Tim Maudlin

FILOSOFÍA de la FÍSICA

I

BREVIARIOS

BREVIARIOS

del

FONDO DE CULTURA ECONÓMICA

588

Traducción
MARIANO SÁNCHEZ-VENTURA

Revisión técnica
ELIAS OKON

Tim Maudlin

Filosofía de la física

I. El espacio y el tiempo

FONDO DE CULTURA ECONÓMICA

Primera edición en inglés, 2012
Primera edición en español, 2014
Primera edición electrónica, 2014

Diseño de portada: Paola Álvarez Baldit

Título original: *Philosophy of Physics. Space and Time*
© 2012, Princeton University Press

D. R. © 2014, Fondo de Cultura Económica
Carretera Picacho-Ajusco, 227; 14738 México, D. F.
Empresa certificada ISO 9001:2008

www.fondodeculturaeconomica.com

Comentarios:
editorial@fondodeculturaeconomica.com
Tel. (55) 5227-4672

Se prohíbe la reproducción total o parcial de esta obra, sea cual fuere el medio. Todos los contenidos que se incluyen tales como características tipográficas y de diagramación, textos, gráficos, logotipos, iconos, imágenes, etc., son propiedad exclusiva del Fondo de Cultura Económica y están protegidos por las leyes mexicanas e internacionales del copyright o derecho de autor.

ISBN 978-607-16-2769-8 (mobi)

Hecho en México - *Made in Mexico*

A la memoria de Robert Weingard,
explorador de las sendas del universo

ÍNDICE

RECONOCIMIENTOS

Las raíces de este volumen se remontan en el tiempo hasta mi ciclo académico de posgrado, época en que Clark Glymour tuvo la amabilidad de impartir un seminario de un año sobre la relatividad a petición de los estudiantes de posgrado en Historia y Filosofía de la Ciencia en la Universidad de Pittsburgh. Sin el beneficio de esa paciente y exhaustiva presentación de los métodos matemáticos modernos, yo no hubiera sido capaz de empezar a pensar en esta teoría de una manera fundamentalmente geométrica. Los seminarios de John Norton nos revelaron el fondo histórico de la cuestión a partir de Newton y Peter Machamer nos guió a través de Galileo. Por entonces, John Earman y el mismo Norton recién habían formulado el argumento del agujero, enigma con que me fogueé teóricamente de tantas vueltas que le di. En suma, este libro es el resultado de mis reflexiones a lo largo de un cuarto de siglo en torno a esa efervescente profusión de ideas que fue la fuente vital de aquel extraordinario programa académico.

Cuando llegué a Rutgers en 1986, tuve la inconmensurable fortuna de encontrar en Robert Weingard a un amigo y un colega. Su naturaleza inquisitiva y su honestidad intelectual hacían que nuestras discusiones fueran siempre una delicia; saqué mucho provecho de su profundo conocimiento de la física. A él le dedico con profunda gratitud y hondo afecto esta obra.

También tengo otra especie de deuda con los muchos estudiantes, tanto de posgrado como de pregrado, a quienes he tenido el privilegio de enseñarles física a través de los años. La exposición de la teoría del espacio-tiempo que aquí se encuentra hubo de evolucionar lentamente a través de muchas de aquellas clases. Al principio asumí las explicaciones estándar, utilizando profusamente las coordenadas y las transformaciones de las coordenadas. Poco a poco, clase tras clase, las referencias a las coordenadas fueron quedando al margen, de modo que la geometría fundamental finalmente quedó al descubierto y fue posible inspeccionarla de forma directa. Mi presentación de la relatividad, en particular, resulta poco ortodoxa, pero creo (toco madera) que es conceptualmente clara. Espero que al menos le haya ahorrado al lector algunas de las confusiones que yo tuve que esforzarme mucho en superar.

