorías se comprende mejor cuando su caracterización, análisis o reconstrucción metateórica se centra en los modelos que determina, no en un particular conjunto de axiomas o recursos lingüísticos mediante los que lo hace. Efectivamente la determinación de los modelos se realiza mediante una serie de axiomas, pero la identidad de la teoría no depende de esa formulación lingüística específica. Si se quiere, las formulaciones lingüísticas son esenciales en el sentido (trivial) de ser el medio necesario para la determinación de los modelos, pero en un sentido verdaderamente importante no lo son, pues nada en la identidad de una teoría depende de que la formulación lingüística sea una u otra. Resumiendo: "De acuerdo con la concepción semántica, presentar una teoría es presentar una familia de modelos. Esta familia puede ser descrita de varios modos, mediante enunciados diferentes en lenguajes diferentes, y ninguna formulación lingüística tiene ningún estatuto privilegiado. Específicamente, no se atribuye ninguna importancia a la axiomatización como tal, e incluso la teoría puede no ser axiomatizable en ningún sentido no trivial" (van Fraassen, 1989, p. 188).

El enfoque semántico, que enfatiza la referencia explícita a los modelos, más que a los enunciados, puede parecer una mera revisión del enfoque sintáctico propio de la

Concepción Heredada. Es efectivamente una revisión, pues pretende expresar más adecuadamente una idea ya contenida en la concepción anterior, aunque insatisfactoriamente expresada. Pero no es una mera revisión si con ello se quiere sugerir que se trata de una revisión sin importancia. En cuanto conceptualización más satisfactoria de una idea esencialmente correcta pero insatisfactoriamente conceptualizada con anterioridad, ejemplifica el tipo de progreso al que se puede aspirar en filosofía. Esta reconceptualización genera inmediatamente otras subsidiarias vinculadas a la idea central, lo que permite reorientar algunos problemas que más dificultades habían planteado a la Concepción Heredada. Uno de ellos será el relativo a la vinculación de los conceptos teóricos con la experiencia. Como se recordará, la Concepción Heredada sostiene que ese vínculo se establece a través de *enunciados,* las reglas de correspondencia, que conectan términos teóricos con términos que, pretendidamente, refieren a entidades directamente observables. Esta cuestión había suscitado todo tipo de problemas y, como vimos, el propio Hempel acaba rechazando la idea de que el vehículo de conexión empírica sea lingüístico. En la perspectiva sintacticista clásica pocas alternativas quedan. Veremos que la referencia a los modelos, característica de la concepción semántica, va a permitir dar una nueva orientación a esta cuestión.

1.2. EL ENFOQUE MODELOTEÓRICO

En el parágrafo anterior hemos visto la motivación y justificación del cambio de estrategia que caracteriza a las concepciones semánticas. En cuanto al desarrollo de esta estrategia, cada miembro de la familia lo hace de un modo específico, no sólo técnicamente sino que también difieren en cuestiones filosóficas fundamentales. No comparten pues una serie de tesis filosóficas sustantivas, sino un modo y un marco en el que plantear los problemas filosóficos. Lo mismo ocurre en el seno de la Concepción Heredada, donde el acuerdo general sobre el enfoque axiomático es compatible con diferencias radicales en temas filosóficos sustantivos, como el del realismo, la explicación o la causalidad.

A pesar de sus diferencias, las diversas caracterizaciones de la noción de teoría que se hacen dentro de la familia semántica tienen algunos elementos comunes:

1. Una teoría se caracteriza en primer lugar, como hemos visto, por determinar un conjunto de modelos; presentar-identificar una teoría es presentar-identificar la familia de sus modelos. La determinación de los modelos se realiza mediante una serie de principios o leyes. Las leyes se deben entender, por tanto, como definiendo una clase de modelos: "x es un sistema ... [un modelo de la teoría _] syss,,*ef*cp(x)", donde cp expresa las leyes en cuestión. Que esto sea una definición, que las leyes definan los modelos, no significa por supuesto que una teoría sea una definición, o que vaya a ser verdadera por definición, o cosas parecidas. Que las leyes definen una serie de modelos significa sólo que las leyes determinan qué entidades son las que se comportan de acuerdo con la teoría; por ejemplo, cierta entidad, cierto pedazo del mundo, es "por definición" un sistema mecánico si y sólo si cumple tales y cuales principios.

2. Una teoría no sólo determina, a través de sus leyes, una clase de modelos. Si sólo hiciera eso, poco tendríamos. Ya sabemos, por ejemplo, qué es en abstracto un sistema mecánico. ¿Qué hacemos sólo con ello? Nada. Definimos los sistemas mecánicos para algo más, quizá p.ej. para explicar el comportamiento del par de objetos Tierra-Luna. Una teoría determina una clase de modelos para algo, para dar cuenta de ciertos datos, fenómenos o experiencias correspondientes a determinado ámbito de la realidad. Parte de la identificación de la teoría consiste entonces en la identificación de esos fenómenos empíricos de los que pretende dar cuenta.

3. Una vez identificados los modelos teóricos abstractos y los fenómenos empíricos de los que se pretende dar cuenta, tenemos lo esencial de la teoría. Lo que hace la teoría es definir los modelos con la pretensión de que representan adecuadamente los fenómenos, esto es, con la pretensión de que los sistemas que constituyen los fenómenos de que queremos dar cuenta están entre los modelos de la teoría; en términos tradicionales, que tales fenómenos concretos satisfacen las leyes de la teoría, que ellos se comportan como las leyes dicen. Esta pretensión se hace explícita mediante un acto lingüístico o proposicional, mediante una *afirmación,* la afirmación o aserción "empírica" de la teoría. La aserción empírica afirma que entre los sistemas empíricos reales de que queremos dar cuenta y los modelos determinados por las leyes se da cierta relación. Esta relación puede ser de diversos tipos, más fuertes o más débiles, según las diferentes versiones. Puede ser la identidad, e.e. que los sistemas empíricos son literalmente algunos de los modelos; o la aproximación, e.e., que los sistemas empíricos se aproximan (en un sentido que hay que precisar) a los modelos; o de subsunción, e.e., que los sistemas empíricos son subsumibles (en un sentido que hay que precisar) bajo los modelos. Pero más allá de los detalles, importantes como veremos, lo esencial es que expresa la pretensión de que nuestra teoría representa adecuadamente la realidad, esto es, que nuestros modelos se "aplican bien" a los sistemas a explicar. Así es cómo la teoría dice cómo es el mundo, esos pedazos del mundo de que quiere dar cuenta en su ámbito de aplicación específico. Dice que el mundo es de cierto modo al afirmar que ciertos sistemas empíricos específicos son (o se aproximan a, o se subsumen bajo) modelos de los que ella ha definido; "el mundo", los sistemas empíricos, se comporta de "ese" modo.

Es importante enfatizar el hecho de que esta afirmación simplemente hace explícita una pretensión ya contenida implícitamente en el par "<modelos definidos, fenómenos>". Es importante para no confundirse en cuestiones importantes, como la contrastación. Algunos representantes de la concepción semántica tienden a identificar las teorías con la aserción empírica (o a incluir la aserción en la identidad de la teoría).[1] Pero, como se verá, hay buenos motivos para no identificar una teoría con su aserción empírica. Hacer eso oscurece la naturaleza estructuralmente muy compleja de las teorías, complejidad que es preciso que se refleje claramente en la noción de teoría para dar cuenta de algunos hechos fundamentales, entre otros los enfatizados por los historicistas. Es más adecuado,

1. No se piense que por eso se destierran de la familia semántica, pues siguen pensando que el mejor modo de describir esa entidad es en términos de modelos y sus relaciones.

por tanto, identificar las teorías con esos pares de conjuntos de modelos (en realidad, como veremos, con secuencias un poco más complejas de conjuntos de modelos). Así identificadas, es obvio entonces que, en un sentido estricto, las teorías no son entidades susceptibles de ser verdaderas o falsas, pues un par (una secuencia) no es una entidad a la que quepa atribuir con sentido los predicados *verdadero* y *falso*. Es cierto pues que, si las identificamos de ese modo, estrictamente las teorías no son verdaderas ni falsas. Pero nada filosóficamente sustantivo se deriva sólo de ello. Las teorías, esos pares, llevan biunívocamente asociadas entidades que sí son susceptibles de ser verdaderas o falsas, a saber, sus aserciones empíricas. Por tanto, aunque no cabe atribuir primariamente valores veritativos a las teorías, sí cabe atribuírselos *derivativamente:* una teoría es "derivativamente verdadera" si y sólo si su aserción empírica es verdadera. Y este sentido derivativo es suficientemente importante desde el punto de vista filosófico.

Insistir en que las teorías deben ser, o incluir esencialmente, aserciones puesto que *decimos* que son verdaderas o falsas, no es un argumento suficiente si hay buenas razones para no identificarlas de ese modo. Pero del hecho de que no se identifiquen con entidades proposicionales no se pueden extraer conclusiones apresuradas sobre problemas filosóficos sustantivos relativos a la "verdad" de las teorías. Por ejemplo (como veremos más adelante en detalle, cap. 12, §5), si hay cierto sentido interesante en el que las teorías no son falsables, no es porque no sean entidades a las que no cabe atribuir los predicados verdadero o falso. No cabe atribuírselo primariamente, pero sí derivativamente y con ello es suficiente para el sentido importante de falsar: si la aserción es falsa la teoría queda "falsada" en el sentido de que no todo puede permanecer igual. Si no son falsables será, como veremos, porque entendemos entonces por teoría sólo la parte esencial, el núcleo lakatosiano que siempre se puede mantener indemne a costa de suficientes reformas en la parte accidental, el cinturón protector de hipótesis específicas.

Una última advertencia antes de ver algunas de las versiones de la familia semántica. Al caracterizar los elementos generales compartidos de esta familia hemos hecho constante y central referencia a los modelos. En la sección 2 del capítulo 8 presentamos la noción intuitiva informal de modelo y una de sus posibles precisiones, la que se establece en la teoría formal de modelos. Debe quedar claro que cuando hemos hablado aquí de modelos nos referíamos a la noción informal. Las diversas versiones de la concepción semántica discrepan, entre otras cosas, en la naturaleza precisa de esas entidades a las que denominan modelos y cuya determinación identifica una teoría. Para Suppes y la concepción estructuralista, se trata de modelos en el sentido genérico de la teoría de modelos, para van Fraassen y Suppe son lo que se denomina *espacios de estado,* para Giere son modelos en cualquier sentido informal aceptable del término.

2. La noción de teoría de Suppes

Patrick Suppes es el primero en criticar la práctica general de la Concepción Heredada de identificar las teorías con determinadas formulaciones lingüísticas. En pleno apogeo de la Concepción Heredada y de su enfoque sintáctico-axiomático, Suppes plantea ya en los años cincuenta las principales objeciones que, como acabamos de ver, se le pueden hacer. Como alternativa a la axiomatización clásica desarrolla un programa alternativo de axiomatización de teorías científicas con el que se inaugura el enfoque semántico. Su propuesta es desarrollada por él mismo y algunos de sus discípulos de Stanford (cf. McKinsey, Sugar y Suppes 1953, Suppes 1954, 1957, cap. 12, 1960, 1967 y 1970b y Adams 1959); en este desarrollo E. Adams tiene, como veremos, una posición especialmente destacada al contribuir con una modificación esencial a la propuesta original de Suppes. Durante cierto tiempo, sin embargo, ese nuevo enfoque no recibe general atención y queda reducido a la llamada *escuela de Stanford*. Es a finales de los sesenta y principalmente durante los setenta, una vez superados los momentos más radicales de la revuelta historicista de los años sesenta, cuando la propuesta modelista iniciada por Suppes se extiende entre la comunidad metacientífica y es aceptada en sus aspectos más generales.

El nuevo procedimiento de axiomatización consiste en la introducción de lo que Suppes llama un *predicado conjuntista:* "axiomatizar una teoría es definir un predicado conjuntista" (1970b, p. 2/25). En esencia, un predicado tal es una manera específica de definir una clase de modelos. En este caso, tal manera se caracteriza básicamente por entender los modelos en el sentido técnico de la teoría de modelos, como sistemas o estructuras constituidas por una serie de dominios básicos y relaciones y funciones construidos sobre ellos. El recurso formal que se utiliza para definir la clase de modelos es entonces el lenguaje semiformal de la teoría intuitiva de conjuntos, completado con todos los recursos matemáticos necesarios propios de la teoría que se está axiomatizando; por ejemplo, para la mecánica clásica se usan en la axiomatización conceptos del análisis. El lema de Suppes es: el instrumento para axiomatizar las teorías científicas no es la metamatemática sino la matemática.

En esta propuesta hay que distinguir dos contribuciones, ambas importantes pero diferentes. Una es la propuesta de caracterizar una teoría definiendo una clase de modelos. Otra es la precisión de la noción de modelo en términos de secuencias de entidades conjuntistas de cierto tipo y la estrategia vinculada de determinar los modelos mediante el lenguaje conjuntista adecuadamente enriquecido. La primera es más general que la segunda, se puede concordar con Suppes en el enfoque modelista general pero discrepar en el desarrollo específico del mismo; de hecho eso es lo que hacen algunos miembros de la familia semántica. Eso no quiere decir que la segunda contribución no sea importante. Para Suppes, y para los que le siguen también en esto, la técnica conjuntista es mucho más dúctil y manejable que la clásica, permitiendo reconstruir efectivamente teorías interesantes de la ciencia real. En la perspectiva clásica, el recurso formal para la axiomatización es exclusivamente la lógica de primer orden, por lo que, si observamos estrictamente tal constricción, la axiomatización de una teoría física matematizada contiene como parte la axiomatización de toda la matemática que presupone, algo que distaba mucho de estar realizado, incluso de ser prácticamente realizable. Por ello, los ejemplos de axiomatizaciones que se manejan casi siempre en la Concepción Heredada son maquetas muy simples y poco interesantes, que no se corresponden con teorías científicas usadas realmente por los científicos.

Un predicado teórico conjuntista es un predicado del tipo "x es un sistema syss$_{ael}$ cp(x)" donde (p especifica:

a) Las entidades que componen x, esto es, que *x* es una estructura o secuencia de conjuntos y relaciones y funciones sobre ellos.

b) *(i)* Los tipos lógicos de las entidades componentes de *x,* esto es, si se trata de dominios de objetos, de relaciones o de funciones; (ii) su constitución relativa, esto es, los dominios y contradominios de relaciones y funciones; y (iii) sus propiedades matemáticas más generales, como que ciertos conjuntos son finitos, o infinitos numerables, o que cierta función es continua, etc. Los axiomas mediante los que se hacen estas caracterizaciones son meras tipificaciones, son por tanto axiomas *suį generis, o* como diremos después, *axiomas impropios.* Los axiomas impropios no imponen constricciones efectivas a las estructuras, simplemente nos dicen de qué tipo de entidades están constituidas, qué propiedades matemáticas tienen y cuáles son las relaciones lógicas de constitución entre ellas.

c) Condiciones restrictivas no puramente constitutivas o lógicas. Esto es, se trata de axiomas en sentido propio que tienen un efecto constrictivo. A las estructuras que satisfacen las condiciones definicionales de *b)* se les impone ahora como condiciones adicionales las leyes, en sentido tradicional, de la teoría. Son efectivamente restrictivas porque las cumplirán sólo algunas de las estructuras especificadas en b), otras no. Muchas veces tendrán la forma de relaciones entre varias de las entidades; por ejemplo, si en la estructura hay dos operaciones, uno de estos axiomas propios puede exigir que una sea distributiva respecto de la otra. Pero a veces pueden afectar a un solo componente; por ejemplo, se puede exigir que cierta operación sea asociativa.

Para fijar las ideas, reproducimos como ejemplo la definición del predicado "x es un sistema de mecánica de partículas" (cf. Suppes, 1957, cap. 12, parcialmente modificado en Adams, 1959; la presente es una versión mixta, con algunas simplificaciones notacionales que suponen algunas deficiencias técnicas, sobre todo en (8), pero es suficiente para los actuales fines ilustrativos).

Definición 10.1:

x es un *sistema de mecánica (newtoniana) de partículas* syss$_{def}$ existen P, T, s, m, f tales que:

(1) $x = \langle P, T, s, m, f \rangle$

(2) P es un conjunto finito no vacío.

(3) T es un intervalo de números reales.

(4) s es una función de $P \times T$ en el conjunto de vectores tridimensionales (tríos ordenados) de números reales, y es dos veces diferenciable sobre T.

(5) m es una función de P en el conjunto de números reales tal que para todo $p \in P$: $m(p) > 0$.

(6) f es una función de $P \times T \times N$ en el conjunto de vectores tridimensionales (tríos ordenados) de números reales.

(N es el conjunto-ayuda de números naturales, que marca con un índice la f para cada p y t; podríamos escribir '$f(p, t)$' en lugar de '$f(p, t, i)$').

(7) Para todo $p \in P$ y $t \in T$: $m(p) \cdot d^2/dt^2 [s(p, t)] = \sum_{i \in N} f(p, t, i)$.

(8) Para todo $p, q \in P$ y $t \in T$:

(i) $f(p, t, i_q) = -f(q, t, j_p)$

(ii) $s(p, t) \otimes f(p, t, i_q) = -s(q,t) \otimes f(q, t, j_p)$.

(Aclaración notacional: indicamos mediante 'i_q' que la f que tiene como uno de sus argumentos dicho índice "se debe a q", así '$f(p,t,i_q)$' denota el valor de f sobre p en t "debido a q", *e.e.*, la fuerza que ejerce q sobre p; '$\otimes$' denota el producto vectorial.)

(1) presenta (el número de) los constituyentes de las estructuras. (2)-(6) son los axiomas impropios, meras tipificaciones lógico-matemáticas de las entidades que constituyen la estructura. La idea es que P es un conjunto específico de partículas: en una estructura x determinada ese conjunto contiene sólo la Tierra y la Luna; en otra, el Sol y los planetas; en otra, la Tierra y un péndulo; en otra la Tierra y dos objetos en una polea; etc. T es un conjunto de instantes temporales. s es la función posición, que asigna a cada partícula del sistema un determinado vector-posición en cada instante; es dos veces diferenciable respecto del tiempo, su primera derivada es la velocidad y su segunda derivada es la aceleración. m es la función masa, que asigna a cada partícula un número real positivo, su masa (que es independiente del tiempo). f es la función fuerza, que asigna a cada partícula en cada instante una serie de vectores-fuerza, las fuerzas actuantes sobre la partícula en ese instante; en vez de tener varias funciones, tenemos una única función que tiene como argumentos, además de partículas e instantes, ciertos índices que distinguen los diferentes vectores-fuerza actuantes sobre p en t; así, $f(p, t, i) = \langle x_1, x_2, x_3 \rangle$ y $f(p, t, j) = \langle y_1, y_2, y_3 \rangle$ $(i \neq j)$ son los valores de dos fuerzas diferentes actuantes sobre la partícula p en el instante t. (7) y (8) son los axiomas propios, expresan las leyes propiamente dichas de esta teoría. (7) expresa el segundo principio de Newton: la suma (vectorial) de las fuerzas actuantes sobre una partícula en un instante es igual a la variación de cantidad de movimiento, o como se suele decir, al producto de la masa de la partícula por su vector-aceleración en ese instante. (8) expresa (con ciertas deficiencias técnicas) el principio de acción y reacción: las fuerzas que se ejercen mutuamente dos partículas son de igual módulo y dirección y de sentidos contrarios.

Éste es un ejemplo típico de la axiomatización suppesiana de una teoría mediante la definición de un predicado conjuntista. Debe quedar claro que lo que se hace es, como habíamos anunciado, definir cierta clase de modelos. Las estructuras que satisfacen (1)-(8) son, *por definición*, sistemas mecánicos newtonianos. Presentar la mecánica newtoniana es presentar (definir) esa clase de modelos. Debe quedar claro también que esos modelos están sometidos a, son caracterizados a través de, algunas condiciones efectivamente restrictivas. Las condiciones (1)-(6), meras tipificaciones, determinan simplemente el tipo lógico-matemático de las entidades que constituyen los sistemas. Las entidades de ese tipo lógico, que satisfacen (1)-(6), son, por decirlo así, candidatos a ser modelos de la teoría; esto es, entidades de las que tiene sentido plantearse si se comportan del modo que dice la teoría, si cumplen las leyes propiamente dichas. Si una estructura no tiene una función que asigne a los elementos del dominio números reales, no tiene sentido preguntarse si cumple o no el segundo principio de Newton, pues tal principio involucra funciones de ese tipo. A las estructuras que satisfacen las tipificaciones las llama Suppes *realizaciones posibles* (cf. 1960, pp. 287-289). Lo que debe quedar claro es que lo esencial de una teoría no son (sólo) sus posibles realizaciones, sino (principalmente) sus *realizaciones efectivas o modelos* en sentido propio. La teoría no sólo contiene tipificaciones, contiene condiciones adicionales que son restrictivas en el sentido de que algunas de las realizaciones posibles las cumplirán, pero otras no. No por tener el tipo de conjuntos y funciones que especifican (l)-(6) toda estructura va a satisfacer (7)-(8); puede ser que tenga ese tipo de entidades, pero que la suma de los vectores-fuerza para una partícula en un instante simplemente no dé el mismo resultado que el producto de su masa por su aceleración, por ejemplo que sea igual al producto de la masa por el cuadrado de la aceleración, o la raíz cuadrada del producto de la masa por la aceleración, o cualquier otra cosa (como ejercicio, el lector puede construir un ejemplo puramente numérico de sistema que cumpla (1)-(6) pero no (7)). Las realizaciones efectivas o modelos de una teoría son aquellas realizaciones posibles que además satisfacen los axiomas propios; el conjunto de modelos será por tanto en general un subconjunto propio del conjunto de realizaciones posibles.

3. Adams y las aplicaciones intencionales

En la sección anterior hemos presentado lo esencial de la nueva caracterización que hace Suppes de las teorías científicas, debemos ver ahora brevemente la importante modificación que introduce su discípulo E. Adams. La modificación de Adams está destinada a subsanar lo que él considera una insuficiencia crucial de la versión original de Suppes.

La insuficiencia que Adams atribuye a la propuesta de Suppes tiene que ver con algo que hemos hecho al presentar el ejemplo y que Suppes mismo hace, y que sin embargo no es claro que se pueda hacer desde sus presupuestos. Una vez presentado el predicado conjuntista, hemos indicado cuál era la *interpretación pretendida* de las entidades componentes de los modelos, esto es, partículas físicas, sus masas, posiciones espaciales, fuerzas incidentes, etc. La cuestión es, ¿quién dice eso?, ¿cómo dice eso la teoría? Puede ocurrir que el predicado sea satisfecho por entidades que ontológicamente nada tengan que ver con esas entidades pretendidas. Por ejemplo, que los ángeles, junto con su "cantidad de espíritu", sus "afinidades" o lo que sea, satisfagan esos axiomas. O, por poner un ejemplo menos absurdo, esos axiomas son satisfechos de hecho por estructuras puramente matemáticas, esto es, estructuras cuyo conjunto P está constituido por números. En otras palabras, entre los modelos efectivos, no meramente entre las realizaciones posibles, sino entre las realizaciones efectivas que cumplen (7) y (8) además de (1)-(6), hay con seguridad sistemas puramente matemáticos (y quizá "angélicos" u otros de parecida rareza), sistemas *de los que no pretende hablar la teoría.* Parece claro que es esencial a una teoría *empírica* el que pretenda aplicarse sólo a algunos de sus modelos efectivos; en el ejemplo visto no se pensaron los principios newtonianos para sistemas puramente matemáticos (o angélicos). Pero si presentar una teoría consiste *exclusivamente* en presentar una clase de modelos definidos mediante la introducción de un predicado conjuntista (con axiomas impropios y propios), no se ve cómo se puede recoger ese hecho.

La cuestión en juego es, como el lector habrá adivinado, la de la interpretación empírica. El predicado conjuntista que define los modelos es un mero formalismo matemático abstracto carente de interpretación empírica, o mejor dicho compatible con interpretaciones muy diferentes, tanto empíricas como no empíricas. El conjunto de modelos que tal predicado determina incluye sistemas de la más variada constitución, tanto empíricos como matemáticos. Efectivamente, estamos de nuevo ante el viejo problema de la conexión del formalismo con la experiencia. Otro modo de presentar la objeción a Suppes es mostrar que su caracterización, sin elementos adicionales, no permite distinguir las teorías empíricas de las teorías matemáticas. Para Suppes eso no es un problema tan grave, pues piensa que en realidad la diferencia entre unas y otras no es siempre tan clara como se pretende, y que una ventaja de su enfoque es justamente que hace explícito ese hecho. Naturalmente Suppes no pretende negar que a veces hay una diferencia. Reconoce que hay casos en que es así y ofrece una vía para dar cuenta de ella. Sin embargo, Suppes no piensa que esa diferencia, cuando se da, haya de reflejarse en la estructura manifiesta de la teoría. La diferencia radica en que, en las teorías empíricas (matematizadas), la determinación-medición de algunas de (o todas) sus magnitudes vincula dicha magnitud con situaciones empíricas cualitativas que fundamentan la medición; por ejemplo, la función masa está ligada a procedimientos de comparación cualitativa mediante una balanza de brazos. Esas situaciones empíricas cualitativas sobre las que descansa en última instancia[2] la medición son estudiadas por las llamadas *teorías de la medición (metrización) fundamental* (para estas y otras nociones relativas a la medición, cf. cap. 6). La interpretación empírica de una teoría se expresa entonces a través de los vínculos que guardan sus funciones métricas con las teorías de la medición fundamental. Por tanto, la interpretación empírica no se manifiesta "inmediatamente" en la caracterización-axiomatización de una teoría, sino sólo en la reconstrucción de sus vínculos interteóricos con las teorías de la metrización fundamental.

Adams plantea esencialmente la misma objeción que hemos presentado, pero de un modo que no se puede resolver apelando a la medición fundamental. La objeción de Adams es que si caracterizamos las teorías, como hace Suppes, *exclusivamente* mediante el conjunto de sus modelos o realizaciones efectivas, entonces no es posible hacer explícito el elemento *veritativo o proposicional* de las teorías; esto es, no es posible hacer explícito el sentido en que las teorías son verdaderas o falsas, o si se prefiere, correctas o incorrectas. El conjunto de modelos caracteriza *cierto modo como pueden ser las cosas,* el modo como

2. En última instancia porque, como vimos en el capítulo 6, algunas veces (la mayoría en realidad) la medición de una magnitud para cierto objeto usa simplemente otros valores. Eso es la medición indirecta. Pero, recuérdese, la medición indirecta no puede ser el único procedimiento de medición, pues los valores previamente disponibles se han tenido que medir con anterioridad, y así sucesivamente. Así, en algún lugar debe empezar la tarea, en algún momento asignamos números a las cosas sin usar números previamente disponibles. Esos son justamente los procedimientos de medición directa o fundamentales, sobre los que descansa *en última instancia* toda medición, y a los que se refiere Suppes.

son las cosas según la teoría. Pero ¿de qué cosas trata? La teoría quiere decir "así son las cosas". Pero ¿de *qué cosas* dice ella que son *así?:* ¿planetas?, ¿péndulos?, ¿países?, ¿ángeles?, ¿simples números? El "así" está expresado por el conjunto de modelos. Pero si eso es todo lo que tenemos, nos falta algo que exprese "las cosas" de las que se pretende que son de ese modo. Sin eso no podemos expresar esa pretensión de la teoría. Como vimos, esta pretensión es esencial a las teorías, pues éstas son ideadas para dar cuenta de parcelas específicas de la realidad. Y esta pretensión contiene el elemento proposicional de las teorías, pues se expresa mediante una afirmación susceptible de ser verdadera o falsa: verdadera si *esas* cosas son efectivamente *así* (si están entre los modelos), falsa si no lo son.

Adams propone "abordar el concepto de *verdad o corrección [...]* a través de la noción de *interpretación pretendida ['intended'] o modelo pretendido* de la teoría, [... que es] cualquier sistema del cual [...] se pretende que se ajusta a los axiomas. Hay siempre en general un enorme número de sistemas que satisfacen los axiomas de la teoría, pero en las teorías de la ciencia empírica, normalmente sólo unos pocos de ellos serán aplicaciones o modelos pretendidos" (1959, p. 258). Son modelos pretendidos de la mecánica newtoniana, por ejemplo, el sistema formado por la Tierra y la Luna, o el constituido por el Sol con los planetas, o un plano inclinado, o un proyectil sobre la Tierra, etc. La identificación o caracterización metateórica de una teoría debe incluir entonces, además del conjunto de modelos que satisfacen el predicado, un conjunto de aplicaciones, de sistemas físicos específicos, de "partes concretas de la realidad", de las que se pretende que se comportan como la teoría dice, esto es, de las que se pretende que están entre los modelos. Resumiendo: "Si la verdad y la falsedad han de ser definidas, hemos visto que se deben tener en cuenta dos aspectos de una teoría: primero, el aspecto formal que corresponde al predicado conjuntista definido mediante los axiomas, [... o mejor,] la extensión de dicho predicado, el conjunto de los sistemas que satisfacen los axiomas; y segundo, el aspecto aplicativo, que corresponde al conjunto de modelos pretendidos. Formalmente, una teoría T se caracterizará como un par ordenado de conjuntos $T = \langle C, I \rangle$ tal que C es el conjunto de todas las entidades que satisfacen los axiomas, e I es el conjunto de modelos pretendidos" *(ibid.).* Como se ve, una teoría no es estrictamente una entidad de la que cabe predicar primariamente la verdad o la falsedad, pero en un sentido lato, derivativo, sí que es adecuado, y esencial, decir que puede ser verdadera o falsa: "La teoría es verdadera si y sólo si todos sus modelos pretendidos satisfacen sus axiomas, en caso contrario es falsa. Si $T = \langle C, I \rangle$, entonces T es verdadera si y sólo si I está incluido en C" *(ibid., pp.* 259-260). "1 $\subseteq$ C" expresa pues sucintamente la aserción o hipótesis empírica vinculada a la teoría, de la cual ésta hereda su valor veritativo.

Ésta es la modificación esencial con la que Adams contribuye al programa de Suppes. En la versión de Adams, esta modificación presenta sin embargo algunas dificultades. La más inmediata es que queda oscuro el modo en que se determinan las aplicaciones pretendidas y, con ello, la forma en que se contrasta la aserción empírica. Por supuesto que las aplicaciones no se "extraen" simplemente de entre los modelos del conjunto C, pues entonces la aserción sería tautológica. Para que quede clara la naturaleza del problema es esencial distinguir dos sentidos de `determinar las aplicaciones'. En un primer sentido significa "seleccionarlas". La cuestión es entonces cómo se seleccionan los sistemas empíricos, las partes concretas de la realidad a la que se pretende aplicar la teoría. El único modo de responder a esta cuestión es apelando a las intenciones de la comunidad de científicos: *I* es el conjunto de sistemas empíricos *x* tales que la comunidad científica *CC pretende o intenta* aplicar *T a x*. Por ejemplo, en las fases iniciales de la Mecánica Clásica, los físicos pretendían que la teoría se aplicaba a cuerpos en caída libre, tiros parabólicos, trayectorias de cuerpos celestes, y muchas otras cosas, entre ellas los rayos de luz; la luz fue inicialmente una aplicación intencional de la mecánica (al menos de los partidarios de la teoría corpuscular, como el propio Newton), aplicación que terminó por excluirse del dominio de aplicaciones cuando se impuso la teoría ondulatoria rival. Simplemente, qué sistemas específicos están en I depende exclusivamente de las pretensiones o intenciones de los científicos (en un momento dado, cf. cap. 13).

En un segundo sentido, `determinar las aplicaciones' significa, una vez seleccionadas, "determinar sus parámetros", típicamente en los casos de teorías cuantitativas, determinar en cada aplicación los valores precisos de cada una de las magnitudes involucradas. Y aquí es donde aparece el problema, pues, si en la determinación de las aplicaciones, en la medición de los valores de las magnitudes del sistema-aplicación x del que se quiere contrastar si se ajusta o no a las leyes de *T,* se usaran las leyes de *T,* estaríamos ante un expediente autojustificativo. Esto es, si en la determinación de los hechos o base empírica de aplicación se usaran las leyes de la teoría, la aserción se *autojustificaría.* El problema con la caracterización de Adams es que no es lo suficientemente fina para abordar esta cuestión. Nótese que según Adams la aserción empírica es de la forma I c C, y por tanto cada aplicación concreta x es un sistema del mismo tipo lógico que los modelos actuales, tienen los mismos componentes, las mismas funciones. Eso supone que determinar una aplicación seleccionada exige medir en dicho sistema los valores de todas las funciones de las que habla la teoría. Como veremos más adelante, si eso fuese efectivamente así, estaríamos irremisiblemente condenados al problema de la autojustificación, pues *algunas* de las funciones de las que habla la teoría no se pueden medir sin usar sus propias leyes. En la medida en que las teorías no son localmente autojustificativas, en esa misma medida el análisis de Adams es insatisfactorio, no puede ser que la contrastación de una teoría exija disponer en los sistemas-aplicación de los valores para todas las magnitudes de que habla la teoría. Veremos que una de las motivaciones por las que surge el estructuralismo en Sneed es precisamente caracterizar las aplicaciones pretendidas de un modo más adecuado que permita elucidar el carácter no autojustificativo de la aserción empírica.

Antes de concluir con la escuela de Stanford, hay que señalar que el propio Suppes se plantea en cierto momento la cuestión de la aplicación empírica de las teorías empíricas desde una perspectiva que guarda algo de semejanza con el espíritu de la propuesta de Adams. En un trabajo de 1960 publicado dos años más tarde, 'Models of Data', defiende que lo que cuenta como datos para una teoría se presenta también en forma de modelos, los *modelos de datos.* La diferencia entre las teorías empíricas y matemáticas es que en las primeras, y no en las segundas, los modelos de datos son de distinto tipo lógico que los modelos teóricos. Aunque no es totalmente explícito en este punto, parece que la diferencia de tipo lógico a que se refiere en el caso de teorías empíricas consiste en que los modelos de datos son subestructuras de los modelos teóricos. A juzgar por el ejemplo que presenta, de este modo parece que se debe interpretar su afirmación de que "en la teoría [empírica] se usan nociones teóricas que no tienen un análogo directo observable en los datos experimentales" (§1). En su ejemplo, la teoría del aprendizaje Estes-Suppes (cf. Suppes y Estes, 1959), los modelos de la teoría están constituidos por ciertas entidades, algunas consideradas observables y otras no; los modelos de datos están constituidos entonces por los constituyentes *observables* de los modelos teóricos, de modo que resultan ser subestructuras de aquéllos. Los modelos de datos, además, son definidos por sus propias teorías, y es a través de su conexión con estas teorías de datos como adquiere contenido empírico la primera. "Lo que he intentado argüir es que se establece una jerarquía completa de modelos entre los modelos de la teoría básica y la base experimental completa. Más aún, para cada nivel de la jerarquía hay una teoría por derecho propio. A la teoría de cierto nivel le es dado su significado empírico al hacer conexiones formales con la teoría de un nivel más bajo" (§3).

La propuesta de Suppes está sólo esbozada en este artículo, y no llegó a desarrollarla en trabajos posteriores (de hecho, posteriormente parece contradecirla parcialmente, pues exige que los datos sean del mismo tipo lógico que los modelos teóricos, cf. Suppes 1989, p. 264). En esa versión es muy imprecisa, está poco articulada con el resto de su programa y contiene elementos problemáticos que no se tratan. Aunque puede encontrarse cierta semejanza de espíritu con las ideas de Adams, sus modelos de datos no se corresponden exactamente con las aplicaciones pretendidas de Adams. Aquéllos son observacionales y plenamente determinables teóricamente (mediante otra teoría de bajo nivel); éstas se determinan intencionalmente y no tienen por qué ser plenamente observacionales, de hecho no lo pueden ser si deben tener el mismo tipo lógico que los modelos teóricos. Veremos que el análisis satisfactorio de la base empírica incorpora elementos de ambos.

4. La familia semanticista

Como indicamos, el enfoque semántico inaugurado por Suppes se mantiene en principio circunscrito al ámbito de su grupo en Stanford, pero a finales de los años sesenta comienza a expandirse y durante los setenta se va asentando poco a poco hasta convertirse en dominante a partir de los ochenta. Veremos ahora brevemente los elementos específicos de los representantes más destacados de este nuevo enfoque: van Fraassen, Suppe, Giere y la Concepción Estructuralista (para la escuela polaca, cf. Przelecki, 1969, y Wójcicki, 1977 y 1979; para la escuela italiana, cf. Dalla Chiara y Toraldo di Francia, 1973 y 1976). Aunque la implantación general se realiza bajo la influencia de los trabajos de Suppes, no todos los miembros de la familia están directamente influidos por él o le siguen en los aspectos específicos de su propuesta. Se trata más bien de que a la estela de la propuesta específica de Suppes se desarrollan una serie de otras propuestas que en muchos casos comparten con aquél sólo la orientación modelística. Comparten tan sólo una estrategia general y una preferencia por determinada forma, la modelística, de presentar y analizar los problemas, pero, como también advertimos, no comparten tesis filosóficas sustantivas.

Casi todos los miembros de esta familia realizan contribuciones importantes en varios ámbitos de la filosofía de la ciencia, y algunas de ellas se presentan en detalle en otras partes de esta obra. En relación al tema que ahora nos ocupa, la estructura de las teorías, la concepción estructuralista es la que ha realizado un análisis más detallado de la estructura fina de las teorías, ejemplificando tal análisis con numerosas reconstrucciones de teorías específicas. Los otros miembros de la familia se limitan en este tema a presentar los aspectos más generales de su propuesta semántica particular, sin desarrollar en detalle la estructura fina de las teorías. Veremos aquí cuáles son esos aspectos más generales característicos de cada una de las propuestas y en la próxima sección presentaremos en detalle el análisis estructuralista.

4.1. VAN FRAASSEN: ESPACIOS DE ESTADO; BASE EMPÍRICA Y OBSERVABILIDAD

Van Fraassen coincide con Suppes en que el modo filosóficamente más iluminador de caracterizar una teoría es presentándola como definiendo una clase de modelos. Discrepa de él, sin embargo, en la naturaleza matemática de estas entidades. Frente a los modelos como estructuras conjuntistas de Suppes, van Fraassen opta por los modelos como "puntos" o "trayectorias" en un *espacio de estados,* idea cuya aplicación a las teorías físicas atribuye a Beth. Beth (cf. 1960) propone un análisis semántico de las mecánicas newtoniana y cuántica en términos de sistemas constituidos por estados gobernados por las ecuaciones mecánicas fundamentales. Van Fraassen desarrolla y generaliza esta idea a principios de los años setenta (cf. 1970 y 1972). Aunque los detalles son complicados y no podemos verlos aquí, el núcleo de la idea es el siguiente (van Fraassen advierte sobre las limitaciones para el caso de teorías físicas relativistas, pero no nos detendremos en ello).

Un estado de un sistema está definido por los valores de ciertas magnitudes en cierto momento (cf. cap. 5, § 1.3). Por ejemplo, un estado de un gas queda definido por los valores del volumen, la presión y la temperatura; se puede identificar por tanto con una triada ordenada $<v, p, t>$ de números reales, donde cada componente es, respectivamente, el valor de la correspondiente magnitud. En mecánica, el estado de cada partícula en un instante lo determina su posición $q = (q_x, q_y, q_z)$ *y su* momento $p = (p_x, p_y, p_z)$; el estado se puede identificar con el séxtuplo ordenado $<q_x, q_y, q_z, p_x, p_y, p_z>$. Los estados se identifican por tanto en general con puntos en un determinado sistema de coordenadas, de tantas dimensiones como componentes tengan los estados, tridimensional en el primer ejemplo, hexadimensional en el segundo. A cada tipo de sistema le corresponde entonces un *espacio de estados,* el conjunto de todas las posibles n-secuencias (*n* es la dimensión del espacio) de valores; los estados posibles de los sistemas de ese tipo son pues los puntos de ese espacio. Lo que hacen los postulados y leyes de una teoría es imponer constricciones sobre las relaciones entre estados, permitiendo ciertas transiciones (leyes de sucesión) o coexistencias (leyes de coexistencia) entre estados y excluyendo otras (sobre las leyes de sucesión y coexistencia, cf. cap. 5, § 1.3). Las transiciones se identifican con determinadas trayectorias en dicho espacio, y las coexistencias con regiones específicas del mismo. Las leyes de una teoría permiten ciertas trayectorias y regiones y excluyen otras; de entre todas las trayectorias y regiones *lógicamente* posibles, la teoría determina sólo algunas de ellas, las *nómicamente* posibles. Así, el conjunto completo de puntos del espacio es el análogo al conjunto de realizaciones posibles de Suppes, y el subconjunto del mismo permitido por las leyes es el análogo al conjunto de realizaciones efectivas de Suppes. En ambos casos tenemos un espacio de modelos *lógicamente posibles* en relación con el cual las leyes de la teoría determinan el subespacio de modelos *físicamente posibles.*

Como en Suppes, por tanto, la teoría define mediante las leyes una clase de modelos, pero ahora tales modelos son trayectorias o regiones permitidas en un espacio de estados de determinada dimensión. Esta diferencia en la caracterización de los modelos no tiene consecuencias filosóficas sustantivas. En concreto, la forma de antirrealismo que van Fraassen defiende, su llamado *empirismo constructivo,* no depende de las preferencias sobre la forma de los modelos. El empirismo constructivo es una tesis epistemológica acerca de qué creencias implica la aceptación de una teoría. En la defensa de esta tesis epistemológica, van Fraassen desarrolla toda una variedad de tesis, de orientación general también antirrealista, sobre muchas cuestiones filosóficas sustantivas, como la causalidad, la explicación, las leyes, la modalidad o la observabilidad (cf. especialmente 1980 y 1989). No es éste el lugar de revisarlas, ni siquiera someramente. Nos limitaremos para concluir a presentar la idea de base empírica sobre la que sostiene parte de su argumento general.

"La parte `pura' de la teoría define el tipo de sistemas a los cuales se aplica; las aserciones empíricas tendrán la forma de que cierto sistema empírico dado pertenece a tal clase" (1970, p. 311). En realidad la aserción no dice, como en Adams, exactamente que los sistemas empíricos pertenecen a dicha clase, que son algunos de los modelos, sino sólo que son "subsumibles". La diferencia radica en que los sistemas a los que se aplica la teoría son submodelos, subestructuras de los modelos determinados por las leyes consistentes en quedarnos con la parte observacional de los modelos: "ciertas partes de los modelos [son] identificadas como *subestructuras empíricas,* y esos [son] los candidatos para la representación de los fenómenos observables con los cuales la ciencia se puede confrontar en nuestra experiencia, [...] la adecuación empírica consiste en la subsumibilidad de esas partes en algún modelo único del mundo permitido por la teoría" (1989, pp. 227-228). Lo que hace la teoría es postular la existencia de ciertas entidades inobservables, "ocultas", cuya (supuesta) interacción con las entidades observables produce (pretendidamente) los efectos observables, los fenómenos. Parte de lo que la teoría sostiene es que esas subestructuras empíricas son subsumibles bajo uno de sus modelos, esto es, que se comportan del modo en que lo harían si el mundo fuese uno de sus modelos, con sus entidades ocultas interaccionando con las observacionales del modo específico indicado en las leyes. Ése es el contenido de la aserción empírica y si dicha aserción es verdadera decimos que la teoría es *empíricamente adecuada* (que "salva los fenómenos").

Van Fraassen insiste en que eso es sólo parte de lo que la teoría dice, porque quiere defender que la teoría dice también algo más, dice que el mundo contiene tales y cuales entidades además de las observables: "Es claro que podemos discutir dos cuestiones separadas: ¿qué dice la teoría sobre cómo es el mundo? y ¿qué dice la teoría sobre cómo son los fenómenos? Puesto que los fenómenos son la parte observable del mundo, y es contingente que haya o no otras partes, se sigue que estas preguntas no son la misma" (1989, p. 191). Lo que quiere defender es que *la teoría misma,* y no sólo su aserción empírica, puede ser verdadera o falsa. Por eso insiste en que la teoría debe ser una entidad en cierto sentido proposicional, con valor veritativo y susceptible de ser o no creída. Hay un sentido débil en que la teoría puede ser verdadera o falsa, a saber, que su aserción es verdadera o falsa, que *la parte observacional del mundo* es como dice la teoría. Pero hay un sentido más fuerte en que la teoría puede ser verdadera o falsa, a saber, es verdadera si y sólo si *el mundo* es como dice la teoría, esto es, *si el mundo es uno de sus modelos.* En el primer sentido prefiere hablar, más que de verdad de la teoría, de *adecuación empírica;* sólo en el segundo sentido la teoría es propiamente *verdadera.* Este doble sentido se aplica también a las actitudes proposicionales que los sujetos epistémicos podemos tener hacia las teorías. Podemos creer sólo que la teoría es empíricamente adecuada, que su aserción empírica es verdadera; o podemos creer algo más, a saber, que la teoría misma, toda ella, es verdadera.

En estos términos puede formular ahora van Fraassen su antirrealismo sucintamente. En su opinión, el realismo no es una tesis ontológica sobre lo que hay, sino una tesis epistemológica sobre lo que estamos justificados en creer que hay. Su antirrealismo sostiene que al aceptar una teoría estamos justificados sólo en creer en su adecuación empírica, no en su verdad. Aceptar una teoría nos compromete sólo a creer que lo que afirma de la parte observable del mundo es verdad, no a creer que lo que *también* afirma acerca de inobservables es verdad. A esta posición antirrealista hacia lo inobservable la denomina van Fraassen *empirismo constructivo: "Uso* el adjetivo `constructivo' para indicar mi concepción de que la actividad científica es una actividad de construcción y no de descubrimiento: construcción de modelos que deben ser adecuados a los fenómenos, y no descubrimiento de la verdad acerca de lo inobservable" (1980, p. 5).

Este antirrealismo es en opinión de van Fraassen la conclusión ineludible de dos premisas en su opinión irrechazables: *a)* la tesis empirista según la cual la justificación de toda creencia empírica debe descansar en los fenómenos, en la experiencia, *y b)* el hecho lógico de que puede haber teorías diferentes *incompatibles* entre sí pero *empíricamente equivalentes,* con las mismas consecuencias contrastacionales (en esto consiste la *infradeterminación de la teoría por la experiencia,* sobre la que volveremos por extenso en el capítulo 12 dedicado al problema de la inducción). De *b)* se sigue que la creencia en una teoría frente a otra incompatible empíricamente equivalente no está basada en la experiencia y, por tanto, por *a),* no será una creencia justificada. En general, pues, sólo estamos justificados en creer en la adecuación empírica, no en la verdad de una teoría, de *toda* ella.

Aunque no podemos discutir aquí a fondo este argumento, debe notarse que para que concluya lo que pretende van Fraassen ha de aceptarse una premisa implícita adicional. De *a) y b)* se sigue que sólo estamos justificados en creer las afirmaciones que las teorías hacen sobre entidades "dadas en la experiencia", pero para concluir que sólo estamos justificados en creer las afirmaciones que las teorías hacen sobre entidades observables, hace falta la premisa adicional según la cual *c) la parte empírica de las teorías, su base de contrastación, es siempre observacional. El* reto todavía pendiente es ofrecer una noción precisa y plausible de *observabilidad* que sustente c). Van Fraassen defiende un concepto antropocéntrico de *observable*. Afirma que, en tanto que organismos biológicos, somos cierto tipo de mecanismos de medición o detección, que como tales tenemos ciertas limitaciones inherentes y "son estas limitaciones a las cuales refiere el 'able' de `observable'" (1980, p. 17). Por ser antropocéntrico, este concepto no puede tener relevancia ontológica, para saber lo que hay, pero sí epistemológica, para saber qué estamos justificados a creer que hay. Reconoce además que el concepto es hasta cierto punto vago. Por ejemplo, sostiene que las lunas de Júpiter son observables, pues los astronautas serían capaces de verlas directamente si se acercaran a ellas, pero que las partículas en una cámara de niebla no lo son, pues el juicio sobre su presencia incluye inferencias teóricas; pero entonces, ¿qué decir de la observación con microscopio electrónico?, ¿y de una estrella lejana que quizá ya ha desaparecido? Pero antropocentrismo y vaguedad no son los problemas principales, pues son asumibles por sus tesis. Recuérdese que su tesis antirrealista es epistémica, es una tesis acerca de lo que los humanos estamos justificados en creer que hay, y por eso no es objetable que su antirrealismo esté relativizado a nuestras capacidades epistémicas, esto es, que dependa de una noción antropomórfica de `observable'.