El manuscrito se ha beneficiado de la retroalimentación y los comentarios de muchas personas. Agradezco particularmente la contribución de dos lectores anónimos de la editorial: espero que consideren mejor la versión final que la que ellos leyeron. Sean Carrol insistió acertadamente en que yo revisara a fondo los detalles respecto a la relatividad general y gracias a sus recomendaciones el capítulo vi ha mejorado mucho. Adam Elga aportó útiles comentarios, al igual que el seminario de filosofía de la ciencia en la Universidad de Nueva York. El infatigable Bert Sweet revisó los detalles y los cálculos con el característico cuidado que tanto lo distingue.

Estoy agradecido también con Scott Soames, Rob Tempio y la editorial Princeton University Press por haberme permitido ampliar a dos volúmenes el manuscrito que

pretendía abarcar sólo uno sobre filosofía de la física. No me es grato pensar en los compromisos e imperfecciones que la imposición de un solo volumen hubiera requerido.

En un nivel más práctico, el tiempo necesario para completar el manuscrito me lo brindó un semestre sabático en Rutgers. *Merci*.

Finalmente, como siempre, Vishnya Maudlin ha hecho mucho más que soportar la obsesión que me acompaña en la creación de un libro. Ella siempre ha estado allí, tanto en el aula como en el hogar, dispuesta a discutir y criticar y clarificar. No me puedo imaginar asumiendo solitariamente una tarea como ésta. No hubiera sido posible sin su ayuda.

INTRODUCCIÓN: OBJETIVO Y ESTRUCTURA DE ESTOS VOLÚMENES

La filosofía de la física se ocupa de la realidad física en su totalidad, considerada en una forma provechosamente genérica. Por ejemplo, el mundo físico al parecer posee aspectos espaciales y temporales; por lo tanto, la existencia y la naturaleza del espacio y el tiempo (o el espacio-tiempo) es un tema esencial. También lo es la materia, esa cosa de que se componen las mesas y las sillas y los planetas. Al decir "una forma provechosamente genérica" doy a entender lo siguiente: la pregunta más general que sea posible hacer respecto a la materia consiste en inquirir qué tipo de cosa es. Por ejemplo, podríamos argüir que la materia se compone de partículas parecidas a puntos, o de campos, o de cuerdas unidimensionales, o de una combinación de tales cosas, o de algo totalmente diferente. Dada cualquiera de estas explicaciones generales, todavía existen cuestiones específicas en todos los casos: ¿cuántos tipos de campos hay?, ¿cuál es la masa de las partículas?, y así por el estilo. Nosotros nos dedicaremos a las cuestiones más generales, en vez de a las más específicas.

La filosofía de la física, como disciplina, es una prolongación de la física propiamente dicha. Los tipos de preguntas que aquí plantearemos se encuentran entre las que los físicos mismos se plantean y a las que muchas teorías de la física han tratado de responder a lo largo de la historia. Pero una asombrosa cantidad de investigación física puede elaborarse sin poseer las respuestas a semejantes preguntas. Por ejemplo, la ciencia de la termodinámica inicialmente pretendía (como su nombre lo sugiere) explicar matemáticamente la forma en que el calor se extiende a través de un objeto y de un objeto a otro. Sin embargo, se pueden encontrar ecuaciones bastante detalladas de cómo el calor fluye, sin tener idea de lo que el calor *es*. ¿Se trata de una especie de fluido (como sostiene la teoría calórica) que literalmente fluye de un objeto a otro, o se trata de un tipo de movimiento (como sostiene la teoría cinética) que se transmite de un cuerpo al otro al interactuar? Si todo lo que te importa es saber cuánto tardará en enfriarse hasta 37 °C una barra de hierro de 5 kg con una temperatura de 93 °C cuando se le sumerge en un recipiente con agua a una temperatura de 10 °C, las ecuaciones del flujo del calor pueden darte la respuesta. Pero no habrás de saber nada en realidad sobre la naturaleza fundamental del calor cuando hayas hecho el cálculo. A un herrero la naturaleza del calor le podría importar un bledo; asimismo, a un filósofo de la física poco le importaría saber cuánto tiempo tardaría en enfriarse la barra de hierro. A un físico usualmente le importarían ambos aspectos, aunque enfocara más uno que el otro en diferentes ocasiones. Hoy en día es característico en la enseñanza de la física que se invierta mucho más tiempo en aprender cómo solucionar una ecuación que en discutir las cuestiones más "filosóficas" sobre la naturaleza del calor, o sobre la naturaleza del espacio y el tiempo, o sobre la naturaleza de la materia. Los estudiantes de física a quienes fascinan estas cuestiones fundamentales suelen sentirse defraudados por las clases de física en que se dejan totalmente de lado estos temas. Esta obra está dedicada tanto a ellos como a los filósofos que se interesan en la realidad física.