El problema principal es si se puede sostener que los sistemas empíricos que ejercen de datos en las teorías están constituidos por entidades observables *en su sentido de `observable'*. *Él* mismo reconoce que "la teoría no se confronta con datos brutos sino con modelos de datos, y la construcción de estos datos es un proceso sofisticado y creativo" (1989, p. 229). De nuevo, como ocurría con la Concepción Heredada, incluso si en términos globales nuestro conocimiento se origina en situaciones observables en dicho sentido, hace falta un argumento adicional para establecer que la base empírica de cada teoría tiene esas características. Más bien parece que no siempre es así; en realidad casi nunca es así, o nunca si hablamos de teorías científicas mínimamente desarrolladas. Por seguir con su propio ejemplo: reconoce que las partículas no son observables en una cámara de niebla, que lo observable son los rastros en la niebla; afirma que los modelos de datos que ejercen de base empírica son partes, subestructuras, de los modelos de la teoría; pero, simplemente, sucede que los modelos de la mecánica cuántica no incluyen entre sus entidades cosas como rastros en la niebla.

Si c) no es cierto, entonces para que su argumento concluya lo que pretende hay que reinterpretar *a)* de modo que se refiera a la observación: la justificación de toda creencia *descansa* en la "observación directa". Pero el problema ahora es con `descansa'. Si es "descansa inmediatamente", entonces su antirrealismo se aplica también a la base empírica de contrastación cuando no sea directamente observable. Si es "descansa en última instancia", entonces hay que elaborar en detalle cuál es la relación entre la base empírica y la observación y qué se considera "en última instancia" Esto es esencial, pues dependiendo de qué aceptemos como "descansar en última instancia", vuelven a abrirse toda serie de estrategias a los realistas para recuperar la justificación de la creencia en las entidades "teóricas" postuladas por la teoría para dar cuenta de los modelos de datos de experiencia. En definitiva, el antirrealismo de van Fraassen parece, sin especificaciones adicionales, inestable: o se aplica también a la base de contrastación (cuando ésta no sea directamente observable), o no tiene por qué aplicarse a las entidades teóricas.

4.2. SUPPE: SISTEMAS RELACIONALES; FENÓMENOS, DATOS Y TEORÍAS

Suppe inicia su propio enfoque semántico en su tesis doctoral (cf. 1967) dedicada al significado y uso de los modelos en la ciencia, influido por los trabajos de von Neumann y Birkhoff sobre fundamentación de la mecánica cuántica y por los de Suppes sobre modelos de datos. En dos trabajos clásicos sobre la Concepción Heredada, prácticamente ignorada en su tesis, contrasta los aspectos centrales de dicho enfoque con la concepción axiomática clásica (cf. 1972 y 1974), y durante finales de los años setenta y en los ochenta desarrolla su concepción aplicándola a los principales temas de la filosofía de la ciencia (cf. 1989).

Suppe sigue a Suppes en la aproximación modeloteórica general pero, como van Fraassen, influido en su caso por los trabajos de von Neumann y Birkhoff, prefiere caracterizar los modelos mediante estados en un espacio de estados, no al modo conjuntista de Suppes. El instrumental matemático es prácticamente coincidente con el de van Fraassen y no abundaremos en él. Una teoría se analiza ahora como un *sistema relacional (cf.* 1989 p. 84), consistente en *a)* un dominio que contiene todos los estados lógicamente posibles de los sistemas de que trata la teoría (e.e. el espacio de estados entero) y *b)* una serie de relaciones entre los estados, determinadas por los postulados o leyes de la teoría, que especifican las trayectorias y regiones físicamente posibles. El sistema relacional contiene lo que Suppe denomina *sistemas físicos causalmente posibles,* que son los que hacen de modelos teóricos. Una teoría, entonces, determina, a través de alguna de sus formulaciones, una clase de tales sistemas, una clase de modelos. Para su identidad no es esencial la particular formulación sino la clase de modelos.

Mediante la determinación de los sistemas físicos causalmente posibles, la teoría pretende dar cuenta de cierto ámbito de la experiencia, lo que Suppe llama el *alcance pretendido ('intended scope').* Este ámbito de aplicación está constituido por sistemas físicos que ejercen de "datos duros" *("hard" data')* para la teoría. Pero los datos no son en ningún sentido relevante "observables": "Las teorías tienen como su principal objeto los informes de datos duros, no informes de observación directa. [...] La necesidad de una dicotomía observacional/teórico desaparece. La reemplaza la distinción entre datos duros aproblemáticos sobre sistemas físicos y condiciones de entorno y los más problemáticos asertos teóricos acerca de ellos" (1989, pp. 69, 71). Los datos son *relativamente* aproblemáticos en dos sentidos: primero, porque son aproblemáticos *relativamente a una teoría,* aquella teoría para la que son datos; segundo, porque, incluso para la teoría en cuestión, no son *totalmente* aproblemáticos, en caso de contrastación negativa pueden ser problematizados, esto es, revisados. Ello es posible porque los sistemas físicos que presentan los datos son réplicas altamente abstractas e idealizadas de los fenómenos. En la réplica se seleccionan sólo los parámetros del sistema relevantes para la teoría y se abstraen los demás, y los que se seleccionan se idealizan. Por ejemplo (¡bid., *p.* 65), en la determinación del sistema-dato en un caso de caída libre en mecánica se prescinde de parámetros como el color, etc., y otros relevantes como la velocidad se seleccionan en condiciones ideales, como ausencia de rozamiento, masa puntual, etc. La determinación de los datos es pues un complejo proceso de elaboración a partir de los fenómenos, que involucra un gran número de supuestos teóricos en la selección de los parámetros, su medición, la idealización, la determinación de las condiciones de entorno, etc. En ciertas circunstancias puede ser más adecuado revisar este proceso que los postulados teóricos. Quizá se piense que esta caracterización de los datos, *obtenidos a partir de los fenómenos,* abre la puerta trasera a la distinción que se ha abandonado, pues aunque los datos no serían observables, los fenómenos "de los que se extraen" sí lo serían. La distinción volvería a ser fundamental, sólo que un peldaño más abajo. Pero según Suppe no es así. Los fenómenos están constituidos por particulares que poseen ciertas propiedades y que están en ciertas relaciones, pero "estos particulares, sus propiedades y relaciones no necesitan ser observables" (ibid., *p. 93).*

Así caracterizada, una teoría es *empíricamente verdadera* si los datos coinciden con los modelos de la teoría, si los sistemas físicos del alcance pretendido coinciden con *los sistemas físicos causalmente posibles* determinados por la teoría, esto es, si en los sistemas de datos los valores de los atributos son los determinados por la teoría (quizá con ciertas idealizaciones). En realidad esa es una condición sólo necesaria, pues Suppe añade otra condición "antinominalista", que aquí sólo podemos presentar imprecisamente y sin comentario: los parámetros de los sistemas de datos corresponden *a clases naturales* (cf. *ibid., p.* 98; sobre este concepto, cf. *supra,* cap. 5, §2). Suppe coincide con van Fraassen en que la aceptación de la teoría no supone aceptar su verdad, la verdad de *toda ella.* Pero no coincide con aquél en sus motivos. Esta diferencia es la que le permite defender, contra van Fraassen, lo que califica de *cuasi-realismo.* Las teorías, afirma, no dan descripciones literales de cómo funciona el mundo real, sólo pretenden describir cómo *funcionaría* el mundo si los parámetros seleccionados *fuesen* independientes de los desestimados. "Las teorías proporcionan descripciones contrafácticas de cómo *sería* el mundo *si* los parámetros desestimados *no influyesen* en los fenómenos que la teoría pretende describir. Pero típicamente los parámetros desestimados influyen al menos a veces en los fenómenos, y por tanto las caracterizaciones ofrecidas por las teorías no son literalmente verdaderas, sino como máximo contrafácticamente verdaderas, de los fenómenos de su alcance. Ésta es la postura cuasi-realista que he defendido" (ibid., *pp.* 348-349).

4.3. GIERE: MODELOS E HIPÓTESIS TEÓRICAS

Giere desarrolla su propia versión de la concepción semántica en el marco de un programa metacientífico más amplio de análisis de los diversos elementos de la ciencia desde una perspectiva *cognitiva* (cf. especialmente 1988; también 1979, su libro de texto clásico sobre la argumentación científica, con nueva edición muy revisada en 1991). Desde esta perspectiva, propone considerar las teorías como medios para definir modelos abstractos de los que se postula su aplicación a ciertos sistemas reales. "Mi sugerencia preferida es que entendamos una teoría como compuesta de dos elementos: (1) una población de modelos, y (2) varias hipótesis conectando esos modelos con sistemas en el mundo real' (1988, p. 85).

Los modelos ahora no se caracterizan como entidades conjuntistas, ni mediante espacios de estado, ni de ninguna otra forma específica. No se les atribuye una naturaleza matemática determinada. La noción de *modelo teórico* es aquí extremadamente amplia, son entidades abstractas definidas mediante ciertos recursos expresivos, generalmente, pero no necesariamente, lingüísticos (p.ej. se pueden usar grafos o croquis). A veces los modelos pueden ser "modelos a escala" físicamente construidos, como en el caso del modelo de doble hélice de Watson y Crick para el ADN. Pero en general no son así y, lo que es más importante, *en tanto que modelos teóricos* no tienen por qué ser (no cuentan como) entidades físicas. "Un modelo teórico es parte de un mundo imaginado. No existe en ningún lugar excepto en las mentes de los científicos o como sujetos abstractos de las descripciones verbales que los científicos escriben" (1991, p. 26). Por ejemplo, si antes de ir a una fiesta nos "imaginamos" quién viene con quién, estamos determinando, definiendo, una entidad abstracta que es un modelo de (algunos aspectos de) la fiesta; otro ejemplo, el preferido por Giere, son los mapas. "Un modelo es por tanto, como en estos ejemplos, una entidad abstracta y estructurada que representa algo distinto. Los postulados, leyes y ecuaciones que aparecen en los textos científicos *definen* estas entidades. La ecuación "$md^2s/dt' = -kx$" define lo que es un oscilador armónico simple; la ecuación "$md^2s/dt^2 = -(mg/l)x$" define un tipo de oscilador armónico simple, el péndulo sin fricción. Osciladores, péndulos, son por tanto modelos definidos mediante esas ecuaciones, y en tanto que tales son "entidades *socialmente* construidas [y] no tienen realidad más allá de la atribuida a ellas por la comunidad de físicos" (1988, p. 78).

Una vez definidos los modelos teóricos, la teoría formula ciertas *hipótesis teóricas.* Una hipótesis teórica es un enunciado o proposición que afirma cierto tipo de relación entre un modelo y un sistema real determinado (o una clase de sistemas tales). Giere enfatiza que a diferencia de los modelos, las hipótesis teóricas sí son entidades lingüísticas (proposicionales), verdaderas o falsas. La relación que se afirma en la hipótesis teórica no es la de identidad, no se afirma que cierto sistema es el modelo; nótese que los sistemas son entidades físicas y los modelos no lo son, son entidades abstractas. La relación afirmada en la hipótesis es la de *similitud o semejanza.* Pero toda relación de semejanza debe ser cualificada para ser mínimamente precisa. Debe relativizarse a determinados *aspectos y,* en ellos, a cierto *grado.* La forma general de la hipótesis teórica es pues la siguiente: "Tal sistema real identificable es similar al modelo designado en los aspectos y grados indicados" (¡bid., *p.* 81). Es esencial notar que no todos los aspectos del sistema real se desean reflejar en el modelo. En el caso del modelo para nuestra fiesta, no nos interesa quizá el color de las ropas, o incluso la hora de llegada. Lo mismo ocurre en la ciencia, p.ej. en la mecánica no nos interesa el color de los objetos, o incluso a veces tampoco la forma ni el tamaño. Así, las hipótesis contenidas en los textos científicos formuladas en términos identificatorios expresan en realidad afirmaciones de similaridad. Cuando los físicos dicen "la Tierra y la Luna constituyen un sistema gravitacional newtoniano de dos partículas", lo que están afirmando es: "las posiciones y velocidades de la Tierra y la Luna en el sistema Tierra-Luna se aproximan mucho a las de un modelo newtoniano de dos partículas con fuerza central cuadrático-inversa".

Giere desea enfatizar que, en su perspectiva, los enunciados contenidos en la formulación de la teoría no están en conexión directa con el mundo real, sino que se conectan indirectamente con el mundo a través de los modelos. Los enunciados definen los modelos, y los modelos están directamente conectados con el mundo físico a través de la relación de similaridad. Esta relación de similaridad-en-ciertos-respectos-relevantes-y-hasta-cierto-grado es expresada por la hipótesis teórica, que sí es una entidad lingüística. La relación puede darse o no darse; si se da la hipótesis, es verdadera, si no, es falsa. Podría pensarse que la abstracción, aproximación e idealización de la relación de similaridad se pueden reducir, hasta eventualmente eliminarse, mediante la definición de modelos más completos y precisos. Al aumentar los respectos relevantes, disminuye la idealización y se afina la aproximación. Por ejemplo, se puede definir un modelo para el oscilador armónico que incluya la fricción; este modelo incluye un nuevo aspecto para la relación de semejanza, es por tanto menos idealizado y puede aumentarse el grado de semejanza o aproximación a los valores del sistema real. Pero eso sólo reduce o estrecha la semejanza, por lo general no es posible convertirla en correspondencia exacta, en correspondencia entre el sistema y el modelo en todos los aspectos y con una precisión completa.

Una consecuencia de este enfoque es, en opinión de Giere, que las teorías científicas son entidades que no están bien definidas. El motivo es que no está bien determinado, al menos no formalmente, cuáles son los modelos vinculados a una teoría específica, por ejemplo, qué cuenta propiamente como modelo newtoniano. En su opinión, todo lo que se puede decir es que los modelos de la mecánica comparten "un parecido de familia". Según Giere, este parecido es innegable, pero no consiste (sólo) en algo estructuralmente identificable en los modelos. Los modelos por sí solos no muestran en qué consiste dicho parecido. La única determinación posible es en términos sociológicos: "Nada en la estructura de los modelos mismos puede determinar que el parecido es suficiente para pertenecer a la familia. Esta cuestión es decidida exclusivamente por los juicios de los miembros de la comunidad científica en un momento. Eso no quiere decir que haya un parecido objetivo susceptible de ser juzgado correcta o incorrectamente. Lo que quiere decir es que el conjunto de los juicios de los científicos *determina* si el parecido es suficiente. Éste es un aspecto en el que las teorías son no sólo construidas, sino además socialmente construidas" *(ibid., p. 86).*

Giere defiende sobre estas bases cierto tipo de "realismo", que él denomina *realismo constructivista,* que tan sólo podemos enunciar aquí superficialmente. La ciencia tiene un aspecto esencialmente constructivo, la definición de los modelos, y modelos diferentes pueden ser representaciones alternativas de un mismo sistema físico. Hay modelos mejores que otros, pero eso no se puede especificar apelando exclusivamente al mundo. Nada en el mundo mismo fija los aspectos a representar, ni cuán buena es la representación. La especificación debe apelar necesariamente a intereses humanos, y no sólo epistémicos o científicos, sino también a intereses prácticos de diverso tipo. Eso supone una cierta dosis de relativismo, pero no es un relativismo radical: podemos circular por Nueva York, mejor o peor, con dos mapas de Nueva York diferentes, pero no con uno de San Francisco. Este relativismo es compatible en su opinión con cierto realismo, en el sentido de que los modelos representan "hechos del mundo". Pero éste es un sentido muy impreciso asumible por los antirrealistas. Precisarlo requiere al menos dos cosas. Primero, caracterizar más finamente los sistemas físicos "del mundo" de los que se predica su similaridad con los modelos, y lo que dice Giere al respecto sobre los datos es muy poco (cf. 1991, pp. 29-30). Segundo, imponer constricciones claras a la similaridad predicada que permitan, p.ej., decir por qué cierto mapa no es un mapa de Nueva York; ¿acaso un mapa de San Francisco no es similar a Nueva York en *algunos* respectos? Si las únicas constricciones posibles apelan esencialmente a intereses o prácticas humanas, entonces difícilmente se puede calificar esta posición de realista.

4.4. SNEED Y LA CONCEPCIÓN ESTRUCTURALISTA

La concepción estructuralista aúna y desarrolla de un modo específico dos tradiciones anteriores. De un lado, el programa Suppes-Adams de análisis y reconstrucción de teorías mediante el instrumental modeloteórico de la teoría informal de conjuntos. De otro, los trabajos de los historicistas, en especial de Kuhn y Lakatos, donde se analizan las teorías como entidades estructuralmente complejas y susceptibles de evolución, con un "núcleo" central inmutable y un "entorno" complementario cambiante. Ambos elementos se encuentran ya en *The Logical Structure of Mathematical Physics* (1971). Uno de los principales problemas de los historicistas es la vaguedad de sus nociones centrales, que consideraban casi siempre ineliminable. En esta obra, Sneed ofrece ya una primera precisión formal, todavía muy tosca, de esas ideas aplicando el aparato conjuntista de Suppes-Adams. La propuesta de Sneed la recoge Stegnlüller (cf. 1973 y 1979), dando lugar a toda una serie de trabajos que desarrollan las diversas partes del programa y lo aplican a la reconstrucción de un considerable número de teorías científicas. Estos trabajos culminan parcialmente a mediados de los años ochenta con la publicación de *An Architectonic for Science,* de Balzer, Moulines y Sneed, *summa* del programa que contiene sus principales elementos y algunas reconstrucciones de teorías. El programa estructuralista continúa su desarrollo en los años ochenta y noventa, tanto extendiéndose a nuevos ámbitos y problemas metacientíficos como aplicándose a la reconstrucción de nuevas teorías (Balzer y Moulines (eds.) 1996 y 1998 recogen, respectivamente, los principales resultados en ambas tareas).

La concepción estructuralista es, dentro de la familia semántica, la que ofrece un análisis más detallado de la estructura fina de las teorías. En la próxima sección vamos a ver los principales elementos de dicho análisis con cierto detalle. Para concluir ésta avanzaremos tan sólo sus rasgos generales.

a) Se rechaza la distinción "teórico/observacional" y se sustituye por otra, "teórico/no teórico', relativizada a cada teoría.

b) En términos de esa nueva distinción se caracteriza la base empírica y el dominio de aplicaciones pretendidas. Los datos están cargados de teoría pero no de la teoría para la que son datos.

c) Con esta nueva caracterización se da una formulación de la aserción empírica que claramente excluye la interpretación "autojustificativa" de la misma.

d) Se identifican como nuevos elementos en la determinación de los modelos, además de las tradicionales leyes, otros menos manifiestos pero igualmente esenciales, las ligaduras o restricciones cruzadas.

e) Se identifican los vínculos entre los modelos de diversas teorías.

f) Se caracteriza la estructura sincrónica de una teoría como una red con diversos componentes, unos más esenciales y permanentes y otros más específicos y cambiantes. La evolución de una teoría consiste en la sucesión de tales redes.

g) Se analizan en términos modelísticos las tradicionales relaciones interteóricas de reducción y equivalencia.

5. La concepción estructuralista de las teorías

Una teoría tiene, como en la versión de Adams del programa de Suppes, una parte *formal* y otra *aplicativa.* Pero ambas partes se articulan a su vez, como en Kuhn y Lakatos, en diversos niveles de especificidad. Esta idea de los diversos niveles de especificidad se expresa mediante la noción de *red teórica,* que describe en toda su riqueza la estructura sincrónica de las teorías, su imagen "congelada" en un momento dado de su evolución. Las redes están formadas por diversos elementos estratificados según su especificidad. Cada uno de estos elementos tiene una parte formal y otra aplicativa. La parte formal global de la teoría-red queda expresada por el conjunto de las partes formales de los elementos constituyentes; su parte aplicativa global por el conjunto de las partes aplicativas de sus constituyentes. A estos elementos constituyentes se les denomina *elementos teóricos.* La parte formal de los elementos teóricos se denomina *núcleo* y su parte aplicativa, *dominio de aplicaciones pretendidas (o intencionales).*

5.1. EL NÚCLEO *K*

El núcleo, al que denotamos mediante la letra `K', expresa la parte formal de la teoría, las tradicionales leyes. Como en la familia semántica en general, las leyes no se expresan en términos lingüísticos sino modelísticos, entendiendo los modelos, siguiendo aquí a Suppes, como estructuras conjuntistas definidas mediante la introducción de cierto predicado. El núcleo K contiene entonces una serie de modelos, las estructuras que satisfacen los axiomas del predicado. Sin embargo, a diferencia de Suppes y Adams, para el estructuralismo no es adecuado identificar el núcleo con un único conjunto de modelos. Es conveniente que la expresión modelística de la parte formal de la teoría recoja y haga explícitos los diversos elementos distintivos; algunos de ellos ya están implícitos en la caracterización de Suppes, otros sin embargo son nuevos. Para referirnos a ellos vamos a recurrir al ejemplo de Suppes de la mecánica de partículas presentado en la sección 2. Hay algunas diferencias técnicas y de matiz entre esa versión y la estándar en el estructuralismo, pero a los efectos actuales se pueden obviar. Tenga pues el lector de nuevo presente a partir de ahora aquella definición de los modelos de la mecánica.

Modelos potenciales y modelos actuales

Ya vimos entonces que algunos de los axiomas del predicado conjuntista, en ese caso los axiomas (1)-(6), son meras caracterizaciones o tipificaciones de los modelos. Esos axiomas "impropios", *solos,* definen efectivamente entidades o modelos, pero sólo el tipo lógico-matemático de los mismos, por lo que toda estructura de ese tipo será modelo de ellos, *sin importar qué pase después de sustantivo o específico a sus constituyentes.* Los axiomas (7) y (8) no son así, imponen constricciones efectivas adicionales no meramente lógicas, expresan las leyes en sentido propio de las teorías. Eso significa que de todas las estructuras que satisfacen (1)-(6), sólo algunas satisfacen además (7) y (8). Llamaremos *modelos potenciales* (de la teoría en cuestión), y denotaremos su conjunto mediante 'Mp', a las estructuras que satisfacen los axiomas impropios o tipificaciones, y *modelos actuales* (de la teoría en cuestión), y denotaremos su conjunto mediante `M', *a* las estructuras que satisfacen *además* los axiomas propios que expresan constricciones no meramente lógicas. Los modelos potenciales son *potenciales* porque *pueden* ser modelos efectivos de la teoría, porque son las entidades de las que tiene sentido preguntarse si satisfacen o no las leyes propiamente dichas. Aquellos modelos potenciales que, además de las tipificaciones, satisfacen las leyes propiamente dichas son los modelos actuales o efectivos; es inmediato, por tanto, que $M \subset Mp$.

Definición 10.2:

$x \in Mp(MC)$ $syss_{def}$ x satisface (1)-(6) de Def. 10.1.

Definición 10.3:

$x \in M(MC)$ $syss_{def}$ $x \in Mp(MC)$ *y* x satisface (7)-(8) de Def. 10.1.

Es conveniente expresar esta diferencia incluyendo en el núcleo *ambos* conjuntos de modelos. En primer lugar, porque la diferencia expresa un hecho importante, a saber, la diferencia entre la parte meramente conceptualizadora de la teoría, *Mp, y* la parte efectivamente restrictiva, *M.* Pero además, porque los modelos actuales no constituyen la única constricción efectiva de la teoría. Hay otros elementos de la teoría, menos manifiestos, pero igualmente restrictivos, cuya expresión requiere también hacer referencia a los modelos potenciales. Es importante pues tener singularizados los modelos potenciales, el aparato conceptual de la teoría, con relación a los cuales se expresan diversos tipos de restricciones teóricas efectivas. De momento vamos a presentar una, en el último apartado veremos otra.

Condiciones de ligadura

Las restricciones a que nos referimos son lo que el estructuralismo denomina *ligaduras o restricciones cruzadas ('constraints').* La idea es que las leyes usuales no son las únicas que imponen condiciones adicionales efectivas a los modelos potenciales. Si consideramos modelos sueltos, sí, pero si tenemos en cuenta varios modelos a la vez, no. Por ejemplo, según la mecánica clásica no puede ser que una partícula p tenga una masa en un modelo x y otra masa diferente en otro modelo y (por supuesto que la mecánica clásica permite los cambios de masa, por ejemplo "si se quita un trozo" a un objeto, pero se considera siempre que eso corresponde a la generación de otra partícula); por ejemplo, si cierto cohete está en el dominio de dos sistemas, uno el sistema Tierra-cohete y el otro el sistema Luna-cohete, en ambos modelos ha de tener la misma masa. Ésta no es la única constricción intermodélica. La teoría tampoco permite que si un modelo x contiene una partícula p, (p.ej. conductor-más-coche), que es la combinación de dos partículas p_2 (conductor solo) y p_3 (coche solo), haya modelos que asignen a p_2 y p_3 masas cuya suma no coincida con la asignada a p, en x. La primera condición expresa simplemente que la masa de una partícula es constante, y la segunda que la masa es aditiva, esto es, la masa de un compuesto es la suma de las masas de los componentes. Este tipo de condiciones *intermodélicas* son las que permiten "transportar la información" de unos modelos a otros. Si tengo la masa del cohete en el modelo que forma con la Tierra, puedo calcular ciertos valores dinámicos de la Luna gracias a que exporto la información sobre la masa del cohete al modelo que forma con la Luna (cf. cap. 6, §6 sobre la trascendencia de estos hechos para la medición indirecta).

Debe quedar claro que no hay manera de expresar este tipo de constricciones mediante los axiomas usuales, pues éstos se aplican a modelos sueltos. La condición que define la ligadura de identidad para la masa es la siguiente: "para toda partícula p, y modelos potenciales x, y (que tengan a p en su dominio): $m_x(p) = m_y(p)$". Esta condición no es satisfecha o insatisfecha por modelos potenciales sueltos sino por grupos de ellos: si un conjunto tiene dos modelos con una partícula común a ambos dominios y en cada uno la función m asigna a esa partícula valores diferentes, no satisface la condición; si todos los modelos del conjunto asignan a las partículas comunes de sus dominios la misma masa, sí que la satisface. El efecto que tiene esta condición, por tanto, no es determinar un conjunto de modelos, sino un conjunto de conjuntos de modelos; esto es, agrupa los modelos en grupos, grupos tales que, en cada uno, sus modelos asignan a una misma partícula una misma masa; cada grupo se caracteriza porque en él los modelos asignan a cada partícula determinada masa. Una condición que es satisfecha o no por modelos sueltos define un conjunto de modelos, *el* conjunto de los modelos que la satisfacen; éste es el caso de los axiomas (7)-(8). Una condición que es satisfecha o no por conjuntos de modelos, define un conjunto de conjuntos de modelos, *el* conjunto de los conjuntos de modelos que la satisface. Éste es el caso de la ligadura de identidad para la masa. La condición define pues un conjunto de conjuntos de modelos potenciales, al que denotaremos mediante `$C_{=m}$'.

Definición 10.4:

$$C_{=m}(\mathrm{MC}) =_{def} \{X \subseteq Mp(\mathrm{MC}) \,/\, \forall x,y \in X\ \forall p \in P_x \cap P_y : m_x(p) = m_y(p)\}.$$

Debe estar claro que, mientras que $M(\mathrm{MC}) \subseteq Mp(\mathrm{MC})$, $C_{=m}(\mathrm{MC}) \subseteq \mathrm{Pot}(Mp(\mathrm{MC}))$. Análogamente procede la condición de aditividad, que define otro conjunto de conjuntos de modelos. Ahora en cada uno de esos grupos la masa de una partícula compuesta es la suma de la masa de sus componentes, en cualesquiera modelos del grupo en que estén el compuesto o los componentes ('o' denota aquí la composición de partículas).

Definición 10.5:

C o,,,(MC) =d f { X c Mp(MC) / V x, y, z E X V p E P_x V q E P,. V r EP_z (r = p o q --> mz(r) = m.,t(p) + m.,(q)) }

Estas dos ligaduras cuentan por tanto como constricciones efectivas adicionales de la teoría, que, a diferencia de las leyes usuales, no operan a nivel de modelos aislados sino de grupos de modelos, por eso se califican de restricciones *cruzadas.* Como en nuestro ejemplo, puede haber varias ligaduras en una misma teoría, y lo que interesa es tener identificado el efecto combinado de todas ellas. A este efecto combinado o suma de las ligaduras se la denomina *ligadura global y* se denota mediante 'GC'. Puesto que cada fi- gadura es determinado subconjunto { {x, y,, zi, ..J, {x2, Y2, ..J 1 .... } de Pot(Mp), la ligadura global se identifica con su intersección conjuntista, pues los elementos de dicha intersección satisfarán a la vez todas las condiciones de ligadura.

Definición 10.6:

GC(MC) *=,,rfC_,,,(MC) n C.,,,(MC)*

Así, en general, si *C,, ..., C,,* son las *n* ligaduras de una teoría (C; c Pot(Mp)), entonces GC = *C, n ... n C,,. GC* se incorpora pues como un nuevo componente del núcleo *K,* junto con *Mp y M.*

T-teoricidad y modelos parciales

Falta un último elemento para que el núcleo contenga todo lo que es relevante de "la parte formal" de la teoría (último provisionalmente, pues como hemos anunciado en el último apartado haremos referencia a otro). Este elemento tiene que ver con la recurrente cuestión de la teoricidad. El estructuralismo rechaza la distinción "teórico/observacional" por ambigua. Esta distinción esconde en realidad dos: "observable/inobservable" de un lado, y "no teórico/teórico" de otro. Ambas distinciones no coinciden intensionalmente *ni* extensionalmente. La primera distinción no tiene relevancia alguna para el análisis *local* de la estructura de las teorías (aunque por supuesto es relevante para la cuestión general de cómo se relaciona el conjunto de las teorías con la observación). Para el análisis local de la estructura de las teorías la distinción relevante es la segunda, pero en este caso no se trata ya de una distinción absoluta, sino que está relativizada a las teorías. Un término, o un concepto, o una entidad, no es teórico o no teórico sin más, sino *relativamente a una teoría dada.* Por eso no se debe hablar tanto de teoricidad cuanto de T-teoricidad, teoricidad relativamente a una teoría T. La idea que hay detrás es, expresada en términos modeloteóricos, similar a la distinción que vimos en el último Hempel entre vocabulario antecedente y vocabulario propio (aunque formulada ya con anterioridad en la obra fundacional del estructuralismo, Sneed, 1971). La idea es que un concepto es T-teórico si es un concepto *propio* de la teoría *T*, "introducido' por ella, y es T-no teórico si es un concepto disponible previamente a T. La cuestión es precisar esta intuición.

La formulación precisa del criterio de T-teoricidad usa de la noción técnica de *procedimiento de determinación*, que no podemos presentar aquí en detalle. Bastará de momento con la siguiente caracterización informal. Como vimos en el capítulo 4, los conceptos se aplican o no a las cosas, o si son cuantitativos, asignan valores a ciertas cosas. Determinar un concepto es determinar si se aplica o no a un objeto particular dado, o si es cuantitativo, determinar el valor de la magnitud para el objeto. Los modos para proceder a ello son los procedimientos de determinación de los conceptos. Puedo determinar la distancia entre la Tierra y la Luna haciendo ciertos cálculos a partir del período de rotación y las masas correspondientes. Puedo determinarlo también mediante ciertos procedimientos óptico-geométricos. Puedo determinar la masa de un objeto mediante una balanza de brazos. También mediante una balanza de muelle. O viendo cuánto se desplaza otra masa tras chocar con ella a cierta velocidad. Todos ellos son procedimientos de determinación, unos de la distancia, otros de la masa, etc. Pues bien, si un concepto es T-no teórico, si es "anterior" a T, entonces tendrá al menos algunos procedimientos de determinación *independientes* de *T;* en cambio si es T-teórico, si es propio de *T*, su de- terminación *depende siempre* de *T.* Un procedimiento de determinación se considera dependiente de la teoría *T* si presupone la aplicabilidad de *T*, la validez de sus leyes, esto es, si usa o presupone modelos actuales de T. La idea es que un concepto es T-teórico si no se puede determinar sin presuponer la aplicabilidad de *T*, si *todo* procedimiento para su determinación la presupone; y es T-no teórico si tiene *algún* procedimiento de determinación T-independiente, si es posible determinarlo sin suponer la aplicación de la teoría, por más que también tenga otros T-dependientes.

En el caso de la mecánica que venimos usando como ejemplo, la posición es MC-no teórica. Es cierto que, como ilustra el caso de la distancia Tierra-Luna, se puede determinar por procedimientos que usan las leyes de la mecánica, como el efecto gravitacional, pero también se puede determinar sin usar leyes mecánicas, por procedimientos óptico-geométricos. Lo mismo ocurre con el tiempo o duración. Sin embargo no ocurre así con la masa: todos los procedimientos de determinación de esta magnitud presuponen la aplicabilidad de la mecánica, usan modelos mecánicos. Ello es obvio de los procedimientos de medición indirectos (mediante dinamómetro, o a través de la alteración en la trayectoria de otro cuerpo, etc.). Pero también lo es respecto de la medición directa mediante balanza, pues a menos que se considere que la balanza satisface ciertas leyes mecánicas no se puede considerar que lo que se mide es *la masa de la que habla la mecánica* (cf. cap. 6, §7). Faltaría más, se dirá, la masa *es* un concepto mecánico. Pues bien, eso es justamente lo que queríamos, precisar el sentido exacto en que lo es, en que es un concepto "propio de" o "introducido por" la mecánica. En eso consiste la distinción "T-teórico/T-no teórico". En el caso de la mecánica clásica de partículas, *espacio y tiempo* son MC-no teóricos, conceptos cinemáticos previos, *masa y fuerza* son conceptos MC-teóricos, los conceptos propiamente mecánicos, dinámicos. Es probable que para todo concepto T-no teórico haya otra teoría T' respecto de la cual el concepto sea T'-teórico, pero eso es una hipótesis metaempírica que se debe confirmar.

La noción de T-teoricidad permite precisar el último componente del núcleo. Hemos visto que los modelos potenciales expresan el aparato conceptual de la teoría. Es conveniente ahora distinguir en el núcleo entre el aparato conceptual global de la teoría y el aparato conceptual específico de ella. Esto es, distinguir los modelos que usan todo el aparato conceptual de la teoría de aquellos que usan sólo conceptos previamente disponibles, en esa diferencia radica la contribución conceptual específica de la teoría (además de para estas consideraciones generales, la necesidad de distinguir entre ambos tipos de modelos se hará patente cuando discutamos la base empírica). La determinación de esos modelos que no contienen el aparato específico de la teoría es sencilla una vez se dispone de la noción de T-teoricidad presentada, pues tales modelos contienen como constituyentes exclusivamente las entidades correspondientes a los conceptos T-no teóricos; esto es, estos modelos se obtienen a partir de los modelos potenciales "recortando" de ellos las entidades T-teóricas. A estos modelos se les denomina *modelos (potenciales) parciales*, y se denota su conjunto mediante *'Mpp'*. Así, en general, se puede definir una *función recorte r* que genera los modelos parciales a partir de los potenciales. Si los modelos potenciales de T son estructuras del tipo $x = \langle D_1, ..., D_k, ..., R_1, ..., R_n, ..., R_m \rangle$ y $R_{n+1}, ..., R_m$ son T-teóricos, entonces $r(x) = \langle D_1, ..., D_k, ..., R_1, ..., R_n \rangle$. El conjunto *Mpp* de los modelos parciales es entonces simplemente el conjunto de los modelos potenciales una vez que hemos recortado de ellos las funciones T-teóricas: $Mpp =_{def} \{y \,/\, \exists x \in Mp : y = r(x)\}$ o, abreviadamente, $Mpp =_{def} r[Mp]$, donde `r[...]' denota la función recorte aplicada a conjuntos de modelos (recuérdese, cf. Apéndice, que *r[X]* es el recorrido de *r* restringido a X, en este caso el conjunto formado por los modelos de X una vez recortados). En nuestro ejemplo, los modelos parciales de la mecánica son entidades del tipo $\langle P, T, s \rangle$, que no contienen parámetros MC-teóricos, contienen sólo parámetros cinemáticos; mientras que los modelos potenciales $\langle P, T, s, m, f \rangle$ incluyen además los parámetros dinámicos, los propiamente mecánico-teóricos.

Definición 10.7:

$$Mpp(MC) =_{def} \{ \langle P, T, s \rangle \,/\, \exists\, m, f : \langle P, T, s, m, f \rangle \in Mp(MC) \}.$$

Con ello concluimos la presentación del núcleo, la parte formal de los elementos teóricos. El núcleo *K* se expresa mediante la tupla $\boldsymbol{K} = \langle Mp, Mpp, M, GC \rangle$, donde *Mp* es el conjunto de modelos potenciales, *Mpp* el de los modelos parciales $(Mpp = r[Mp])$, *M* el de los modelos actuales $(M \subseteq Mp)$ y GC la ligadura global $(GC \subseteq \text{Pot}(Mp))$.

5.2. APLICACIONES INTENCIONALES

El núcleo *K* es el componente formal de la teoría, pero no el único. Como hemos visto en general en las concepciones semánticas, las teorías *empíricas* pretenden que las constricciones de K lo son de ciertas *partes de la realidad física,* los sistemas empíricos a los que se pretende aplicar el núcleo. Estos sistemas empíricos se denominan en el estructuralismo, como en Adams, *aplicaciones pretendidas o intencionales ('intended applications'), y* se denota su conjunto mediante `I'. En nuestro ejemplo de la mecánica clásica, son aplicaciones pretendidas cosas como el sistema Tierra-Luna, el sistema Solar, un trapecista en su balancín, dos bolas de billar chocando, una balanza, un esquiador deslizándose por una pendiente, un niño saltando en una colchoneta elástica, un satélite de comunicaciones en órbita, etc.

La caracterización estructuralista de los dominios de aplicaciones contiene sin embargo elementos específicos, especialmente los dos siguientes. En primer lugar, las aplicaciones pretendidas de una teoría T se individualizan y describen mediante el vocabulario previo a T, esto es, mediante el aparato conceptual T-no teórico. Así, en los ejemplos mecánicos mencionados, la descripción de las aplicaciones incluye exclusivamente valores de las magnitudes *posición y tiempo,* es decir, son descripciones de los sistemas en términos puramente cinemáticos que presentan sus trayectorias espaciales a lo largo del tiempo. Por tanto, las aplicaciones pretendidas que conforman la base empírica de la teoría, los "datos" de la teoría, ciertamente están cargados de teoría, pero no de la teoría para la que son datos sino, en línea con las observaciones de Lakatos, de otra previa o antecedente. Los datos de la mecánica, a los que se pretende aplicar y sobre los que se contrasta, están *cinemáticamente* cargados, pero *no dinámicamente* cargados. Esto es esencial para dar cuenta del carácter no autojustificativo de la aserción empírica mediante la que se contrasta la teoría. Formalmente, ello se traduce en que cada aplicación pretendida es un determinado sistema que contiene exclusivamente entidades T-no teóricas. Cada aplicación pretendida es entonces un determinado *modelo parcial y* el conjunto *I* de todas ellas es por tanto cierto subconj unto de *Mpp: I c Mpp.*

El segundo hecho a destacar (parcialmente apuntado por Adams y, dentro de los historicistas, por Kuhn) es que la selección de las aplicaciones, la determinación de I, contiene elementos pragmáticos ineliminables, pues tal determinación es esencialmente *intencional y paradigmática.* La determinación es intencional porque lo que hace de un sistema específico que sea una aplicación pretendida es que sea un objeto intencional de los usuarios de la teoría, que la comunidad científica *pretenda* que las constricciones-leyes se aplican a tal sistema (cf. más arriba §3). Y es paradigmática porque el conjunto I no se presenta "listando" todos y cada uno de los sistemas físicos que son aplicaciones pretendidas, sino "paradigmáticamente". No sólo es una aplicación pretendida de la mecánica un cierto esquiador deslizándose por una pendiente determinada en cierto momento específico, sino cualquier esquiador en cualquier pendiente en cualquier momento; y, por supuesto no sólo los esquiadores, también los ciclistas, y los niños bajando por las barandillas, etc. Para determinar el dominio I no hemos de listar todos y cada uno de los sistemas cinemáticos particulares de plano inclinado, sino algunos paradigmáticos y añadir: "y cosas como ésas"; o, alternativamente si se prefiere, referirse de modo general y relativamente impreciso a "todos los sistemas en que un objeto desciende por una superficie inclinada". Y lo mismo con los objetos vibrantes, con las órbitas estacionarias, con los objetos chocando y separándose después, con los objetos chocando y siguiendo unidos después, etc. Esto sugiere que quizá sería mejor caracterizar al dominio de aplicaciones *I,* no simplemente como un conjunto de aplicaciones sueltas *(I (-- Mpp),* sino como un conjunto de conjuntos de aplicaciones *(I c Pot(Mpp))* que tiene por elementos conjuntos que son *grupos de aplicaciones de un mismo tipo.* Pero aquí no vamos a introducir esta complicación (para un estudio detenido de la misma, cf. Moulines, 1982, cap. 2.4) y seguiremos considerando la versión más sencilla según la cual los elementos de I son directamente las aplicaciones individualmente consideradas.

5.3. LAS TEORÍAS COMO ELEMENTOS TEÓRICOS. CONTENIDO Y ASERCIÓN EMPÍRICA

Elementos teóricos

Ahora podemos presentar ya la noción estructuralista mínima (y provisional) de teoría, la noción de *elemento teórico.* Un elemento teórico, una teoría en este sentido mínimo, está constituido por (1) una parte formal que expresa los recursos conceptuales a diferentes niveles y las constricciones-leyes que según la teoría rigen su ámbito de estudio, y (2) una parte aplicativa que especifica en términos preteóricos los sistemas físicos a los que la teoría pretende aplicarse, de los que pretende que son regidos por sus constricciones-leyes. Haciendo uso del aparato previamente introducido, un elemento teórico T se identifica entonces con el par formado por el núcleo *K,* la parte formal, y el dominio de aplicaciones *I,* la parte aplicativa: *T = <K, I>.*

Ésta es la noción más simple de teoría, y, como veremos, resulta parcialmente inadecuada por su "rigidez", pero ya es suficientemente rica y útil para.expresar de modo preciso la naturaleza de la aserción empírica de una teoría. Para ello es conveniente presentar primero la noción de *contenido* de una teoría.

Contenido teórico y contenido empírico

Hemos visto que el núcleo *K* expresa la parte matemático-formal de la teoría. Es en ella donde se presentan las condiciones que, según la teoría, rigen las "partes de la realidad" de que ella trata. Estas condiciones consisten básicamente en las leyes propiamente dichas de un lado, y las condiciones de ligadura de otro, que en el núcleo se corresponden, respectivamente, con los conjuntos M y GC. Sin embargo, la teoría, al aplicarse, no pretende que estas condiciones rigen aisladamente o separadas, sino que las aplicaciones satisfacen todas las restricciones a la vez, tanto las leyes como las ligaduras. Es conveniente entonces "juntar" ambos tipos de condiciones, presentar su efecto restrictivo conjunto. Esto se expresa mediante la noción de *contenido teórico,* a la que nos referiremos mediante `Con,'. El contenido teórico, esto es, el efecto combinado de leyes y ligaduras, queda representado mediante la apropiada intersección conjuntista de los conjuntos M y *GC.* Como *M* es un conjunto {x,, x_2, x_3, ..., x_{9i} ..., $x_{,5}$, ...} de determinados modelos potenciales *(M c Mp) y GC* es un conjunto ({X,, x_2, x_5, ...}, {x_4, x_7, **x9, ... }, ..., {..., x,5, ... } }** de conjuntos de modelos potenciales (GC c Pot(Mp)), la intersección apropiada correspon-

diente a la combinación de ambos tipos de condiciones no es la de GC con M, sino la de GC con Pot(M), esto es: Con, = $_{def}$ *Pot(M) n GC. Es inmediato que Con, C Pot(Mp), el contenido teórico de T, es un conjunto de conjuntos de modelos potenciales, el conjunto cuyos elementos son conjuntos tales que: (1) satisfacen las ligaduras; y (2) están formados por modelos que satisfacen las leyes de la teoría, los axiomas propios del predicado conjuntista.*

La noción central para expresar la aserción empírica es la de contenido empírico, que se deriva de la de contenido teórico. El contenido empírico es el "contenido contrastacional"; en la versión tradicional, las consecuencias empíricas de la teoría. En nuestros actuales términos, las consecuencias empíricas del contenido teórico, el efecto a nivel empírico, esto es, T-no teórico, de las condiciones restrictivas de la parte formal de la teoría. El contenido empírico recoge entonces los (conjuntos de) modelos parciales que resultan de recortar los componentes T-teóricos de los modelos potenciales que satisfacen las restricciones. O de otro modo, los modelos parciales que es posible aumentar con componentes T-teóricos de forma que se cumplan las restricciones (y si las restricciones son efectivamente tales, no todo modelo parcial es aumentable de esta forma). Así, si denotamos mediante `Con' el conjunto que expresa el contenido empírico, dicho conjunto es el resultado de recortar los componentes T-teóricos en los modelos que aparecen en Con,, abreviadamente: Con = r[[Con,]] (r(...) se aplica a modelos sueltos, r[...] se aplica a conjuntos de modelos, r[[...]] es la función recorte aplicada a conjuntos de conjuntos de modelos, como Con,).

Aserción empírica

Ahora podemos expresar de modo preciso la naturaleza que, según el estructuralismo, tiene la aserción empírica de una teoría. La teoría pretende que ciertos sistemas físicos, T-no teóricamente descritos, satisfacen las condiciones impuestas por la teoría en el sentido siguiente: ésos son los datos de experiencia que se deberían obtener si la realidad operase como la teoría dice. Esta pretensión se expresa en la aserción empírica de la teoría. Por todo lo anterior debe ser claro que la forma lógica que corresponde a la aserción es "I E Con", esto es, el dominio de aplicaciones pretendidas I es uno de los conjuntos de modelos parciales, T-no teóricos, que las constricciones del núcleo K determinan a nivel empírico. Ésta es la versión modeloteórica precisa de la idea intuitiva de que las aplicaciones pretendidas satisfacen individualmente las leyes y, además, satisfacen colectivamente las condiciones de ligadura. Mejor dicho, no que "ellas mismas" satisfacen esas condiciones, pues ellas son estructuras T-no teóricas y tales condiciones involucran esencialmente constituyentes T-teóricos de los modelos. La aserción afirma que ciertos sistemas empíricos concretos, descritos T-no teóricamente, tienen el comportamiento que las restricciones legales determinan a nivel T-no teórico. Tomemos un sistema empírico que se comporta de cierto modo según ciertos parámetros T-no teóricos. Que la aserción sea cierta significa que ése es justamente el modo en que le corresponde comportarse si están presentes en él los parámetros T-teóricos que la teoría postula y éstos se relacionan con los T-no teóricos de la forma que establecen las leyes. Es decir, los sistemas de I son mo-

delos parciales que pueden ampliarse con funciones T-teóricas de modo que se obtengan modelos que satisfacen aisladamente las leyes y conjuntamente las ligaduras. En este sentido, la aserción afirma que la experiencia es *subsumible o encaja* en la teoría.

Aplicada al ejemplo de la mecánica, la aserción entendida en estos términos expresa de modo sucinto lo siguiente: los sistemas físicos particulares intencionalmente seleccionados (planos, péndulos, muelles, poleas, órbitas, etc.) son tales que sus valores cinemáticos (posiciones, velocidad y aceleración en ciertos instantes) coinciden con los que deberían tener si en los sistemas estuvieran además presentes ciertos parámetros dinámicos (masas, fuerzas) interactuando con los cinemáticos del modo especificado en la mecánica, esto es, *a)* del modo que especifican el segundo principio de Newton y la ley de acción y reacción, *y b)* manteniendo la misma masa para las partículas que aparecen en diversas aplicaciones y respetando la aditividad de las masas cuando una partícula esté compuesta de otras (sean cuales sean las aplicaciones en que aparezcan).

Es importante darse cuenta de que, aunque la experiencia o los datos están "cargados de teoría", eso no tiene consecuencias autojustificativas para la aserción. Se seleccionan intencionalmente ciertos sistemas físicos. Primero, se hacen ciertos cálculos suponiendo que en los sistemas está actuando todo lo que postula la teoría y del modo como ella establece. Segundo, *e independientemente,* se determinan en los sistemas los valores de ciertas magnitudes cuya medición no presupone la aplicación o validez de la teoría. Por último, se comprueba si esos valores coinciden con los calculados. No hay autojustificación en absoluto (al menos en sentido local). La aserción puede ser perfectamente falsa, lo es si los valores simplemente no coinciden.