En esta obra la filosofía de la física se ha dividido en tres partes que ocupan dos volúmenes. Cada uno puede leerse independientemente del otro. Pero ciertos temas —en especial la necesidad de una explicación totalmente física de los procedimientos de la "medición"— se plantean en ambos volúmenes, de manera que leerlos en orden es lo más recomendable. El primero se relaciona con la naturaleza del espacio y el tiempo. Contiene una breve historia de los debates en torno al espacio y el tiempo, desde la física clásica hasta la relatividad general. En la física, el espacio y el tiempo (posteriormente, el espacio-tiempo) constituyen el escenario en que se desarrolla la historia del universo físico. Pero el espacio y el tiempo mismos son entes elusivos. El mundo físico se nos presenta como una serie de cosas y eventos en el espacio que coexisten o que se suceden uno tras otro en el tiempo. Pero el espacio y el tiempo por sí mismos no se muestran a nuestros sentidos: carecen de color o de sabor o de sonido o de aroma o de sustancia tangible. Lo que sí parecen tener el espacio y el tiempo es una estructura geométrica. Habremos de examinar varias teorías acerca de lo que *es* precisamente esta estructura, así como sobre qué es lo que *tiene* esa estructura. La teoría de la relatividad se expone en primer lugar, y ante todo, como una teoría de la geometría del espacio-tiempo. La relatividad especial se explica en suficiente detalle como para resolver problemas específicos sobre el comportamiento de los relojes y los objetos rígidos en un mundo relativista. La relatividad general, sin embargo, no se expone con tanto rigor. Mi objetivo ha sido presentar con absoluta claridad los fundamentos conceptuales de estas teorías, enfocando particularmente la forma en que la utilización de las coordenadas en la física se relaciona con la estructura geométrica subyacente.

En el segundo volumen se recoge la teoría de la materia. La primera parte de éste presenta la actual teoría de la materia: la teoría cuántica. A diferencia de lo que ocurre con la relatividad, los físicos no se han puesto de acuerdo en cómo entender la teoría cuántica. De hecho, el nombre mismo de "teoría cuántica" es erróneo: no existe tal teoría. Más bien, existe un formalismo matemático y algunas (muy eficaces) reglas empíricas que indican cómo utilizar ese formalismo para efectuar ciertos tipos de predicciones. Aquí la diferencia entre el herrero y el filósofo de la física se agudiza. Al herrero (o al físico en cuanto forjador) en verdad poco le importa la naturaleza de la realidad física: le basta calcular el resultado de diversos experimentos. El filósofo de la física se interesa en la realidad subyacente y sólo le importan las predicciones en la medida en que puedan servir como pruebas de la validez de una explicación científica de la realidad subyacente. En este volumen examinaremos algunas de las explicaciones concurrentes sobre la naturaleza de la materia. Estas teorías tienen en común buena parte de las matemáticas de la teoría cuántica, pero sus explicaciones de la existencia de la naturaleza difieren radicalmente.