Esta caracterización de la aserción es parcialmente insatisfactoria por excesivamente rigurosa. Pretende que los valores coincidan exactamente, en cuyo caso toda aserción resulta falsa, pues siempre hay errores de aproximación. Ésta es en realidad una versión exacta o idealizada de la aserción, versión que no se corresponde con las pretensiones reales en la actividad científica. Los científicos nunca pretenden la coincidencia plena, sino el acuerdo aproximado con los datos dentro de ciertos límites. Para reflejar este hecho el estructuralismo ofrece una versión modificada de la aserción empírica que recoge los aspectos aproximativos indicados. No vamos a presentarla aquí (cf. p.ej. Moulines, 1982, cap. 2.7), para la idea central basta con la versión idealizada.

5.4. ESPECIALIZACIÓN. LAS TEORÍAS COMO REDES TEÓRICAS

Los elementos teóricos expresan la estructura sincrónica de las teorías sólo parcialmente, pues hay un aspecto estructuralmente relevante a nivel sincrónico que ellos no recogen. Se trata de un aspecto que, como vimos, enfatizaban especialmente Kuhn y Lakatos con la idea de que las teorías contienen partes esenciales o inamovibles donde descansa su identidad y partes más accidentales que pueden perderse o modificarse permaneciendo, en un sentido diacrónico relevante, la misma teoría. Para capturar y formular en términos precisos esta idea, el estructuralismo ha desarrollado el concepto de *red teórica,* que expresa la naturaleza sincrónica de las teorías en toda su riqueza estructural, y que el propio Kuhn ha reconocido que es una buena precisión semiformal de sus matrices disciplinares en cierto momento de su evolución (cf. Kuhn, 1975).

Especialización

Una *red teórica* es un conjunto de elementos teóricos que guardan cierta relación entre sí. La idea es que el conjunto represente la estructura (sincrónica) de una teoría en sus diferentes estratos, esto es, en sus diversos *niveles de especificidad.* Tal conjunto, partiendo de elementos muy generales, se va concretando progresivamente en direcciones diversas cada vez más restrictivas y específicas, las "ramas" de la teoría-red. La relación que se ha de dar entre los elementos teóricos para considerar el conjunto una red ha de ser de "concreción" o "especificación" o, como se dice en terminología estructural, una *relación de especialización.* Podemos ilustrar esta situación con el ejemplo de la mecánica que hemos venido manejando. Volvamos a la definición de los modelos de la mecánica tal como vimos que la presentaba Suppes. Suppes exige que los modelos actuales de la mecánica satisfagan tanto el axioma (7), el segundo principio de Newton, como el (8), el principio de acción y reacción. Desde un punto de vista histórico eso es correcto, si por mecánica entendemos mecánica *newtoniana,* esto es, la que concibió y en la que creía Newton. Pero desde un punto de vista estructural, la estrategia es inadecuada. El segundo principio y la ley de acción y reacción no están al mismo nivel, y es importante que este hecho se refleje en la estructura de la teoría. En contra de lo que creía Newton, no todo sistema que se ajusta a su segundo principio satisface además esa ley de acción y reacción. Hay sistemas mecánicos que satisfacen el segundo principio y que sin embargo son "no newtonianos", en el sentido de que incumplen dicha ley, por ejemplo sistemas que incluyen partículas moviéndose en un campo electromagnético (aunque este hecho queda algo oscurecido en la versión, como advertimos, técnicamente imperfecta que dimos de la ley). Así, mientras todo sistema mecánico satisface (7), no todos ellos satisfacen (8), sólo lo hacen algunos de ellos. Los modelos actuales que satisfacen (8) además de (7) son una *especialización* de los que sólo satisfacen (7). Los modelos actuales más generales de la mecánica son los que satisfacen (7). A partir de ahí se pueden abrir varias líneas de especialización. Algunos satisfarán además (8). Otros no satisfarán (8) pero satisfarán otro u otros principios específicos, etc. Y esto puede pasar también en niveles inferiores. Por ejemplo, no todos los sistemas de acción y reacción satisfacen otros principios adicionales. Unos satisfarán el principio de las fuerzas cuadrático-inversas de la distancia, otros el principio de oscilación armónica, etc. A partir del segundo principio, general, la mecánica clásica se va especializando en diversas direcciones específicas imponiendo progresivamente condiciones adicionales en diversas direcciones con la intención de dar cuenta de aplicaciones específicas.

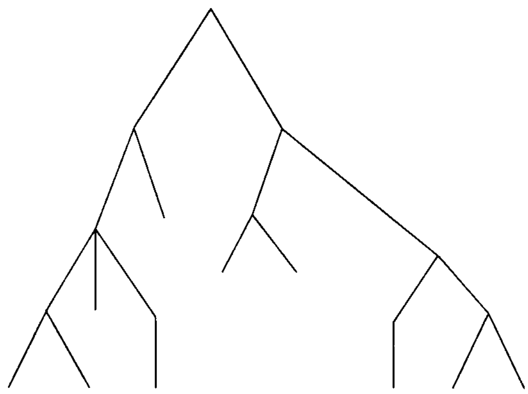
Éste es el panorama que pretende recoger y expresar la noción estructuralista de *red teórica. El* primer paso es definir de modo preciso la relación de especialización. Un elemento T' es una especialización de otro T si la parte formal (las constricciones) de T' es una concreción de la de T y está destinada a dar cuenta de una parte de las aplicaciones pretendidas de T. En términos modeloteóricos, ello significa lo siguiente: (1) los modelos determinados por las constricciones (leyes y ligaduras) del núcleo *K'* son parte de los determinados por *K,* esto es, los correspondientes conjuntos *M'* y GC' de *K'* están incluidos respectivamente en *M y* GC de *K* (pues se van imponiendo condiciones adicionales), mientras que la parte conceptualizadora de los elementos teóricos, los conjuntos *Mp y Mpp,* queda igual; y (2) las aplicaciones de *I'* son algunas de las de *I.* La definición es pues la siguiente, donde `*T' a T'* abrevia `*T'* es una especialización de *T': T'* a T syss *f(1) M'p = Mp, M'pp = Mpp, M' c M, GC' c GC y (2) I'c I.* Como puede verse, la relación de especialización es reflexiva, antisimétrica y transitiva, esto es, de orden parcial (no estricto).

Redes teóricas

Con la noción de especialización disponible podemos precisar la noción de *red teórica.* Una red teórica *N* es simplemente un conjunto de elementos teóricos (parcialmente) ordenado por una relación de especialización: **N =** <{*T,*}, *a*> es una *red teórica* syss*def* (1) {T;} es un conjunto no vacío de elementos teóricos y (2) a es una relación de especialización sobre {T}. A cada red le corresponde un conjunto I_N de aplicaciones pretendidas, la unión de los dominios *I;* de los elementos *T;* que la constituyen.

Mediante el concepto de red teórica se captura la estructura de *una teoría en un momento dado* en toda su complejidad; este concepto expresa adecuadamente la naturaleza de las teorías desde un punto de vista sincrónico o estático. Sin embargo, el concepto es en cierto sentido demasiado débil, pues, al no exigir a a condiciones adicionales, se acepta (como en todo orden parcial) la posibilidad de que haya órdenes "extraños", con partes desconectadas entre sí, esto es, de que partes de una teoría estén totalmente aisladas de otras. El estructuralismo, que adopta por lo general una postura lo más liberal posible, considera que ello no es *conceptualmente* insatisfactorio. Se reconoce que en las teorías conocidas no ocurre *de hecho* tal cosa, pero se considera que se trata de una cuestión *(meta)empírica* que no hay que prejuzgar *a priori.* Aunque en parte es una cuestión abierta, se opta en general por limitarse a la versión débil y definir después un *tipo* de redes-teorías, las *conectadas,* constatando como cuestión *de hecho* que las teorías conocidas son de ese tipo. Una red conectada es una red "no degenerada", sin partes aisladas. Para ello no es necesario exigir que a sea conexa en el sentido lógico usual, esto es que cualesquiera dos elementos diferentes estén relacionados; eso daría lugar a un orden lineal, identificando, contra lo que se pretende, las redes conectadas con redes de una sola línea de especialización. Hay que exigir algo más débil, que sea una "malla", que siempre haya un camino que conecte dos elementos cualesquiera. Formalmente ello se garantiza si podemos "circular-vía-d' entre cualesquiera dos elementos de la red: *N* = <*{Ti},* CY> es una red teórica *conectada syss,,ef* para todo *T, T'* E *{T}* hay *T,, ..., T,,* E *{Ti}* tales que *(T a T,oT,aT)y(T,aT$_2$ oT$_2$ aT,)y...y(T,,.,aT,oT,aT,,.,)y(T'aT,,oT,,aT').*

Un tipo especialmente interesante de redes (conectadas) son aquellas que presentan un único elemento superior, del cual "emana todo". Estas redes (que tienen forma de pulpo o de árbol invertido) se caracterizan formalmente por tener algún elemento teórico del que todos son especializaciones (es inmediato que si tiene alguno, tiene sólo uno). El estructuralismo llama *arbóreas* a tales redes: *N* = <*{T;},* a> es una red teórica *arbórea* syss$_{def}$ hay $T \in \{T_i\}$ tal que para todo $T' \in \{T_i\}$ $T' \sigma T$. Las teorías arbóreas son especialmente interesantes pues en ellas, por así decir, la "esencia" está concentrada en un único elemento teórico básico. El siguiente gráfico ilustra esta situación.



Las redes arbóreas reflejan parcialmente la imagen de la ciencia que se desprende de los análisis de Kuhn y Lakatos, parcialmente porque faltan por ver los aspectos diacrónicos, el tipo de evolución de las redes que constituye la ciencia normal kuhniana. Resumamos cuáles son los principales elementos estructurales sincrónicos descubiertos por los historicistas que son recogidos en la noción estructuralista de red teórica. Las teorías tienen, en los elementos teóricos de la red, un componente formal, el núcleo K, y otro aplicativo, el dominio I de aplicaciones pretendidas. Una parte del núcleo, Mpp, conceptualiza la *experiencia, los* hechos, esto es, $I \subseteq Mpp$. Otra parte explica lo así conceptualizado, explicación que introduce aparato conceptual nuevo propio de la teoría (Mp): las leyes M ($\subseteq Mp$) y ligaduras GC ($\subseteq$ Pot(Mp)) intentan "subsumir" las aplicaciones, pretensión expresada por la aserción empírica de la teoría. Así, los hechos a explicar están cargados de teoría, pero no de la parte de la teoría que pretende explicarlos. El núcleo, que en sí mismo es puramente formal, se carga entonces de contenido empírico al aplicarse-a-las-aplicaciones. Además, todo esto no ocurre de modo "rígido", como en un bloque indiferenciado. Las redes tienen partes esenciales (si son arbóreas, concentradas en un elemento teórico básico) cuyo componente formal es por lo general muy débil, muy poco o nada restrictivo *en sí mismo* (sin especializarlo), y partes accidentales que desarrollan mediante σ la parte esencial especializándola en diversas direcciones, tanto en su componente formal, imponiendo restricciones más fuertes, como en el aplicativo.

5.5. VÍNCULOS INTERTEÓRICOS Y HOLONES

Concluiremos señalando brevemente un último componente de la concepción estructuralista de las teorías que hemos obviado hasta aquí para simplificar la exposición. Este componente pretende dar cuenta de un hecho usual y esencial de la ciencia, a saber, que las teorías no son entidades aisladas sino que mantienen estrechas relaciones entre sí. Algunas de esas relaciones se expresan mediante "leyes mixtas" o "leyes puente", mediante postulados que involucran conceptos de diversas teorías. Las teorías mantienen pues *vínculos interteóricos.* En principio los vínculos pueden relacionar varias teorías a la vez, pero lo usual parece ser que relacionen dos teorías; en todo caso nos limitaremos aquí a este caso, el más sencillo, para evitar complicaciones de una presentación generalizada a cualquier número de teorías. Un ejemplo típico de vínculo interteórico binario lo constituye el que se da entre la hidrodinámica y la termodinámica expresado en la ecuación "P=dE/dV' que relaciona presión, volumen y energía, siendo la presión una magnitud específicamente dinámica y la energía una magnitud específicamente termodinámica.

Los vínculos interteóricos tienen, como las leyes propias de la teoría, efectos restrictivos sobre los modelos, pero a diferencia de ellas no son satisfechas o insatisfechas por modelos potenciales de una única teoría sino por *pares* (en el caso de los vínculos binarios) de modelos potenciales de teorías diferentes. Las leyes propias determinan un subconjunto de modelos potenciales, aquellos que las satisfacen (e.e. los modelos actuales). Los vínculos interteóricos no determinan *directamente* un subconjunto de modelos potenciales de una teoría. Si *Mp y Mp'* son respectivamente los conjuntos de modelos potenciales de dos teorías *T y* T', entonces el producto cartesiano *Mp x Mp'* contiene todos los pares posibles de modelos de ambas. Pues bien, dado un determinado principio puente entre T y T', sólo algunos de esos pares satisfarán dicho principio, por lo que se puede considerar que el principio en cuestión determina o define cierto subconjunto *L* de *Mp x Mp',* el conjunto de pares de modelos que lo satisfacen. Por tanto, los principios puente determinan *primariamente* conjuntos de pares de modelos. Pero eso supone una restricción efectiva adicional para cada una de las teorías, tiene como efecto la determinación de cierto subconjunto de modelos potenciales en cada una de las teorías: para *T* ese conjunto es el de los primeros miembros de los pares de *L*, para *T'* es el de los segundos miembros de los pares. Denotemos mediante `L_T' al conjunto de modelos potenciales de *T* determinado-en-T por el principio puente *L (y* análogamente con T'). Pues bien, si el principio es efectivamente restrictivo, L_T será un subconjunto propio de *Mp.* Como *T* puede tener varios vínculos interteóricos *L;* con diversas teorías, cada uno de ellos determina de este modo indirecto un cierto subconjunto $L;_T$ de modelos, que representa el efecto constrictivo del vínculo en la teoría T. El efecto combinado o conjunto de todos los vínculos se recoge entonces en la intersección de todos esos conjuntos, el *vínculo global* que se denota mediante 'GL', y que es la intersección de todos los vínculos $L;_T$ para *T.*

Una caracterización completa del núcleo *K* que exprese *todas* las condiciones que la teoría impone a los modelos debe incluir también este tipo de constricciones derivadas de las leyes puente. Así, hay que completar la anterior caracterización provisional del núcleo con este nuevo elemento: ***K*** = *<Mp, Mpp, M, GC, GL>;* el contenido teórico es entonces *Con,* = Pot(M) n GC n Pot(GL). Nótese que si no se incluyesen en la caracterización de las teorías este tipo de leyes-restricciones empíricas no aparecerían en la reconstrucción de ninguna teoría y por tanto "desaparecerían" en una eventual reconstrucción total de la ciencia resultante de reconstruir todas y cada una de las teorías. El motivo es que estas leyes empíricas no se formulan con el vocabulario *exclusivo* de una única teoría, involucran conceptos de diferentes teorías, y por ello no aparecen como axiomas propios que determinan los modelos actuales. Pero no por ello son menos constrictivos empíricamente, sor tan parte de lo que la teoría afirma de la experiencia como los axiomas propios de cada teoría y por tanto deben hacerse manifiestos en la reconstrucción de cada teoría. Nótese que no sería una buena estrategia alternativa "ampliar" por este motivo los conceptos propios de la teoría incluyendo cualquier concepto con el que se vinculen mediante leyes "los originales". Eso permitiría recoger las leyes-puente como axiomas propios pero al precio de unificar inaceptablemente diferentes teorías. Si la energía debiera incluirse como magnitud propia de la hidrodinámica por su relación con la presión mediante la ley mencionada, entonces también debería incluirse la entropía, dadas las leyes que la conectan con la energía. Pero eso tendría la consecuencia inaceptable de convertir la hidrodinámica y la termodinámica en una misma teoría. Y no sólo a ellas, sino que convertiría en la misma teoría gran número de teorías físicas (dadas las conexiones de hidrodinámica y termodinámica con otras), o quizá todas las teorías físicas o, si por distintos caminos se conectan con otras disciplinas, todas las teorías empíricas.

Es obvio que no todas las teorías dentro de una disciplina, o de toda la ciencia, son la misma teoría. Por tanto, considerar los vínculos interteóricos exactamente del mismo modo que los axiomas propios de las teorías es inaceptable. Lo adecuado es reconstruirlos, e incluirlos en la caracterización de las teorías, como lo que son, a saber, leyes puente que vinculan teorías diferentes. Su existencia genera cierto tipo de "unidad", pero no puede convertir teorías diferentes en la misma teoría. Esas unidades que generan no son teorías individuales sino grupos de teorías interconectadas, o lo que el estructuralismo denomina *holones ("totalidades ") teóricos.* Estas macro-unidades científicas pueden englobar partes de una disciplina, o incluso de disciplinas diferentes, y son fundamentales para elucidar algunas cuestiones relativas a la estructura global de la ciencia. El examen de estas cuestiones, sin embargo, sobrepasa los límites de este libro (para un estudio detallado, cf. Balzer, Moulines y Sneed, 1987, cap. VIII).

6. Consideraciones finales

Con este capítulo concluimos el análisis de la estructura sincrónica de las teorías. La reconstrucción o análisis de una teoría debe poner de manifiesto todos los aspectos que sean relevantes para elucidar su naturaleza. Independientemente del formalismo que se prefiera usar para ello, la revisión que hemos hecho permite establecer *al menos los siguientes* elementos relevantes para la dimensión sincrónica de las teorías.

1. Las teorías tienen una parte formal, las leyes, y otra aplicativa, los sistemas físicos concretos a los que se pretende aplicar las leyes. Tal pretensión es expresada por la aserción empírica de la teoría.
2. Es más adecuado identificar las teorías a través de sus modelos que a través de sus enunciados. Para dar cuenta de algunas intuiciones hemos de referirnos siquiera

implícitamente a los modelos, y lo preferible es presentar el análisis metateórico, así como las cuestiones vinculadas al mismo, directamente en términos de modelos.

3. El aparato conceptual con el que se describen y determinan los modelos de datos es sólo parte del usado por la teoría. La determinación de los modelos de datos no puede depender de conceptos cuya aplicación presuponga la validez de la teoría. Los conceptos mediante los que se determinan los datos son pues previos, anteriores o no-teóricos en relación a la teoría para la que son datos. Los conceptos mediante los que la teoría explica o subsume esos datos son los conceptos propios o teóricos en relación a la teoría. La distinción "teórico/no-teórico" es relativa a cada teoría.

4. La caracterización del componente formal debe hacer manifiesta la diferencia entre aparato meramente conceptualizador y aparato propiamente constrictivo.

5. En cuanto al aparato conceptualizador, se debe hacer manifiesta la diferencia entre los conceptos previos, T-no teóricos, y los conceptos propios, T-teóricos.

6. En cuanto al aparato propiamente constrictivo, la reconstrucción debe hacer manifiesta la diferencia entre: a) constricciones que se imponen a sistemas aislados e involucran conceptos exclusivos de la teoría en cuestión (leyes propias); b) constricciones que se imponen a sistemas aislados e involucran conceptos de diferentes teorías (leyes puente); c) constricciones que se imponen a grupos de sistemas (condiciones de coherencia o ligaduras).

7. La parte aplicativa, los sistemas de datos, seleccionada intencional y paradigmáticamente y determinada T-no teóricamente, contribuye esencialmente a la determinación del significado empírico de los términos teóricos.

8. Todo lo anterior se debe considerar conformando una estructura dúctil, con unas partes más genéricas y esenciales que constituyen el núcleo firme de la teoría, y otras partes más específicas y accidentales que pueden ir modificándose como resultado de la contrastación de la aserción empírica. En qué sentido se pueden producir estas modificaciones lo examinaremos en detalle en el capítulo 13.

CAPÍTULO 11

RELACIONES INTERTEÓRICAS

Las teorías de las ciencias empíricas en general (a diferencia, quizá, de algunas teorías de la matemática pura y de las teorías metafísicas) no son "mónadas" conceptuales y metodológicas; es decir, ni desde el punto de vista de su armazón conceptual, ni tomando en cuenta el modo como funcionan, como se aplican y ponen a prueba, pueden ellas existir de manera completamente aislada unas de otras. En el capítulo anterior hemos visto ya un modo en que las teorías empíricas están conectadas unas con otras, a través de los vínculos interteóricos *o leyes puente.* En este capítulo examinaremos otros tipos de relaciones interteóricas de naturaleza más global, en especial la *teorización,* la *reducción y* la *equivalencia.* Después de una introducción a la noción general de relación interteórica, examinaremos cada una de estas relaciones y concluiremos con un apéndice dedicado al reduccionismo entre ciencias especiales y ciencia básica.

1. Concepto general de relación interteórica

Cada teoría de las diversas disciplinas científicas se halla en relaciones más o menos estrechas y de diversa índole con otras teorías, con frecuencia de la misma disciplina, pero a veces también de disciplinas bastante distintas. No se puede entender y aplicar una teoría mecánica, pongamos por caso, sin tomar en consideración su relación con la geometría física; las relaciones de la termodinámica con la química son esenciales a ambas disciplinas; no sabremos realmente qué dice la genética sobre los seres vivos si no tomamos en cuenta conceptos esenciales de la taxonomía, etc. Es muy dudoso que, en el estado actual de la ciencia empírica, exista una sola teoría, por elemental que sea, que no conlleve relaciones significativas empírica y conceptualmente con otras varias teorías. En muchos casos, estas relaciones son incluso absolutamente esenciales a la teoría en cuestión en el sentido de que no podemos identificar esa teoría o determinar plenamente de qué trata si desconocemos algunas de sus relaciones con otras teorías. Por ejemplo, la relación de la mecánica con la geometría física es esencial para la primera (aunque no para la segunda): no comprenderemos lo esencial de una teoría mecánica si no aprehendemos la vinculación de algunos de sus conceptos básicos con conceptos provenientes de la geometría.

En otros casos, aunque sería quizá exagerado afirmar que la identificación de una teoría dada presupone su relación con otras teorías, sin embargo, las relaciones interteóricas resultan esenciales a la hora de someter a prueba empírica la teoría en cuestión. Probablemente no haya una sola teoría empírica cuya contrastación con la experiencia no requiera del concurso de otras teorías, aunque sólo sea por el hecho de que los instrumentos utilizados para poner a prueba esa teoría vienen controlados por las leyes de otras teorías. Así, por ejemplo, cuando ponemos a prueba las predicciones experimentales de la termodinámica mediante un termómetro, presuponemos implícitamente que éste funciona correctamente, y ello quiere decir que funciona de acuerdo a leyes mecánicas, hidrodinámicas, electrostáticas, etc.

La constatación de que nunca podemos poner a prueba una teoría empírica aisladamente, sin tomar en cuenta que forma parte de toda una familia de teorías coadyuvantes, la hizo ya Pierre Duhem a principios del siglo xx. Este autor formuló esta tesis sólo para las teorías de la física y dudaba de que fuera aplicable a otras disciplinas (a la fisiología, por ejemplo). Sin embargo, hoy día sabemos ya lo bastante acerca de la estructura de otras disciplinas, además, de la física como para que nos atrevamos a suponer el mismo efecto en todas las ciencias empíricas: ninguna teoría empírica puede ser contrastada sin tomar en consideración sus relaciones interteóricas. Esta visión de la problemática de la contrastación de teorías fue radicalizada posteriormente por W. V. Quine, quien postuló que, en la contrastación de cada teoría particular, interviene una madeja inextricable y prácticamente inabarcable de relaciones de esa teoría con la totalidad de la ciencia (incluso las ciencias formales). A tal tesis se la suele caracterizar como *holismo (metodológico)* (de la palabra griega *holos,* que significa "totalidad"); también se la suele llamar "tesis Duhem-Quine", dando a entender que ambos autores, Duhem y Quine, defendieron prácticamente el mismo punto de vista. Sin embargo, como acabamos de indicar, el "holismo" de Duhem es mucho más moderado (y verosímil) que el holismo extremo de Quine. A efectos de la discusión presente nos basta con dar por bien establecida la versión duhemiana del holismo: al contrastar una teoría con la experiencia siempre hay que tener en cuenta al menos *algunas* de sus relaciones con *algunas* otras teorías.

Así, pues, tanto respecto a la cuestión de la identidad de teorías empíricas como respecto a su contrastación, sus relaciones mutuas juegan un papel de primer orden. Por ello es que el estudio de las relaciones interteóricas representa un capítulo muy importante de la filosofía de la ciencia, un capítulo largo tiempo negligido, pero que en las últimas décadas ha pasado cada vez más al primer plano de la discusión. El estudio de las relaciones interteóricas resulta imprescindible para comprender los aspectos más globales de la ciencia, tanto en una perspectiva sincrónica como en una diacrónica. Aquí podemos tratar sólo de los tipos más importantes y discutidos de relaciones interteóricas, y lo haremos sólo desde un punto de vista sincrónico; algunos aspectos de relevancia diacrónica de las relaciones interteóricas entrarán en juego en el último capítulo.

Otra restricción en el examen que asumiremos es la siguiente. Si consideramos un grupo de *n* teorías, *T,, ...,* *T„* (con *n* > 2), que constatamos relacionadas entre sí, podría ocurrir que hubiera una relación n-ádica R(T,, ..., T,,) que no se pudiera descomponer en relaciones parciales entre pares de teorías del grupo. Sin embargo, numerosos análisis de ejemplos reales de relaciones interteóricas parecen indicar que la eventualidad mencionada es meramente una posibilidad lógica en la inmensa mayoría de casos, y que los tipos realmente relevantes de relaciones interteóricas son (casi) siempre relaciones establecidas sobre un par de teorías, es decir, relaciones diádicas. En cualquier caso, aquí restringiremos nuestra atención a las relaciones interteóricas diádicas. De éstas, a su vez, hay de tipos diversos, según su forma lógica y su función metodológica. Muchos de esos tipos ni siquiera han recibido una denominación especial en la literatura, y los dejaremos de lado. Aquí nos limitaremos a examinar tres grandes tipos, que han sido objeto de amplias investigaciones, y que tienen también especial relevancia epistemológica: la teorización, la reducción y la equivalencia.

La reconstrucción formal de los diversos tipos de relaciones interteóricas dependerá naturalmente, en parte, de la noción formal de teoría que se presuponga. Si se adopta una concepción axiomática o enunciativa de las teorías como cálculos interpretados (cf. capítulo 8), entonces está claro que los diversos tipos de relaciones interteóricas aparecerán como relaciones entre (sistemas de) enunciados o axiomas; en cambio, si adoptamos una concepción semántica de las teorías (cf. capítulo 10), y en especial si las definimos como estructuras modeloteóricas (que es el punto de vista favorecido en este libro), entonces las relaciones interteóricas también se verán como relaciones entre modelos o conjuntos de modelos. En lo que sigue, y para el examen de cada uno de los tipos considerados, primero adoptaremos la idea más clásica de las teorías como sistemas de enunciados para pasar luego a la versión modeloteórica. En realidad, las dos formas de reconstrucción no son incompatibles entre sí, sino que la primera puede servir a modo de sugerencia "elemental" para la segunda que, como veremos, permite un análisis más diferenciado y complejo de las relaciones interteóricas.

2. Teorización

La teorización, vista como relación interteórica, se da entre dos teorías T, y T_0 cuando algunos de los conceptos que aparecen en las leyes de T, vienen determinados en la teoría T_0, o sea, le son "provistos" a T, por T_0; a tales conceptos podemos llamarlos "conceptos T,-no-teóricos", mientras que a los demás conceptos de T, que no vienen determinados por ninguna teoría independiente de T,, los llamamos "T,-teóricos".[1] En tal caso, cuando algunos de los conceptos de T, vienen determinados por una teoría T_0 independiente de T, y otros, en cambio, no vienen determinados por ninguna teoría independiente de *T,,* decimos que *T,* es una *teorización* de T_0 o que T_0 es una teoría *subyacente a*

1. La distinción entre conceptos T-teóricos y T-no-teóricos que establecemos aquí está inspirada en las ideas básicas de la concepción estructural, expuestas en el capítulo 10 (§5). Sin embargo, en la forma en que aquí la discutimos es independiente de dicha concepción.

T,. También podemos decir que *T*„ es una teoría *metodológicamente previa* a T,, pues sin ella algunos de los conceptos de T, no quedarían determinados y por tanto no sabríamos cómo aplicar T, ni, en definitiva, de qué trata dicha teoría.

Así, por ejemplo, si no dispusiéramos de los conceptos cinemáticos de distancia, tiempo, velocidad y aceleración (y de maneras de determinarlos de acuerdo a ciertos principios cinemáticos y geométricos), no tendría sentido tratar de utilizar, aplicar o poner a prueba una teoría mecánica. Por ello podemos decir que la mecánica es una teorización de la cinemática. O bien, si no dispusiéramos del concepto de volumen, no podríamos ni siquiera entender de qué trata la termodinámica, por lo que hay que considerar esta última como una teorización de la geometría física. Finalmente, está claro que la distinción entre fenotipo y genotipo es esencial para cualquier teoría genética; pero la noción de fenotipo viene determinada por los rasgos anatómicos y fisiológicos de los seres vivos, por lo que la genética será una teorización de la anatomía y la fisiología.

En general, se suele suponer que, si T, es una teorización de T_0, es porque T_0 está más próxima a la experiencia inmediata del sujeto epistémico, puede servir como "base empírica" para poner a prueba T,, la cual por lo general se considerará más "abstracta", más alejada de la experiencia. Algunos autores también contraponen el lenguaje en que está formulada T_0, considerado como "lenguaje observacional", al lenguaje propio de *T*, considerado como "lenguaje teórico" (cf. cap. 8). Podemos aceptar este modo de hablar siempre y cuando tengamos presente que se trata de una distinción relativa al par <T,, T_0> : T_0 es "observacional" con respecto a T,, pero no tiene por qué serlo en un sentido absoluto; es decir, Ti, no tiene por qué considerarse una teoría basada únicamente en "observaciones puras", suponiendo que haya tal cosa. Basta simplemente que las determinaciones de los conceptos en T_0 hayan de presuponerse antes de pasar a utilizar T,. Pero por supuesto que T_0 puede ser, a su vez, teorización de otra teoría aún más "elemental" T_2, y por otro lado T, puede servir de "base empírica" a otra teoría aún más "abstracta" T_3, etc.

La teorización puede ser *total o parcial.* Diremos que *T*, es una "teorización total" de *T*, cuando T, es la *única* teoría de la cual *T*, es teorización, o sea, T_0 es la única teoría que subyace a T,. Es plausible suponer que un ejemplo de teorización total lo constituye la relación entre la mecánica y la cinemática, pues todos los conceptos no propios de la mecánica que hay que presuponer para aplicar la mecánica provienen de la cinemática. Sin embargo, la teorización total es más bien la excepción y no la regla. Por lo general, a una misma teoría subyacen varias teorías distintas, o sea, T es teorización de T_0, *T*(,', *Tó'*, ... Así, por ejemplo, la termodinámica es teorización de por lo menos tres teorías: la geometría física (por el volumen), la hidrodinámica (por la presión) y la estequiometría (por el concepto de mol).

Parece muy plausible suponer que la teorización es una relación *asimétrica;* o sea, que si *T*, es teorización de T_0, entonces no podrá ser T_0 también teorización de *T*,. Sin embargo, es importante notar que no hay ninguna razón *a priori* o conceptual para que ello sea así: en principio, podría ocurrir en algún caso que algunos conceptos de T, presupusieran T_0, pero que ciertos conceptos de T_0 presupusieran a su vez la determinación de otros conceptos de T,. En tal caso no tendríamos un círculo lógico vicioso, pero sí lo que podríamos denominar un "círculo metodológico vicioso". Está claro que la praxis científica está constituida de tal modo que, *en principio,* tratará de evitarse una situación así. No obstante, que realmente consiga evitarse siempre, es otra cuestión. Puede ocurrir que, en la práctica del uso de teorías, se introduzcan inadvertidamente tales círculos. Ello puede ocurrir especialmente cuando las "cadenas de teorizaciones" son relativamente largas. En efecto, supongamos que tuviéramos una serie de teorías T,,, *T,, ..., T,, , T,,,* tal que T_n sea teorización de T,,, ..., *T,* teorización de T_0 y finalmente que T_0 sea teorización de T,,; admitamos además que la relación de teorización es transitiva, o sea que, si T_3 es teorización de T_2 y T_2 es teorización de *T,,* entonces también habrá que considerar T_3 como teorización de T, (lo cual es un supuesto muy plausible); entonces tendríamos en el caso de esa "cadena" de teorías que *T,,* es teorización de *T,, y* T_0 es teorización de T,,, precisamente el círculo que tratábamos de evitar.

Es una cuestión todavía abierta la de si una situación como la descrita puede realmente darse en las ciencias empíricas, y qué consecuencias epistemológicas y metodológicas tendría ella; esta cuestión, como el lector habrá adivinado, está emparentada con las tesis del holismo señaladas al principio, en particular en su forma extrema debida a Quine. Aquí no podemos detenernos a fondo en este problema y nos limitamos a apuntarlo tan sólo. En general, supondremos que tales círculos no se dan, y que la constitución de la mayoría de disciplinas (al menos desde el punto de vista *sincrónico)* es tal que la teorización es realmente una relación transitiva y asimétrica. Ello implica, a su vez, la existencia de un orden jerárquico entre las teorías, desde las más "básicas", que no son teorizaciones de otras teorías, hasta las más "teóricas", que revelan tener tras de sí largas cadenas de teorizaciones. Ésta es la alternativa *fundacionista.* Según la alternativa opuesta, *coherentista,* no habría teorías básicas y globalmente considerado "todo estaría presupuesto en todo". Caben alternativas intermedias, con la presencia tanto de algunas teorías básicas como de algunos círculos metodológicos. Aunque hemos supuesto que en general tales círculos no se dan (fundacionismo), debe quedar claro que ello no es algo que se pueda establecer *a priori,* sino que se debe resolver (meta)empíricamente mediante un detallado y exhaustivo trabajo de análisis y reconstrucción de conjuntos de teorías.

Hemos iniciado la discusión de la relación de teorización caracterizándola como la relación que existe entre dos teorías *T,* y T, cuando algunos de los conceptos de *T,* vienen determinados por T_0, mientras que otros conceptos de T, no vienen determinados por ninguna teoría independiente de T, y son por tanto "T-teóricos". Esta caracterización es más o menos intuitiva pero por ello mismo también más o menos vaga. Conviene que nuestra caracterización sea más precisa.

La noción clave aquí, que aún no hemos dilucidado formalmente, es la de *determinación.* Hemos dicho que, cuando *T,* es una teorización de T_0, algunos conceptos de *T,* vienen determinados en T,, y otros no. Pero ¿qué quiere decir exactamente que los conceptos de una teoría son "determinados" en otra? Para elucidar esta cuestión haremos uso de la concepción modeloteórica de las teorías tal como la hemos expuesto en el capítulo anterior, especialmente en su versión estructural. Antes, sin embargo, conviene introducir la noción general de *subestructura.*

Definición 11.1:

> Sean dos estructuras $x = \langle D_1, ..., D_m, R_1, ..., R_n \rangle$ y $y = \langle D'_1, ..., D'_p, R'_1, ..., R'_q \rangle$.
> Diremos que y es una *subestructura* de x, $y\mathbf{S}x$, si y sólo si:
> (1) $p \leq m \wedge q \leq n$
> (2) $\forall i \, \exists j \, D'_i \subseteq D_j \; (1 \leq i \leq p, 1 \leq j \leq m)$
> (3) $\forall i \, \exists j \, R'_i \subseteq R_j \; (1 \leq i \leq q, 1 \leq j \leq n)$

Intuitivamente: la estructura y es subestructura de x cuando todos los dominios de *y* son subconjuntos (propios o impropios) de algunos dominios de x y *mutatis mutandis* para las relaciones (y funciones) respectivas. La noción de subestructura es pues simplemente una generalización de la noción elemental de subconjunto. Un caso extremo de subestructura es naturalmente la identidad de dos estructuras; en el otro extremo tenemos que el conjunto vacío es subestructura de cualquier estructura; un caso intermedio de subestructura es lo que en el capítulo anterior llamamos submodelo o "recorte" de un sistema, esto es, el resultado de suprimir algunas de las relaciones del sistema original. Esta noción de *subestructura* es pues extremadamente general (a veces se usa el mismo término para otra noción más estrecha, a saber, como Def. 11.1 pero exigiendo en *(1)p=myq=n).*

Supongamos ahora que los modelos (potenciales) de la teoría T, tienen la forma *x = <D,, ..., D,,,, R,, ..., R,,> (D;* son los dominios básicos de *T,* y *R;* las relaciones construidas sobre ellos), análogamente supongamos que los modelos (potenciales) de la teoría T_0 tienen la forma y = *<D',, ..., D',,, R',, ..., R'$_y$>,* con *p <_ m y q <_ n.* Podemos definir ahora exactamente qué significa que T, sea teorización de T_0. La idea básica es la siguiente: cuando *T,* se considera teorización de *T,,* es porque toda aplicación intencional *x* de *T,* (es decir, toda estructura que representa un "pedazo de realidad" al que se pretende aplicar *T,,* cf. cap. 10, §5) tiene una subestructura *y* "determinada por T_0" en el sentido de que cumple sus leyes, esto es, y es *un modelo actual* de T_0 (o parte de un modelo actual de T_0). Por otro lado, para que T, sea una teorización genuina deberá haber un "excedente" de conceptos no provistos por Ti,, es decir, todos los modelos (potenciales) de T, contendrán una subestructura "ajena" a los modelos de T_0.

Definición 11.2:

> T_1 es una *teorización* de T_0 si y sólo si:
> (1) $\forall x \in I(T_1) \, \exists y,z \, (y\mathbf{S}x \wedge y\mathbf{S}z \wedge z \in M(T_0))$
> (2) $\forall x \in Mp(T_1) \, \exists y \, (y\mathbf{S}x \wedge \forall z(z \in Mp(T_0) \rightarrow \neg \, y\mathbf{S}z))$.

En el caso en que en la condición (1) ocurra y = x, tendremos que cada aplicación intencional "completa" de T, se concibe como un modelo o parte de un modelo de una determinada teoría subyacente, en cuyo caso sería superfluo buscar otras teorías subyacentes para *T,,* situación que se corresponde a lo que hemos descrito antes como *teorización total.* Pero, por lo general, las aplicaciones intencionales de una teoría *T,* estarán compuestas de diversas subestructuras y, y', ... determinadas como modelos de diversas teorías subyacentes T_0, T'_0, ...

3. **Reducción**

La reducción de una teoría a otra es probablemente el tipo de relación interteórica que más se ha discutido en la filosofía de la ciencia. Ello se debe a que la relación de reducción se ha conectado con cuestiones epistemológicas y metodológicas de largo alcance, como son las del realismo (epistemológico), la unidad de la ciencia, el progreso científico, etc. En efecto, si todas las disciplinas científicas existentes pudieran reducirse a una sola (por ejemplo, todas las ciencias sociales a la biología, la biología a la química, la química a la física), y dentro de esa disciplina hubiera una sola teoría que redujera a todas las demás (por ejemplo, la "gran teoría unificada" que persiguen los físicos de partículas), entonces podríamos considerar el desarrollo científico como un "progreso" hacia una "unidad" cada vez mayor, en la que todas las teorías quedarían al fin reducidas a una sola que explicaría todos los fenómenos del universo y que se podría considerar "la verdadera representación" de "la realidad" tal cual es; tal situación parecería una garantía de conocimiento definitivo (cf. más adelante la última sección).

Frente a este programa reduccionista se han planteando objeciones de diversa índole. Entre ellas, quizá las más frecuentes dentro de la filosofía de la ciencia provienen de una perspectiva diacrónica: se señala que la repetida manifestación de *revoluciones científicas,* en tanto que rupturas dramáticas en el aparato conceptual y metodológico de una disciplina, con la concomitante *inconmensurabilidad* de las teorías involucradas (cf. cap. 13), dan al traste con la idea de reducir las teorías anteriores a las posteriores en una revolución; al menos históricamente, según estos críticos, no resulta verosímil el programa reduccionista para teorías diferentes (y aún menos, si cabe, para las diversas disciplinas).

Aquí no podemos entrar a fondo en esta discusión. Baste hacer notar, no obstante, que tanto las tesis reduccionistas como las antirreduccionistas han adolecido a menudo de cierta falta de rigor conceptual, y que en realidad se puede objetar al reduccionismo radical sin necesidad de apelar a "revoluciones" e "inconmensurabilidades". Tan pronto como se ofrece un concepto exacto y verosímil de reducción se comprueban dos cosas: *a)* que las consecuencias epistemológicas y ontológicas de las reducciones, caso de existir, son mucho menos importantes de lo que la discusión ha sugerido; *y b)* que hay muchos menos casos *genuinos* de reducción de lo que parece y de lo que en obras de divulgación científica suele sugerirse. Y para darse cuenta de ello no es necesario constatar ninguna "inconmensurabilidad", sino que basta con percatarse de que, incluso en el caso de teorías que pertenecen a una misma "familia" y que están vinculadas conceptualmente, reducir una teoría a otra es mucho más arduo de lo que puede esperarse, es una empresa que pocas veces ha culminado en un éxito total. Con otras palabras, incluso prescindiendo de la problemática de las revoluciones científicas y de la inconmensurabilidad, lo cierto es que se han sobrevalorado las posibilidades de reducir unas teorías a otras.

A esta dificultad se añade el hecho (debido precisamente a la falta de rigor en el tratamiento del problema) de que muchos supuestos ejemplos de reducciones no corresponden en realidad al concepto de *reducción exacta,* que es la reducción propiamente dicha, sino a lo sumo a lo que podemos llamar una *reducción aproximativa.* De hecho, la relación de aproximación como relación interteórica, ya sea de carácter reductivo o no, es mucho más importante y frecuente que la reducción exacta, y aunque en algunos casos la aproximación revela ciertas semejanzas estructurales con la reducción, sería erróneo equiparar y aún más identificar ambos conceptos. Muchos ejemplos que se han dado en la literatura científica o filosófica de reducciones revelan ser, ante un examen más cuidadoso, solamente aproximaciones: éste es el caso para la supuesta reducción de la teoría planetaria de Kepler a la teoría de la gravitación de Newton, de la termodinámica a la mecánica estadística, de la mecánica clásica a la relativista, de la genética mendeliana a la genética de poblaciones, etc. La relación interteórica de aproximación es, sin embargo, de naturaleza esencialmente más complicada que otras relaciones interteóricas, en especial la reducción, y su tratamiento requeriría de cierto nivel de tecnicismos que no podemos desarrollar en este libro.

No obstante las prevenciones que hemos formulado sobre la tendencia a sobrevalorar el tema de la reducción en la ciencia, no cabe duda de que se trata de un tipo importante de relación interteórica, que conviene precisar y para el cual hay ejemplos concretos e interesantes. Casos claros de reducción (exacta) de teorías son: la reducción de la mecánica (cartesiana) del choque a la mecánica (newtoniana) de partículas, de la mecánica del sólido rígido a la mecánica de partículas, de la teoría de los gases ideales a la teoría cinética, de la electrostática a la electrodinámica y de la genética mendeliana a (cierta versión de) la biología molecular; probablemente haya otros varios casos que aún no han sido reconstruidos con detalle. Estos casos paradigmáticos de reducciones y las intuiciones asociadas a ellos pueden guiamos a la hora de formular un concepto viable y bien fundado de reducción, que además nos pudiera servir más adelante como base para tratar adecuadamente su "pariente próximo", la aproximación reductiva, la cual sin duda reviste cierta analogía con la reducción exacta.

La intuición básica de la reducción puede ser interpretada tanto en una perspectiva diacrónica como en una sincrónica. Diacrónicamente, la teoría reducida *T precede* a la teoría reductora T* en el sentido de que representa un estadio más "elemental", más "simple", de nuestro conocimiento de determinada parcela de la realidad. En cierto modo, T ha de quedar "cubierta" por T* en el sentido de que los logros positivos de Testarán contenidos también en los logros positivos de T*, aunque probablemente no a la inversa. Podemos decir que, sobre el mismo dominio empírico, *T** dice lo mismo que ya decía *T,* pero lo dice mejor, y además dice otras cosas que nunca dijo T. Desde el punto de vista sincrónico, la teoría reducida T con frecuencia representa un modo más rápido y expedito, pero también "más grosero" de resolver los mismos problemas que se plantean en la teoría reductora T*. Es decir, la teoría reducida simplifica la formulación de los problemas y las aplicaciones propuestas, haciéndolas más asequibles que su teoría reductora, aunque al precio de negligir ciertas informaciones relevantes. Así, por ejemplo, podemos tratar del choque de dos esferas macizas olvidándonos de cómo esas esferas están compuestas de partículas unidas entre sí por ciertas fuerzas de cohesión; o podemos predecir el cambio de volumen que sufrirá un gas al ser sometido a cierta presión sin preocuparnos del movimiento de las moléculas en el interior del gas. La cuestión que nos planteamos ahora es la de cómo desarrollar un concepto general de reducción que responda a estos ejemplos y a la idea intuitiva que ellos sugieren.

Hemos dicho que la teoría reductora se refiere en lo esencial al mismo campo de la experiencia y que contiene la misma información, y más, que la que provee la teoría reducida. Ello sugiere dos cosas. Por un lado, que ambas teorías estarán vinculadas semánticamente, y por tanto que habrá una conexión entre los conceptos de ambas. Y por otro, que las aseveraciones sobre el mundo que hace la teoría reductora son "más fuertes" que las que hace la reducida, pero no incompatibles con ellas. Estos dos requisitos intuitivos de la reducción han sido explicitados en la concepción axiomática de las teorías como las dos condiciones fundamentales de toda reducción: la condición de *conectabilidad y* la de *derivabilidad.* Cuando en el capítulo 8 presentamos la noción de teoría axiomática ya dimos una primera idea de esta noción de reducción (sin tener entonces en cuenta los aspectos empíricos). Recuérdese (cap. 8, §1) que lo esencial consistía entonces en que una teoría reduce a otra si se pueden definir los términos primitivos de la segunda mediante términos primitivos de la primera de modo que los axiomas de la segunda se deriven de los axiomas de la primera más estas definiciones. Éste es el núcleo de la idea clásica de reducción (dos referencias básicas para la misma son Kemeny y Oppenheim, 1956, y Nagel, 1961, cap. 11).

El requisito de conectabilidad exige que, para disponer de una formulación explícita de la reducción de T a T*, se establezcan ciertas "definiciones coordinadoras" entre *todos* los conceptos básicos de *T y* al menos algunos conceptos básicos de *T*.* Estas definiciones tendrán en general la forma de condicionales que afirman que, si cierto concepto *C de T* se aplica a cierto dominio de objetos *D,* entonces necesariamente a este dominio *D* se aplicará(n) también cierto o ciertos conceptos C*, ..., C* de T* "coordinados" con C. El segundo requisito, el de derivabilidad, exige que las leyes de *T* sean *todas* deducibles de las leyes básicas de T* junto con las definiciones coordinadoras (y eventualmente algunos enunciados más particulares sobre condiciones iniciales). Tomemos el ejemplo de la reducción de la mecánica del sólido rígido a la mecánica newtoniana de partículas. En la primera, un concepto básico es el de sólido rígido y una ley básica es la de conservación del momento angular. En la segunda, tenemos como concepto básico el de partícula y las leyes básicas son el Segundo Principio de Newton y la ley de acción y reacción. Pues bien, para reducir la primera teoría a la segunda hay que establecer primero una definición coordinadora del concepto de sólido rígido en términos del concepto de partícula, por la cual se define un sólido rígido como un conjunto de partículas que mantienen distancias constantes entre sí (y análogamente con las restantes nociones propias de la teoría reducida); y luego hay que demostrar que, de las leyes de Newton, más la mencionada definición coordinadora, se deduce la ley de la conservación del momento angular. Debe notarse que aunque las definiciones coordinadoras son afirmaciones generales cargadas (si la reducción es viable) de cierta *nomicidad* ("necesidad" en virtud de la naturaleza), no se trata de leyes usuales; se trata más bien de relaciones de *constitución* (sobre esto, cf. más adelante la última sección).

Este análisis de la noción de reducción apunta, en lo esencial, en la dirección correcta; sin embargo, cuando la queremos aplicar a casos concretos, nos percatamos de que adolece aún de deficiencias, de que es demasiado simplista o idealizada. Ella enfrenta sobre todo dos problemas: (i) muchas veces es difícil o inverosímil establecer *para cada uno* de los conceptos básicos de *T* una definición coordinadora con conceptos de *T**; (ii) la deducción de las leyes de *T* a partir de las de *T** muchas veces no puede llevarse a cabo formalmente, ya sea porque nos faltan precisamente las definiciones coordinadoras (o las que se han propuesto son intuitivamente inaceptables), o bien porque la derivación requiere, además, de ciertos postulados o supuestos adicionales difíciles de formular o variables según el tipo de aplicación.