Si el primer volumen abarca el espacio-tiempo y la primera parte del segundo volumen el contenido material del espacio-tiempo, parecería que ya no habría nada más que discutir. ¿No hemos lamido ya todo el plato? En cierto sentido sí: todo lo que existe en el mundo físico —en un nivel fundamental— se explica por la teoría del espacio-tiempo y por la teoría de la materia. No obstante, hay fenómenos físicos que se entienden más perspicazmente y se explican mejor utilizando un conjunto de conceptos que son distintos a los de la teoría del espacio-tiempo y a los de la teoría cuántica. Un ejemplo señalado de esta situación es la termodinámica. Aun cuando los fenómenos que se corresponden con la termodinámica, en esencia, no son más que los movimientos de la materia en el espacio-tiempo, cierto tipo de intuición, comprensión o explicación se basa en su análisis mediante las herramientas conceptuales de la mecánica estadística. Las mismas herramientas arrojan luz en el aspecto de la probabilidad en la física, en la explicación de patrones conductuales estadísticos y en la aparente irreversibilidad o asimetría-temporal de muchos fenómenos. Nuestra investigación de la relación entre la termodinámica, la entropía, la mecánica estadística y la irreversibilidad proporciona un ejemplo de cómo es posible desarrollar nuevas ideas sobre los fenómenos físicos aun cuando se conozcan ya tanto su ontología fundamental como sus leyes.

Estos dos volúmenes presentan un panorama muy personal. El tema es demasiado amplio y el espacio demasiado angosto como para hacerle justicia a todas las teorías físicas y posiciones filosóficas que se han propuesto en torno a estos asuntos, y yo no pretendo intentarlo. Más bien, pongo a consideración un conjunto circunscrito de planteamientos disyuntivos que me parecen a la vez claros e ilustrativos. Y apoyo sin ambages los que me parecen los más promisorios y bien fundados. No es ésta una visión de conjunto imparcial. Pero espero que mi selección de propuestas ilustre qué requiere una teoría física para ser clara y comprensible. Desafortunadamente, la física se ha contaminado con muy bajos estándares de claridad y precisión respecto a cuestiones fundamentales, y los físicos se han acostumbrado (y se les ha instado) a "cerrar la boca y hacer cálculos", a abstenerse de exigir una explicación clara del alcance ontológico de las teorías de la física. Esta actitud ha sido tan preponderante que acaso ya hemos olvidado cómo debería ser una explicación clara y concisa de la realidad física. De manera que si a usted no le atraen las teorías de la física que voy a discutir (y que no serán del agrado de muchos físicos), espero que al menos aprecie su *inteligibilidad*. Estas teorías podrán ser correctas o erróneas, perspicaces o desatinadas, pero será evidente lo que proponen respecto al mundo físico. Los físicos y los filósofos debemos exigir tal claridad si alguna vez hemos de entender el universo que habitamos.

I. EXPLICACIONES CLÁSICAS DEL ESPACIO Y DEL TIEMPO

EL NACIMIENTO DE LA FÍSICA

La tradición intelectual que produjo la física teórica moderna comienza en la antigua Grecia. Los astrónomos babilonios y egipcios habían compilado voluminosos y precisos datos sobre las posiciones visibles del Sol y los planetas, y además habían ideado modelos matemáticos que podían predecir fenómenos del tipo de los eclipses. Pero los filósofos griegos introdujeron una nueva línea de teorización especulativa en esa empresa observacional. Tales, Anaxágoras y Demócrito, por ejemplo, propusieron sus conjeturas sobre la estructura última de la materia: que todos los objetos materiales se derivan del agua, que son amalgamas de tierra, aire, fuego y agua, o que se componen de una infinita variedad de átomos de diferentes formas. El comportamiento observable de los objetos familiares se explicó luego sobre la base de esta constitución material. Según Demócrito, las cosas dulces se componen de átomos lisos y redondos; las cosas amargas, de átomos angulares, etc. La idea de que las propiedades y el comportamiento perceptibles de los objetos grandes deberían poder explicarse por la estructura y la naturaleza de sus partes imperceptibles subyace a la física hasta estos días.