Por ello, aun cuando podemos conservar la noción general de reducción estipulada antes, es conveniente tomar un enfoque "más global", que no adolezca de las dificultades señaladas. De nuevo nos ayudará aquí la versión modeloteórica. Los requisitos fundamentales serán ahora, dicho de manera intuitiva, los siguientes. Primero, en vez de estipular una coordinación para cada uno de los conceptos de T tomado singularmente, requeriremos simplemente una "correspondencia global" entre el marco conceptual de *T* y el de *T**; ella será formalmente una relación entre *Mp(T)* y *Mp(T*)*. Ahora bien, tal correlación no sólo deberá existir a nivel de los modelos potenciales respectivos, sino también a nivel de las aplicaciones *I(T)* e *I(T*)*, o sea, de las porciones del mundo empírico a las que pretenden aplicarse ambas teorías; toda aplicación intencional de *T* deberá tener su correlato en *T**, pero no necesarimente a la inversa (en general, T* tendrá un mayor campo de aplicación que *T*). La correlación entre *I(T)* e *I(T*)*, formalmente hablando, no será exactamente la misma relación que la que se da entre *Mp(T)* y *Mp(T*)*, pues recuérdese que las aplicaciones intencionales son modelos parciales, esto es, subestructuras resultantes de "recortar" de los modelos potenciales sus constituyentes T-teóricos; sin embargo, es una relación "derivada" de la primera, en el sentido de que es esta misma restringida a las subestructuras en cuestión. Finalmente, el requisito de derivabilidad de las leyes adoptará en esta interpretación modeloteórica la siguiente forma. Aunque no podamos decir, en sentido estricto, que las leyes de *T* se deducen de las de *T**, no obstante podremos postular una condición intuitivamente análoga: siempre que una aplicación cumpla las leyes de T*, es decir, sea extensible a un modelo actual de *T*, y además* cumpla ciertas condiciones específicas, es decir, sea extensible a un modelo actual *de una especialización* de *T**, llamémosla r, entonces en *T* el correlato de esa aplicación cumplirá las leyes de la teoría reducida *T*, o sea, será extensible a un modelo actual de T. Podemos ahora sintetizar estos requisitos en la siguiente definición; en ella, denotamos añadiendo el subíndice '*e*' a la relación que cualquier relación entre modelos potenciales genera a nivel empírico (T-noteórico): si p es una relación entre modelos potenciales, p_e es el resultado de recortar los constituyentes T-teóricos de los modelos potenciales de los pares de p, esto es, p_P = r[p]. La idea que hay detrás es que en la reducción no se usa "toda" la teoría reductora sino sólo parte de ella, determinada especialización (en la definición que sigue, y para simplificar la notación, no usaremos la noción de red teórica sino la de elemento teórico y consideraremos que la teoría reductora es un elemento teórico que tiene especializaciones; recuérdese que la relación σ es la relación de especialización entre elementos teóricos, cf. cap. 10, §5).

Definición 11.3:

Sean $Mp(T)$, $M(T)$, $I(T)$, respectivamente, los conjuntos de modelos potenciales, modelos actuales y aplicaciones intencionales de T, y análogamente $Mp(T^*)$, $M(T^*)$, $I(T^*)$ respecto de T^*.
T es *reducible* a T^* si y sólo si existe una relación ρ tal que:
(1) $\rho \subseteq Mp(T) \times Mp(T^*)$.
(2) $I(T) \subseteq \text{Dom}\ \rho_e$ y $\rho_e[I(T)] \subseteq I(T^*)$.
(3) $\forall y, y^* (<y, y^*> \in \rho_e \wedge y^* \in I(T^*) \rightarrow (\exists T^{*\prime}(T^{*\prime}\ \sigma\ T^* \wedge y^* \in \mathbf{r}[M(T^{*\prime})]) \rightarrow y \in \mathbf{r}[M(T)]\ ))$.

La primera condición establece simplemente que ambas teorías están "globalmente correlacionadas" a nivel de sus marcos conceptuales. La segunda condición establece que la relación ρ_e generada por p a nivel no-teórico (empírico) conecta también globalmente las aplicaciones, con la especificación adicional de que *toda* aplicación intencional de *T deberá* tener un correlato en T* (aunque no necesariamente a la inversa). La tercera condición dice, de cada par de aplicaciones correlacionadas, que si la "aplicación reductora" cumple ciertas leyes especiales de la teoría reductora (más, por supuesto, las leyes fundamentales de la misma), entonces la "aplicación reducida" cumplirá necesariamente las leyes fundamentales de la teoría reducida. En este sentido, dichas leyes se "derivan" de las primeras: que cierta aplicación es subsumible bajo la teoría reductora implica que su correlato es subsumible bajo la teoría reducida; esto es, que la reducida se aplique con éxito "se deriva" de que la reductora se aplica con éxito. Por otro lado, debe notarse que la condición *(3)* no exige que la especialización *T** de *T** sea siempre la misma para cada par de aplicaciones correlacionadas; en algunos casos puede que sea así, pero en otros la especialización escogida puede que varíe según ciertos tipos de aplicaciones intencionales consideradas en una y otra teoría.

4. **Equivalencia**

La relación de equivalencia entre teorías también ha jugado un papel considerable en discusiones epistemológicas generales, aunque quizá no de manera tan controvertida como en el caso de la reducción. La significación de la equivalencia en términos generales estriba en que, cuando ella se da, dos teorías que a primera vista parecen muy distintas por sus conceptos y leyes, resulta, no obstante, que "hablan de lo mismo" o que aportan la misma información sobre la misma porción de realidad. De ahí puede inferirse fácilmente la conclusión epistemológica general de que no tiene por qué haber univocidad en el tratamiento teórico adecuado de la misma parcela de nuestra experiencia. Diversas teorías pueden ser igualmente aptas para explicar el mundo que nos rodea, ninguna de ellas es *la verdadera* en un sentido absoluto. Así, por ejemplo, podemos desarrollar una teoría de las relaciones espaciales en la que partimos del concepto básico de "punto geométrico" y definimos las líneas como sucesiones infinitas de puntos; o bien, alternativamente, podemos partir del concepto de "línea recta" como concepto básico y definir los "puntos" como las intersecciones de líneas rectas. Si escogemos bien los axiomas de una y otra teoría, la que trata primordialmente de puntos y la que trata primordialmente de líneas, constataremos que, aunque aparentemente las dos teorías hablan de cosas distintas, ambas establecen exactamente las mismas relaciones espaciales entre los objetos que podemos comprobar en nuestra experiencia cotidiana, y en este sentido "hablan de lo mismo". En otro campo, el del movimiento de los cuerpos, constatamos que la teoría mecánica de Newton y la teoría mecánica de Lagrange, aunque construidas sobre conceptos y principios distintos, conducen a los mismos resultados empíricos sobre el movimiento de los cuerpos en general; por ello es frecuente leer en los libros de texto de física que la mecánica newtoniana y la mecánica lagrangiana son dos "formulaciones equivalentes" de la mecánica clásica.

Al tratar el tema de la equivalencia de teorías y sus consecuencias epistemológicas conviene, sin embargo, distinguir dos tipos generales de equivalencia que muchas veces se confunden: la que podemos llamar *equivalencia fuerte,* o equivalencia "en sentido estricto", y la que llamaremos *equivalencia empírica,* que es más débil. En el primer caso, aunque conceptos y leyes de una y otra teoría sean distintos, hay una correspondencia plena y biunívoca entre ambas teorías, de modo que todo lo que puede decirse en la primera teoría puede traducirse sin pérdida de información a la segunda, y viceversa. Es decir, hay una correspondencia exacta entre ambas teorías tanto a nivel conceptual como a nivel del contenido de sus afirmaciones respectivas. El ejemplo de la correlación entre una "geometría de puntos" y una "geometría de líneas" es de esta naturaleza.

En el caso de la equivalencia empírica, más débil, ese paralelismo sólo se da a nivel de los datos empíricos que cubren ambas teorías: todo dato predicho por una teoría es también predicho por la otra, y a la inversa. Y, sin embargo, puede que no haya una correlación plena ni entre los conceptos ni entre las leyes de ambas teorías, de modo que no pueden derivarse las leyes de una teoría a partir de las de la otra, ni a la inversa. En tales casos pueden existir serias divergencias teóricas entre ambas teorías, las cuales, no obstante, no se traducen en divergencias en el campo de lo que podemos experimentar: las teorías dicen "más" de lo que dice la experiencia que ellas cubren. Es en este caso en el que piensa Quine cuando insiste en la Tesis de la Indeterminación de la Teoría por la Experiencia: el mismo dominio de datos experimentales es igualmente compatible con dos o más teorías, las cuales, sin embargo, son incompatibles entre sí a nivel teórico (de ello nos ocuparemos por extenso cuando estudiemos en el próximo capítulo el problema de la inducción; recuérdese también el argumento de van Fraassen que examinamos en el capítulo 10).

Si bien la equivalencia *fuerte o estricta* aparece con bastante frecuencia no sólo en geometría, sino en la mayoría de las ramas de la matemática pura, es dudoso que ella juegue un gran papel en las ciencias empíricas propiamente dichas (excepto en casos triviales como el de dos teorías físicas que se distinguen solamente por un cambio de notación). Se ha solido señalar el ejemplo, ya mencionado, de la relación entre la mecánica de Newton y la de Lagrange como caso de equivalencia fuerte en la física; sin embargo, un análisis formal detenido de este ejemplo, como el que se ha realizado dentro de la concepción estructuralista, muestra que la equivalencia fuerte es válida sólo si se hacen ciertos supuestos (generalmente implícitos) acerca de la estructura global de ambas teorías que están lejos de haber sido confirmados (cf. Balzer, Moulines y Sneed, 1987, cap. VI, §5.1). La cuestión de la equivalencia "Newton-Lagrange" sigue, en realidad, abierta. Con más razón aún puede decirse ello de otros ejemplos que suelen aducirse en la física, como la supuesta equivalencia entre la mecánica de Newton y la de Hamilton, o entre la mecánica ondulatoria y la mecánica de matrices en la física cuántica; lo más probable es que éstos sean sólo casos de equivalencia empírica.

Dada la importancia de la distinción entre equivalencia fuerte y equivalencia empírica, conviene establecerla de la manera más rigurosa posible, pues ello también puede facilitar el examen de ejemplos concretos. Utilicemos de nuevo para ello nuestro aparato modeloteórico habitual.

Hemos dicho, de manera intuitiva, que en el caso de la equivalencia fuerte todo lo que puede decirse en una teoría halla su correlato exacto en la otra, y a la inversa; o sea que hay un paralelismo estricto tanto a nivel del aparato conceptual como de las leyes y sus aplicaciones. En nuestros términos modeloteóricos ello significa una correlación tanto a nivel de los modelos potenciales y aplicaciones intencionales como a nivel de los modelos actuales. Y tomando en cuenta la noción de reducción que hemos explicado más arriba es plausible entonces interpretar la equivalencia entre dos teorías como reducción "de doble vía": una teoría es equivalente a otra cuando la primera es reducible a la segunda y la segunda lo es a la primera. Llegamos así a la siguiente definición.

Definición 11.4:

> T_1 es *equivalente en sentido fuerte* a T_2 si y sólo si T_1 es reducible a T_2 y T_2 es reducible a T_1 (en el sentido de la Def. 11.3).

La elucidación de la equivalencia meramente empírica no es tan inmediata y requiere de una decisión previa acerca de qué se debe entender por "igualdad de datos empíricos". En el espíritu de muchos autores está la idea de apelar a situaciones observacionales neutrales; sin embargo, en diversas ocasiones en esta obra hemos señalado el carácter problemático de la idea de una "observación pura", y hemos constatado la necesidad de separar en principio las nociones de *observabilidad y empiricidad y* admitir sólo una "empiricidad" *relativa* a cada teoría. Dentro de nuestro marco modeloteórico, esa noción viene fijada por el dominio de las aplicaciones intencionales de cada teoría. De acuerdo a esta interpretación, la equivalencia empírica entre dos teorías consistirá entonces en una equivalencia meramente al nivel de las aplicaciones intencionales: una correlación entre los dominios de aplicaciones intencionales de ambas teorías de tal naturaleza que, siempre que una aplicación intencional de una teoría sea extensible a un modelo actual de la misma (o sea, cumpla las leyes de esa teoría), entonces su correlato en la otra teoría cumplirá lo mismo, y recíprocamente. He aquí la especificación formal de esta idea (para simplificar, no hacemos mención en ella de las restricciones cruzadas o condiciones de ligadura, cf. cap. 10, §5).

Definición 11.5:

T_1 es *empíricamente equivalente* a T_2 si y sólo si existe una relación ε tal que:
(1) $\varepsilon \subseteq I(T_1) \times I(T_2)$.
(2) $\forall y_1, y_2 (<y_1, y_2> \in \varepsilon \rightarrow (y_1 \in \mathbf{r}[M(T_1)] \leftrightarrow y_2 \in \mathbf{r}[M(T_2)]))$.

Nótese que, en esta definición de equivalencia empírica, no se especifica nada acerca de cómo estén correlacionados los modelos potenciales de ambas teorías, si es que lo están de alguna manera; tampoco se dice nada acerca de los modelos actuales; en particular, no se infiere de ella que si un modelo actual x_1 de T_1 tuviera un correlato x_2 en T_2, este último sería necesariamente también un modelo actual de T_2.

5. Apéndice: Ciencia especial y ciencia básica; reducción, múltiple realizabilidad y superveniencia (*)

En esta última sección vamos a retomar el problema de la relación entre las ciencias especiales y la ciencia básica. Aunque en la literatura se plantea el problema sobre todo en relación a la psicología y la neurociencia (problema mente-cerebro), conceptualmente el problema es general. Se trata de precisar la supuesta "relación de dependencia" entre diversos pares de disciplinas científicas; por ejemplo, psicología/neurología, lingüística/psico-sociología, biología/química, o química/física; en realidad plantear esta cuestión para disciplinas enteras es inapropiado, lo adecuado sería hablar de la relación de dependencia entre teorías concretas de estos pares de disciplinas. La cuestión es pues determinar hasta qué punto las explicaciones de (teorías de) las ciencias especiales "descansan" en explicaciones de (teorías de) ciencias más básicas, hasta llegar eventualmente, mediante una cadena de sucesivas dependencias explicativas, a una supuesta ciencia básica (¿microfísica?). Hicimos algunas consideraciones preliminares sobre esta cuestión en el capítulo 5, cuando examinamos la noción de ley no estricta o ley *ceteris paribus* (§4), y también al comienzo de la sección 3 de este capítulo, al presentar la idea de reducción. Vamos a examinar ahora las diferentes posiciones al respecto con un poco más de detalle.

Aunque algunos aspectos de este problema se pueden tratar más satisfactoriamente desde una perspectiva modeloteórica global, vamos a limitarnos ahora a la perspectiva axiomática "enunciativa" clásica, pues así es como se presenta y discute en la literatura y los aspectos a que nos vamos a ceñir en este apéndice pueden abordarse de modo interesante ya en términos tradicionales.

Recordemos que en la perspectiva axiomática clásica las relaciones de dependencia o reducción se contemplan, no de forma global, sino de forma local, término-a-término (concepto-a-concepto, propiedad-a-propiedad). Se trata de ver hasta qué punto un recurso conceptual de una teoría es "dependiente en su función explicativa" de otros recursos conceptuales de otras teorías "más básicas". O en términos de propiedades, hasta qué punto unas propiedades "macro" dependen de, o se reducen a, propiedades "micro". La intuición, p.ej. en el caso de la psicología, es que puesto que los psiquismos, las mentes, están alojados en los cerebros y éstos están compuestos de neuronas, las propiedades psíquicas dependen de algún modo de propiedades neurológicas; o puesto que las sustancias químicas están formadas por partículas físicas, las propiedades químicas dependen de algún modo (son el resultado) de propiedades físicas; y análogamente en los restantes casos. A las primeras las vamos a considerar *teorías macro, y* a las segundas *teorías micro.* Antes de abordar directamente este problema vamos a presentar dos distinciones importantes en relación al mismo.

5.1. DISTINCIONES PREVIAS: TÉRMINOS GENERALES, CONCEPTOS EXPRESADOS Y ENTIDADES DENOTADAS; ACAECIMIENTO-EJEMPLAR Y ACAECIMIENTO-TIPO

Expresada en términos lingüísticos, la cuestión que vamos a tratar consiste en determinar cuál es la relación entre los conceptos expresados, y las propiedades denotadas, por los predicados de las teorías macro y los predicados de las teorías micro. La primera distinción tiene que ver con los diferentes niveles que se hallan involucrados en las diversas alternativas, el lingüístico, el conceptual o semántico, y el ontológico.

En lo que sigue, distinguiremos cuidadosamente entre: *a) los* términos generales o predicados, como `agua', `rojo', `sentir dolor', o *'H_2O'; b)* los conceptos (significados o contenidos conceptuales) expresados por los términos generales, como el concepto de *agua,* el de *rojo,* el de *sentir dolor, o* el de *molécula formada por dos átomos de hidrógeno y uno de oxígeno; y c)* las entidades, sustancias o propiedades, denotadas por los términos generales, como la sustancia agua, la propiedad de ser rojo, la de sentir dolor, etc. (a diferencia del resto de la obra, en este apéndice no usaremos cursivas para las mayúsculas que refieren a propiedades porque las cursivas se reservan para los conceptos). Puesto que a veces se usa `significado' de modo ambiguo, para referirse unas veces al concepto expresado y otras a la entidad denotada, en general tenderemos a no usar dicho término y hablar directamente de los conceptos expresados o las entidades denotadas; en la medida en que lo usemos, lo usaremos, salvo advertencia en contrario, con el primer sentido. Insistimos en que esta distinción es fundamental, pues se puede defender que aunque los significados conceptuales de dos predicados de dos ciencias son diferentes, ambos denotan la misma entidad.

La segunda distinción, en términos de la cual se suelen presentar las diferentes alternativas en la literatura, es entre *acaecimientos tipo ('type') y acaecimientos ejemplar ('token').* Recordemos (cf. cap. 5, §3) que los acaecimientos son determinada especie de entidades particulares. Un objeto particular es cualquier entidad espacial y/o temporalmente localizada (p.ej. el auto de Adela, esta pantalla de ordenador, el cuerpo calloso del cerebro de Quine, la imagen de la estatua de Colón en el córtex de Pedro ayer a las 14,30, etc.); los acaecimientos particulares (tanto los *procesos* como los *estados)* son cualquier cosa que *ocurre o sucede* en cierto lugar durante cierto intervalo temporal (p.ej. la batalla de Waterloo, el último partido de fútbol Barcelona-Madrid, la salida de Juan de la carretera ayer en la Costa Brava, etc.). Tanto objetos como acaecimientos son entidades particulares que pueden tener diversas propiedades. Un mismo objeto particular puede tener muchas propiedades diferentes (p.ej. esto que está aquí abajo tiene la propiedad de ser una silla, pero también las de ser azul, ser cómoda, estar aquí debajo, o ser mencionado en este libro); también un mismo acaecimiento particular puede tener diversas propiedades (p.ej. eso que ocurrió el martes sobre la estatua de Colón de Barcelona tiene la propiedad de ser la caída de un rayo, pero también las de ocurrir de día, asustar a Rosa, producir un cortocircuito en el funicular, ocurrir sobre la estatua de Colón, o ser mencionado en este escrito).

Cada particular concreto (objeto o acaecimiento) es un caso o *ejemplar ('token')* de las propiedades que ejemplifica. Dos ejemplares son del mismo *tipo ('type')* si comparten determinada propiedad. El auto de José y el de Adela son dos ejemplares diferentes de un mismo tipo (de objeto), Opel Corsa; la enfermedad de Rosa y la de Pedro son ejemplares diferentes de un mismo tipo (de proceso), infección gripal; la disfunción de María y la de Fernando son ejemplares diferentes de un mismo tipo (de estado), amnesia; los estados mentales de Enrique y Eugenia en la Nochevieja de 1996 son ejemplares de un mismo tipo, creencia de que en el Año Nuevo de 1997 lloverá. Dos particulares son o no del mismo tipo dependiendo de las propiedades que se tomen en consideración. Si consideramos cierta propiedad, el vehículo de José y el de Adela son del mismo tipo, un automóvil Opel Corsa, y de diferente tipo que el de Eduardo, un Seat Ibiza. Si consideramos otra propiedad, los tres son del mismo tipo, a saber, "vehículo a motor ligero de cuatro ruedas", y de diferente tipo, por ejemplo, que la motocicleta de Luis. Y aún, según otra propiedad, los coches de Adela, Eduardo y José y la motocicleta de Luis son del mismo tipo "vehículo terrestre a motor", y de diferente tipo que la bicicleta de Pedro o el barco de vela de Ana. Etcétera. Y lo mismo ocurre con los acaecimientos. Según se considere cierta propiedad, lo que le pasó a Juan ayer en la Costa Brava y lo que le pasó en Nochevieja a Rosa son acaecimientos del mismo tipo, accidentes de auto, y de diferente tipo a lo que le ha pasado esta mañana a Luis, un accidente de tren. Pero si consideramos otra propiedad más abstracta o general, los tres acaecimientos son del mismo tipo, accidentes, y de diferente tipo que lo acaecido en Año Nuevo a Marta, recibir un premio de lotería. Etcétera.

Así pues, hablar de tipos de objetos o acaecimientos no es en el fondo sino otro modo de hablar de determinada propiedad que ejemplifican. Esta distinción es importante para no confundir cuestiones diferentes. La pregunta acerca de si la creencia de Enrique de que lloverá en el Año Nuevo de 1997 es o no la misma entidad que el acaecimiento cerebral de tener las neuronas H en el estado 23, es ambigua. Una cosa es si son el mismo acaecimiento *token* y otra si son el mismo acaecimiento *type*, esto es, el mismo tipo de acaecimiento: si la propiedad de ser tal creencia es la misma propiedad que la de tener tales neuronas en tal estado. Como veremos, puede defenderse que son el mismo acaecimiento-ejemplar pero diferentes acaecimientos-tipo, es decir, que es un único acaecimiento particular que tiene dos propiedades *diferentes* (análogamente a como eso que ocurrió sobre la estatua de Colón tiene propiedades diferentes). En lo que sigue será esencial tener presente esta distinción.

5.2. IDENTIDAD CONCEPTUAL. REDUCCIONISMO SEMÁNTICO

El grado máximo de dependencia entre una ciencia especial y una ciencia básica, o mejor entre predicados de la primera y de la segunda, es el reduccionismo semántico: los dos predicados significan lo mismo, son sinónimos, expresan el mismo concepto. O, si se prefiere, uno da el significado del otro: el concepto expresado por el predicado `E' de la ciencia especial se reduce a, se identifica con, el concepto expresado por determinado predicado `B' de la ciencia básica. Se trata pues de una identidad entre los conceptos expresados o significados por ambos predicados.

Ejemplos independientes de la relación entre ciencias especiales y ciencia básica provienen de los casos usuales de sinonimia en los que una expresión explicita el significado conceptual de otra. Por ejemplo, `soltero' y `varón adulto no casado'; o más interesante, `agua' y (supongamos) `sustancia inodora e insípida, que en estado líquido es (sin impurezas) incolora y que (en diversas disoluciones) conforma los lagos, ríos y mares; que en estado sólido constituye las nieves, hielos, etc.'. Así, aunque las expresiones lingüísticas `agua' y `sustancia incolora, [etc.]' son expresiones lingüísticas diferentes, los conceptos *agua* y *sustancia incolora, [etc.]* son el mismo concepto, análogamente a como las diferentes expresiones `silla' y `chair' expresan el mismo concepto *silla* (aunque en este caso una no "da" el significado de la otra).

Ésta es la tesis que defendía el *conductismo lógico* sobre la relación entre lo mental y lo conductual. Según los conductistas lógicos (cf. p.ej. Hempel, 1949; Ryle, 1949 y Wittgenstein, 1958), los predicados mentales expresan conceptos conductuales disposicionales. Por ejemplo, `tener sensación de dolor' y `tener la disposición a chillar en tales circunstancias, a retorcerse en tales otras, a [etc.]' tienen el mismo significado conceptual; el concepto *sensación de dolor* y el concepto *tener la disposición a chillar si ..., a retorcerse si... [etc.]* son el mismo concepto, en el mismo sentido en que *agua* y *sustancia inodora, insípida [etc.]* son el mismo concepto. Y análogamente, por ejemplo, con los predicados `creer que va a llover la próxima hora' y (p.ej.) `tener la disposición a tomar un paraguas si se desea salir de casa, a recoger la ropa si no se quiere que se moje, a [etc.]'. Si el conductismo lógico fuese correcto, esto sucedería con todo predicado mentalista.