Aristóteles acuñó el nombre de esta empresa especulativa. El término “física” se deriva del texto aristotélico Φυσική ἀκρόασις, *physikḗ akróasis*, “discursos sobre la naturaleza”, conocido en español con el título de *Física*. En la lengua griega, *phýsis* se refiere a la naturaleza de una cosa, y Aristóteles definió la naturaleza de un objeto como una fuente interna de movilidad e inmovilidad que le pertenece de manera primaria y propia, no contingente (*Física* II, 192b20-23). Por lo tanto, para Aristóteles la naturaleza de un objeto se revela por cómo se mueve y cómo deja de moverse cuando se le deja completamente solo. Si se deja caer una piedra sin arrojarla hacia ningún lado, ésta empieza a moverse hacia abajo, al parecer por su propia volición. Una burbuja de aire en un recipiente de agua se alza espontáneamente. A la piedra y al aire se les puede obligar a hacer otras cosas, pero sólo mediante la coerción de un agente externo. Sus innatas propensiones hacia la movilidad y el descanso no son atribuibles a agentes externos y por ende deben surgir de la propia naturaleza del objeto mismo.

La definición aristotélica de la “naturaleza” de un objeto no liga a la física de buenas a primeras con el proyecto de explicar la dulzura o la amargura de una cosa. En realidad, el énfasis se hace en el cambio en general y en la locomoción en particular. Aristóteles creía que los tipos diferentes de materia poseen movimientos naturales diferentes, de manera que con el fin de describir esos movimientos naturales tuvo que encontrar una forma de describir y categorizar el movimiento en general, empezando con las descripciones más directas e intuitivas. El movimiento natural del elemento tierra consiste en caer, es decir, moverse hacia abajo. El agua también se esfuerza en moverse hacia abajo, aunque con menos iniciativa que la tierra: una piedra se hunde a través del agua, demostrando así su irresistible y natural tendencia a descender. El fuego asciende naturalmente, como cualquiera que haya observado una fogata puede aseverar; también el aire se eleva, aunque con menos brío.

Está muy bien decir que una piedra cae o se mueve hacia abajo naturalmente, ¿pero qué significa exactamente "hacia abajo"? Es aquí donde Aristóteles se aparta de la opinión común y corriente y comienza un planteamiento teórico. Moverse hacia abajo, según él, significa moverse *hacia un lugar específico*. El movimiento natural de la tierra, según esta perspectiva, tiende a alcanzar una meta: la piedra pretende llegar a un sitio específico, y su movilidad espontánea siempre la lleva más cerca de este objetivo último. El lugar especial al que la piedra se esfuerza por llegar, según Aristóteles, es el centro del universo. Aristóteles pensaba que el cosmos material, en su totalidad, formaba una esfera en cuya superficie externa se hallaban las estrellas fijas. La esfera celeste tiene un solo centro. La dirección "hacia abajo" en cualquier sitio del universo es la dirección hacia ese punto central, y un pedazo de tierra que naturalmente se mueve sin obstrucción habrá de moverse hacia abajo en línea recta, hacia el centro, hasta llegar a la meta. Si logra llegar hasta abajo totalmente, el pedazo de tierra dejará de moverse por propia volición.

De manera similar, "arriba" es la dirección en el espacio que se aleja del centro. El fuego y el aire naturalmente se mueven hacia arriba en línea recta hasta donde puedan hacerlo, desplazando el fuego al aire cuando compiten entre sí. Según Aristóteles, si la esfera sublunar (la parte del universo debajo de la órbita de la Luna) no sufriera ningún tipo de perturbaciones, toda la tierra, aire, fuego y agua de forma natural se segregarían como cuatro esferas concéntricas: en el centro la tierra pura, sucesivamente rodeada por capas esféricas concéntricas de agua, aire y fuego, en este orden. Esta descripción proporciona un esbozo muy tosco del mundo tal como Aristóteles pensaba que era: una rocosa tierra esférica cubierta en gran parte por océanos y rodeada de aire.

La Luna y el Sol y los planetas no caben en este esquema, así que Aristóteles hubo de inventar un quinto tipo de sustancia, una quintaesencia, el llamado αἰθήρ, *aithḗr*, "cielo, firmamento". A diferencia de la tierra, el aire, el fuego y el agua, el éter no se mueve naturalmente en línea recta hacia un blanco: su movilidad natural consiste en un movimiento circular uniforme en torno a un centro. Este movimiento se realiza perfectamente por la esfera de estrellas fijas que gira (hasta donde Aristóteles podía saberlo) con perfecta regularidad, dando una rotación completa en cerca de 23 horas y 56 minutos (un *día sideral)*. Los demás objetos superlunarios —la Luna y el Sol y los planetas— no tienen la misma regularidad: al mismo tiempo que son acarreados por la esfera de las estrellas fijas, también ejecutan sus propios y más complicados movimientos periódicos. Al identificar el movimiento circular uniforme como el estado natural del éter, Aristóteles dejó un problema sin resolver para los astrónomos de las sucesivas generaciones: explicar el aparente movimiento del Sol, la Luna y los planetas como el efecto conjugado de diferentes movimientos circulares uniformes. Esta coerción fundamental en la teoría astronómica permaneció vigente hasta que Kepler hubo de proponer sus dos primeras leyes del movimiento planetario en 1609.

Desafortunadamente, incluso un esbozo ridículamente inadecuado de la historia de la física y la astronomía se encuentra más allá de nuestro alcance y propósito en este volumen. Pero la innovación aristotélica, su enfoque en la locomoción natural como el objeto primario de la física, sigue imperando en el campo actualmente. Nuestra primera orden del día consiste en entender con exactitud el significado de la "locomoción".

El término *locomoción* muestra a las claras su sentido: no se trata de un cambio cualquiera, sino de un cambio de lugar (*locus)*. Y un lugar, para Aristóteles, es una ubicación en un universo espacial que tiene una forma muy especial: la de una esfera. Puesto que es una esfera, el universo aristotélico contiene un centro geométrico esencial al cual Aristóteles hace referencia para caracterizar los movimientos naturales de los diferentes tipos de materia. "Hacia arriba", "hacia abajo" y "movimiento circular uniforme" se definen todos con base en el centro del universo. Si el universo de Aristóteles no hubiera tenido una forma circular, él no hubiera podido formular su física.

Es imposible exagerar la importancia de la explicación del espacio en la física. Si la física estudia en primer lugar el movimiento y si el movimiento consiste en un cambio de lugar, entonces (al parecer) debe haber *lugares* que los objetos materiales pueden ocupar sucesivamente. Un objeto descansa cuando ocupa el mismo lugar durante un tiempo, como la piedra aristotélica en el centro del universo. Se podría decir que si no hubiera algún tipo de espacio donde las cosas se mueven, la física ni siquiera podría haber despegado. Aristóteles adoptó el concepto del espacio, así como el correlativo concepto del movimiento, que todos intuitivamente utilizamos. Él se dio cuenta de que su física necesitaba tal espacio para poseer un cierto tipo de estructura peculiar —una meta, un blanco que los objetos al caer buscan—, y por lo tanto propuso una geometría física que proporcionaba esa estructura. El universo esférico y finito resultante hoy nos parece extraño, pero a cualquier griego de la Antigüedad le habría resultado muy familiar.

En suma, el espacio es el escenario del movimiento y por lo tanto la explicación del espacio tiene que ocupar un rol central en cualquier teoría científica del movimiento. Abandonar el universo esférico de Aristóteles significaba descartar sus principios físicos básicos y repensar la forma que pueden asumir las leyes de la física. Esta tarea fue emprendida por Isaac Newton.