El modo de evaluar una hipótesis sobre una identidad conceptual específica es determinando si es o no *conceptualmente posible* que se ejemplifique una propiedad sin que se ejemplifique otra. Si `E' expresa el mismo contenido conceptual que `B', entonces una situación en la que un particular tenga la propiedad E y no tenga la propiedad B es conceptualmente imposible. Alternativamente: si tal situación es conceptualmente posible (incluso aunque no sea nómicamente posible), entonces los conceptos expresados por ambos predicados no pueden ser el mismo (sobre posibilidad conceptual y nómica, cf. cap. 5, §1). Ésta es la estrategia que usó Putnam para "refutar" el conductismo lógico. Putnam (cf. 1963) diseñó un experimento mental que a su juicio presenta una situación perfectamente concebible (aunque quizá biológicamente imposible) en la que unos sujetos (super-super-espartanos, los denomina) tienen dolor pero no tienen ninguna disposición a la conducta, no ya gestual sino ni siquiera verbal.

El conductismo lógico radical tiene otros problemas, como los derivados del holismo de lo mental, que no vamos a comentar aquí. Tras su, por lo general reconocido, fracaso, algunos filósofos de la psicología han propuesto otra alternativa también reduccionista conceptual pero mucho más plausible. Se trata del *funcionalismo analítico,* según el cual los predicados mentalistas significan conceptos funcionales, donde un concepto funcional es un concepto que establece las relaciones de causa-efecto, el *rol causal,* de los estados de un sistema computacional (cf. p.ej. Putnam, 1967; Fodor, 1968 y Lewis, 1972; para una buena exposición, García-Carpintero, 1995).

5.3. IDENTIDAD DE TIPOS O PROPIEDADES. REDUCCIONISMO ONTOLÓGICO

El reduccionismo semántico, incluso si se da en algunas ocasiones (por ejemplo, entre predicados mentalistas y funcionalistas, si los funcionalistas analíticos tienen razón), es demasiado fuerte para dar cuenta de todas las situaciones en que consideramos intuitivamente que unas explicaciones dependen de otras. Es obvio que si los conceptos expresados son los mismos, las propiedades denotadas también lo serán. Pero muchas veces ocurre lo segundo sin lo primero, los predicados denotan la misma propiedad aunque signifiquen conceptos diferentes. Este tipo de situación en la que términos lingüísticos que expresan contenidos conceptuales diferentes denotan o refieren una misma entidad son conocidas de antiguo y tematizadas en semántica al menos desde Frege. La distinción fregeana entre el sentido y la referencia de una expresión pretende justamente dar cuenta de ella; esto es lo que sucede con las descripciones \`la mujer de Edipo' y \`la madre de Edipo', puesto que ambas nombran a Yocasta, o en el ejemplo preferido de Frege, entre \`la estrella de la mañana' y \`la estrella de la tarde', que nombran a Venus.

Pues bien, algo análogo sucede con algunos predicados o términos generales. Los casos interesantes, a nuestros actuales efectos, son aquellos en los que (contrariamente a lo que sucede con las diferentes descripciones para Venus o Yocasta) uno de los predicados se puede considerar más básico o más fundamental que el otro. Esto es lo que sucede, por ejemplo, con \`agua' *y* '$H_2O$', o con \`temperatura' y \`energía cinética media'. Hemos visto que \`agua' no significa conceptualmente lo mismo que 'H_2O', \`agua' no expresa el concepto *sustancia constituida por moléculas formadas por dos átomos de hidrógeno y uno de oxígeno. Si* significara dicho concepto, todo usuario competente del predicado \`agua' debería poseer dicho concepto, lo que no es el caso; dicho término se ha usado competentemente durante siglos antes del descubrimiento de la química molecular, antes de disponer del concepto de *molécula, y* todavía hoy muchos de los usuarios competentes de dicho término no tienen ni idea de química. Y lo mismo sucede con \`temperatura', un predicado del lenguaje ordinario, y también de una teoría científica sencilla, la termodinámica fenomenológica, y \`energía cinética media', un predicado de la mecánica estadística. Aunque \`temperatura' y \`energía cinética media' significan conceptos diferentes (el primero se usaba correctamente antes de saber nada de mecánica, y menos de mecánica estadística) de hecho denotan la misma magnitud física (algo que ignoran la mayoría de los usuarios competentes del primer predicado). La temperatura *es* la energía cinética media, como el agua *es* H_2O. En este sentido una propiedad se *reduce a, depende de, o descansa en* otra; no son sólo los mismos fenómenos-ejemplar, sino los mismos fenómenos-tipo. Por tanto, no hay en realidad dos propiedades diferentes tales que una descanse en *otra.* Sólo hay diferentes conceptos, la propiedad es la misma.

El reduccionismo ontológico, la identidad de propiedades, es la posibilidad más fuerte después del reduccionismo semántico. El modo más radical de explicar cómo es que fenómenos que describimos *mediante aparatos conceptuales diferentes* son tales que uno descansa en otro, consiste en que no haya en realidad *dos tipos* de fenómenos sino sólo uno. Ésta es la tesis que defienden en filosofía de la psicología los llamados *teóricos de la identidad psicofísica* (cf. p.ej. Feigl, 1958 y Smart, 1959). Según estos autores, aunque el concepto *sentir dolor* no es el mismo concepto que *tener las fibras H activadas,* la propiedad de sentir dolor es de hecho la propiedad cerebral de tener las fibras H activadas, en exactamente el mismo sentido en que ser de agua es la misma propiedad que ser de H_2O, o "tener mayor temperatura que" es la misma propiedad (relacional) que "tener mayor energía cinética que". Los predicados mentalistas son nombres diferentes, que expresan conceptos diferentes, para las propiedades cerebrales.

5.4. MÚLTIPLE REALIZABILIDAD

Las hipótesis sobre identidades ontológicas son hipótesis empíricas y se deben evaluar por tanto empíricamente, investigando si de hecho el predicado `E' denota efectivamente la misma propiedad o sustancia que el predicado `B'. En muchos casos es sencillo ver que no es así, que el reduccionismo ontológico es todavía una hipótesis demasiado fuerte. El ejemplo más sencillo lo ofrecen las propiedades disposicionales, como las denotadas por los predicados `elástico', `soluble', `frágil' o `rojo'. Estos predicados expresan un concepto según el cual un objeto tiene la propiedad en cuestión si en determinadas circunstancias reacciona de cierto modo (cf. cap. 8, §4). Así, por ejemplo, un objeto es frágil si en caso de que se aplicara sobre él determinada presión tangencial el objeto se quebraría; o una superficie es roja si en caso de que incidiera sobre ella luz blanca la superficie absorbería tales frecuencias del espectro. Pues bien, las propiedades disposicionales *descansan* en propiedades físicas. Si un objeto es frágil lo es *en virtud* de ser microfísicamente como es; si una superficie es roja lo es *en virtud* de ser microfísicamente como es. Se dice entonces que las propiedades disposicionales macro *se realizan* mediante propiedades físicas micro. Lo característico de estos casos es que ahora no podemos explicar esta dependencia entre propiedades macro y propiedades micro del modo más sencillo, a saber, mediante la identidad de propiedades, pues claramente *cada propiedad disposicional no se puede identificar con una única propiedad microfísica.* Las propiedades disposicionales *se realizan* mediante propiedades microfísicas, pero simplemente ocurre que la propiedad microfísica que realiza la propiedad disposicional no es la misma en todos los casos. En cada caso particular de objeto frágil, la fragilidad "se debe" a cierta propiedad microfísica del objeto, pero en diferentes objetos la propiedad realizadora es diferente. En unos objetos, por ejemplo los de yeso, la fragilidad se realiza mediante una propiedad física; en otros, por ejemplo los de vidrio, se realiza mediante otra diferente. El yeso y el vidrio son ambos frágiles y sin embargo microfísicamente no tienen nada en común (salvo, claro está, que ambos "realizan" la fragilidad). Lo mismo sucede con el resto de las propiedades disposicionales. Por ejemplo, una determinada tela y una determinada superficie plástica pueden ser ambas rojas, pero microfísicamente no tienen nada en común; la propiedad microfísica que realiza la rojez es diferente en cada caso.

A esta característica, ejemplificada típicamente por las propiedades disposicionales, se la suele denominar *múltiple realizabilidad.* Una propiedad macro es múltiplemente realizable si: *a)* el que cada objeto particular la tenga depende de que el objeto tenga determinada propiedad micro, pero *b)* en diferentes objetos particulares la propiedad corresponde a diferentes propiedades micro. En tal caso, no sólo el concepto expresado por el predicado macro es diferente al expresado por predicados micro, sino que ni siquiera se puede *identificar* la propiedad macro con *una* propiedad micro determinada. Cuando las propiedades macro son múltiplemente realizables no es posible explicar la dependencia entre propiedades macro y propiedades micro *reduciendo o identificando* las primeras con las segundas. En estos casos la identidad de tipos-propiedades, el reduccionismo ontológico, no es una explicación viable. Se puede pretender quizá que la propiedad macro es idéntica, no a una propiedad micro "atómica" sino a una propiedad micro "disyuntiva". Así, si la propiedad macro E se realiza múltiplemente mediante las propiedades atómicas micro *B 1, B2, ..., Bn* se podría identificar quizá la propiedad E con la "propiedad disyuntiva" *B 1-o-B2-...-o-Bn.* Sin embargo esta estrategia del reduccionista requiere aceptar que cualquier disyunción (en general combinación) de propiedades es también una propiedad, lo cual en opinión de muchos requiere a su vez una metafísica de las propiedades inaceptable. Es muy implausible que cualquier predicado molecular denote una propiedad, al menos una propiedad *natural,* que es de lo que aquí se trata. Una vez más surge aquí la cuestión de la diferencia entre propiedades "naturales" y "no naturales" en la que no vamos a detenernos aquí (cf. cap. 5, *§2 y §*6, y cap. 7, §5 y §6).

Según muchos críticos de la teoría de la identidad psicofísica, esta teoría está condenada al fracaso precisamente porque con las propiedades mentales pasa lo mismo que con las disposicionales, a saber, que aunque se realizan mediante propiedades neuro-bio-físicas, son también múltiplemente realizables y por tanto no se puede *identificar* cada propiedad mental con una propiedad neuro-bio-física (de hecho, para muchos, las propiedades mentales son un tipo de propiedades disposicionales, a saber, propiedades funcionales). No sólo atribuimos algunas propiedades mentales básicas (p.ej. percepción de formas) a seres que no tienen una morfología nerviosa como la nuestra (p.ej. pulpos), sino que, limitándonos a los humanos, ocurre que un mismo proceso mental, especialmente en capacidades complejas, puede realizarse mediante procesos cerebrales de diferente tipo; por ejemplo, cuando una zona del cerebro ha resultado dañada, una zona vecina morfológicamente diferente asume la función de la dañada (en esto consiste parte de la denominada *plasticidad* del cerebro).

5.5. DUALISMO DE PROPIEDADES CON IDENTIDAD DE EJEMPLARES Y SUPERVENIENCIA

Cuando las propiedades macro se realizan múltiplemente mediante propiedades micro, no podemos explicar la dependencia de las primeras respecto de las segundas apelando a la identidad de tipos, a que simplemente las propiedades son idénticas. Y sin embargo hay que elucidar esa dependencia de algún modo, pues en un sentido preanalítico intuitivo unas "descansan" en otras (claramente en casos como las disposicionales y quizá también en otros casos). La dificultad reside en que en estos casos las propiedades micro realizadoras son efectivamente *otras,* esto es, *diferentes* a las propiedades macro que realizan.

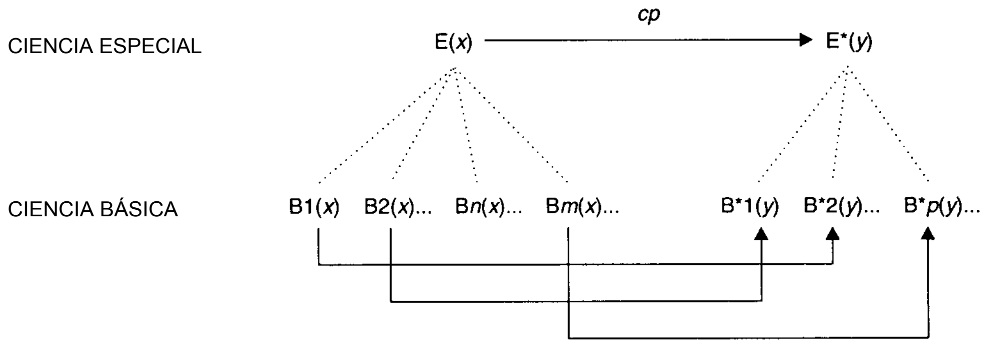
Así pues, cuando hay múltiple realizabilidad la identidad de tipos es inviable. Se pensará que todavía queda la identidad de ejemplares, pero la identidad de ejemplares *por sí sola* es demasiado poco, no puede dar cuenta de la dependencia entre los dos niveles. Pensemos en el acaecimiento en que se vio envuelto Juan ayer en la Costa Brava. Tenemos un mismo acaecimiento-ejemplar con diversas propiedades, por ejemplo, la propiedad de ser un accidente y la propiedad de ocurrir en primavera. Ambas propiedades son propiedades diferentes *del mismo* acaecimiento particular. Y no pensamos que haya una dependencia entre ambas propiedades, no pensamos que el que haya sido un ejemplar del tipo "accidente" depende de que haya sido un ejemplar del tipo "ocurrido en primavera", o viceversa. Sin embargo, en muchos casos, como los que involucran propiedades disposicionales, sí consideramos que hay tal dependencia. Aceptamos que un mismo objeto particular tiene diversas propiedades, la propiedad de ser frágil y la propiedad de tener tal y cual estructura microfísica, pero no nos basta con eso, pues creemos además que el que dicho objeto particular tenga la primera propiedad *depende de* que tiene la segunda. Aceptamos que un mismo acaecimiento-ejemplar tiene diversas propiedades, la propiedad de ser una sensación de dolor y la propiedad de activarse las fibras H, pero no nos basta con eso pues creemos además que el que dicho acaecimiento particular tenga la primera propiedad *depende de* que tiene la segunda. Debe quedar claro que si lo único que tenemos es identidad de ejemplares dicha dependencia queda inexplicada.

Para dar cuenta de estos casos se recurre a la noción de *superveniencia,* un tipo de dependencia más débil que la identidad de tipos (aunque la incluye como caso extremo). La idea es la siguiente. Lo que distingue a los casos en que no hay dependencia (como el del accidente) de aquellos en que sí la hay (como el de la fragilidad) es, por decirlo tautológicamente, que en el primero la ejemplificación de una propiedad no depende de la ejemplificación por el mismo particular de otra propiedad. Es decir: el particular puede tener una propiedad y no tener la otra. El suceso de la Costa Brava *podría* haber sido un accidente sin haber ocurrido en primavera. Que un acaecimiento particular ejemplifique la propiedad de ser un accidente no depende de que se ejemplifique la de ocurrir en primavera. Eso no es así en los otros casos. Que un particular tenga la propiedad de ser frágil depende de que tenga determinada propiedad microfísica en el siguiente sentido: no puede ser que tenga ésta y no tenga aquélla. En general, pues, una propiedad macro E depende de (descansa en, se realiza mediante) una propiedad micro B si *no es (físicamente) posible* que un particular x ejemplifique B y no ejemplifique E. Nótese que la conversa puede no ser cierta, y de hecho no lo es, debido a la múltiple realizabilidad pues E puede realizarse mediante otras propiedades micro B', B", ..., por lo que x puede ser E sin ser B (la conversa sólo es cierta cuando vale la identidad de tipos).

Ésta es la idea de dependencia que intenta expresar la noción de *superveniencia.* Una propiedad A superviene en otra B si no puede ser que un particular ejemplifique B y no ejemplifique A. En realidad la relación de superveniencia se suele caracterizar, no para propiedades sueltas, sino para grupos de propiedades. Así, por ejemplo, decimos que las propiedades cromáticas supervienen sobre propiedades microfísicas si no puede ser que dos particulares tengan las mismas propiedades microfísicas y no tengan la misma propiedad cromática; o que las propiedades mentales supervienen sobre propiedades bio-físicas si no puede ser que dos sujetos estén en el mismo tipo de estado bio-físico *(el mismo* en serio, e.e. compartiendo *todas* las propiedades bio-físicas) y estén en diferente estado mental. En general, las propiedades de la clase (D supervienen sobre propiedades de la clase `F si y sólo si dos particulares (objetos o acaecimientos) que tienen las mismas propiedades T tienen también las mismas propiedades (D. En este sentido "lo (V descansa, depende o se realiza en "lo 'F". Esta idea general se puede precisar de diferentes modos, pero no vamos a detenernos aquí en ellos (cf. Kim 1984 y 1987 y Savellos y Yalcin [eds.] 1995).

Para muchos autores, la "identidad de ejemplares con múltiple realizabilidad (e.e. sin identidad de tipos) más superveniencia" es la situación más común en las ciencias especiales. Las propiedades de las que se ocupan las ciencias especiales supervienen sobre propiedades microfísicas, éste es el grano de verdad que hay en el *fisicalismo.* Pero por lo general la superveniencia no es identidad de tipos, las propiedades macro son múltiplemente realizables; la pretensión contraria es el grano de falsedad que hay en el fisicalismo *reduccionista.* En algunos casos la superveniencia puede derivarse de la identidad, quizá por ejemplo algunas propiedades químicas básicas son idénticas a propiedades físicas complejas. Pero a determinado nivel eso es muy implausible. Propiedades biológicas mínimamente "elevadas" son múltiplemente realizables, no se pueden identificar con una única propiedad bioquímica. Y lo mismo sucede, como hemos visto, con las propiedades mentales respecto de las propiedades neurológicas. Debe notarse que ésta es una tesis empírica, compatible con tesis conceptuales complementarias. Así, por ejemplo, en psicología se puede defender como tesis conceptual cierto reduccionismo semántico, como hacen los funcionalistas analíticos, según el cual los predicados mentalistas expresan conceptos funcionales-computacionales, y a la vez defender como tesis empírica que los predicados mentalistas denotan propiedades (funcionales) que supervienen sobre, y se realizan múltiplemente en propiedades neurológicas.

Además, para algunos de estos autores, como Fodor (cf. para lo que sigue Fodor, 1974), esta situación de las ciencias especiales explicaría el carácter no estricto de las leyes de las ciencias especiales, esto es, que sus leyes sean leyes *ceteris paribus,* que tengan excepciones (sobre estas nociones, cf. cap. 5, §4), aun cuando las leyes de la ciencia básica no las tengan. La existencia de excepciones se derivaría de que alguna de las propiedades micro que realizan la propiedad del antecedente de la ley macro podría no estar causalmente conectada con alguna propiedad de las que realizan la propiedad del consecuente de dicha ley. La situación se recoge en el siguiente gráfico.



Esta situación ilumina, según Fodor, un hecho aparentemente paradójico, a saber, que las leyes de las ciencias especiales tengan excepciones a pesar de que descansan sobre leyes de la ciencia básica que no tienen excepciones. Si la situación de dependencia es la mostrada por el gráfico, entonces podemos explicar las excepciones de la ley especial aun en los casos en que todas las leyes de la ciencia sobre la que descansa sean estrictas. Las excepciones corresponderían a los casos en que la propiedad macro (E) del acaecimiento-tipo antecedente de la ley especial se realiza mediante alguna propiedad micro (p.ej. *Bn)* que no está nómicamente conectada mediante una ley básica con ninguna propiedad micro que realiza la propiedad macro (E*) del acaecimiento-tipo consecuente (de todas formas, el propio Fodor ha cuestionado la viabilidad de esta explicación cuando se trata de propiedades funcionales, cf. Fodor, 1991).

Para concluir con la noción de superveniencia conviene enfatizar que esta noción por sí sola simplemente elucida qué entendemos por dependencia, pero (salvo en el caso extremo de que la superveniencia se derive de la identidad) no explica metafísicamente a qué se debe tal dependencia. ¿Cómo es que unas propiedades supervienen sobre otras en los casos en que no hay identidad?, ¿qué vínculo metafísico hay entre ambas propiedades del que se deriva la superveniencia? Una posibilidad es tomar la relación de superveniencia como un primitivo metafísico bruto, algo que muchos se niegan a aceptar. Los recelos hacia esta posición hacen que algunos terminen cuestionando la legitimidad de las propiedades macro y sostengan que hay *conceptos* macro pero no *propiedades* macro; los predicados macro serían términos como `jade' o `cáncer', que no denotan una única propiedad, son términos ambiguos que en cada ocasión denotan alguna de entre varias propiedades básicas. O de otro modo, si se quiere considerar que denotan propiedades, éstas deben ser en todo caso propiedades "de segundo orden", donde una propiedad de segundo orden es la "propiedad consistente en tener alguna de entre tales y cuales propiedades básicas".

5.6. DUALISMO DE PROPIEDADES CON IDENTIDAD DE EJEMPLARES Y EPIFENOMENISMO

En relación con el último problema apuntado surge otra cuestión relativa a la eficacia causal de las propiedades macro y que aquí sólo apuntaremos. La cuestión es la siguiente (ignoraremos ahora las excepciones de la ley macro). Si cada acaecimiento antecedente de la ley especial es también un acaecimiento antecedente de una ley básica (de diferentes leyes en diferentes casos) y las propiedades micro son sin duda causalmente eficaces, ¿para qué son necesarias las propiedades macro? Un acaecimiento particular causa o produce otro en virtud de que ejemplifica determinada propiedad (cf. cap. 5, §3). Si hay causalidad básica, la propiedad en virtud de la cual un acaecimiento causa otro es determinada propiedad micro Bi, pero entonces parece que la propiedad macro E es superflua, causalmente ineficaz o causalmente redundante. Éste es *el problema de la exclusión explicativa* (cf. p.ej. Kim, 1987), y es lo que motiva las posiciones *epifenomenistas.* Se dice que una propiedad A es un epifenómeno de una propiedad P si A "se debe" a la presencia de B pero A es causalmente ineficaz. En filosofía de la psicología se ha atribuido esta posición al *monismo anómalo* de Davidson (aunque él mismo no acepta ser calificado de epifenomenista, cf. Davidson, 1970 y McLaughlin, 1984).

5.7. ELIMINACIONISMO

El epifenomenismo acepta la existencia de las propiedades macro, la identidad de ejemplares y la superveniencia, pero rechaza que las propiedades macro sean causalmente eficaces. Son "un lujo ontológico", por así decir. Esta posición le parece inaceptablemente conservadora al eliminacionista, que comparte con aquél los recelos sobre la eficacia causal de las supuestas propiedades macro pero no tiene sus escrúpulos conservadores. Si no hay eficacia causal, no hay propiedad. Según el eliminativista, los predicados macro expresan conceptos a los que no corresponde ninguna propiedad en el mundo, como ocurrió con el predicado 'flogisto' de la química del siglo xviii. Hay concepto pero no hay propiedad, por tanto lo mejor, como en el caso del flogisto, es dejar de emplear el término, eliminarlo del lenguaje de la ciencia (exactamente igual que se eliminaron del lenguaje de la meteorología los nombres de los dioses griegos y de sus "estados"). En filosofía de la psicología defienden una postura semejante los Churchland (cf. p.ej. Churchland, 1981 y 1984). Debe quedar claro que el eliminativismo ni siquiera admite la identidad de ejemplares, el motivo es obvio: si no hay propiedades macro (ni idénticas ni diferentes de las propiedades micro), entonces ni siquiera se puede plantear la cuestión de si un evento que ejemplifica una propiedad macro es o no el mismo acaecimiento-ejemplar que ejemplifica una propiedad micro; no hay propiedades macro y no hay por tanto acaecimientos que tienen propiedades macro.

5.8. DUALISMO DE EJEMPLARES

El eliminacionismo (acerca de una determinada ciencia especial) rechaza la identidad de ejemplares, pero porque no hay propiedades macro (denotadas por los predicados de dicha ciencia especial). Pero se puede rechazar la identidad de ejemplares incluso si se acepta la existencia de propiedades macro. Esto es lo que defiende el dualista de ejemplares. Hay efectivamente propiedades de los dos niveles y son propiedades diferentes, pero se ejemplifican en acaecimientos particulares también diferentes. En filosofía de la psicología se puede asimilar a esta posición los diversos tipos de dualismo de sustancias (p.ej. Descartes, Eccles) y de paralelismos (p.ej. la *armonía preestablecida* de Leibniz o el *ocasionalismo* de Malebranche).