La primera ley de Newton y el espacio absoluto

Si pretendiéramos axiomatizar la física de Aristóteles, utilizaríamos diferentes axiomas para los diferentes tipos de materia. "La Tierra, si algo no lo impide, se mueve en línea recta hacia el centro del universo" y "el éter, si algo no lo impide, se mueve con un movimiento circular uniforme en torno al centro del universo". Newton expuso su física como un conjunto de axiomas que él calificó de leyes del movimiento. Estas leyes contienen un tremendo caudal teórico, y no se exagera demasiado al decir que todo lo que necesitamos saber sobre la física newtoniana está implícito en la primera ley del movimiento:

> Primera ley: Todos los cuerpos se mantienen en su estado de descanso o de movimiento uniforme en línea recta, salvo que se les obligue a cambiar su estado mediante fuerzas impresas.

Por sí sola, esta ley hace añicos el universo aristotélico.

Primero, la ley de Newton gobierna a *todos* los cuerpos: tanto a las piedras como a los planetas. Newton oblitera la distinción entre la astronomía y la física terrestre, proponiendo un solo conjunto de principios que explican el comportamiento de ambas. Estamos tan acostumbrados a pensar que la física posee este tipo de universalidad, que tenemos que esforzarnos en apreciar la tremenda importancia de este cambio. Uno de los pasajes más impresionantes en la estructura argumentativa de los *Principia* es la demostración que hace Newton de que la fuerza que mantiene a la Luna en órbita alrededor de la Tierra es precisamente la misma fuerza que hace que una manzana caiga de la rama de un árbol. Newton plantea la existencia de una estructura física común, allí donde la tradición precedente había visto una diversidad fundamental.

Es todavía más notable el hecho de que Newton no le haya adscrito una movilidad natural peculiar a los cuerpos, como Aristóteles lo había hecho. En cambio, la ley de la inercia le atribuye a todos los cuerpos una tendencia innata que los *mantiene* en su estado de movimiento, sea cual fuere. No hay ningún lugar en el universo "hacia el cual se dirija" inherentemente cualquier cuerpo, como una piedra se dirige hacia el centro del universo en la teoría de Aristóteles. La teoría newtoniana no requiere que el espacio posea un punto central especial.

Para Newton el escenario del movimiento es más bien un ente que él denomina *espacio absoluto*. El movimiento según Newton es un cambio de ubicación en este espacio. El rol del espacio absoluto es tan profundo y ubicuo en la teoría newtoniana que probablemente sería imposible entenderla si no se acepta su existencia. Examinaremos varias de las propiedades del espacio absoluto, dejando las más controversiales para el final.

Primero, Newton asume que el espacio absoluto posee la estructura geométrica del espacio euclidiano tridimensional. Llamaremos E^3 a esta estructura. A diferencia del universo físico aristotélico, E^3 es infinita en todas las direcciones y por lo tanto no tiene un centro geométrico. Las figuras en E^3 se gobiernan por los axiomas de la geometría euclidiana: por ejemplo, la suma de los ángulos interiores de cualquier triángulo es igual a dos ángulos rectos.

Será útil para la subsecuente discusión distinguir entre los diferentes tipos de estructura geométrica, los cuales forman una jerarquía. Cada nivel se corresponde con uno de los tres instrumentos utilizados en la geometría euclidiana: el lápiz, la regla, y el compás. ¿Qué tipo de estructura geométrica debe tener un espacio con el fin de que cada uno de estos instrumentos pueda funcionar en él?

El nivel más básico, fundamental de una estructura geométrica en un espacio se describe como su *topología*. La topología de un espacio determina las características de la *continuidad*. Por ejemplo, cuando usamos un lápiz para hacer un diagrama euclidiano, se supone que debemos dibujar líneas *continuas* en un espacio: si de vez en cuando alzamos el lápiz para luego posarlo nuevamente en otro punto al dibujar lo que debía ser una sola línea, no obtendremos una línea continua, conectada, única. Pero para que sea posible distinguir entre una sola línea en un espacio y un par de líneas inconexas, los puntos en el espacio deben tener algún tipo de organización geométrica. Este nivel de organización es la topología del espacio.

A la topología a veces se le llama “geometría de hoja de hule”, y esta denominación es adecuadamente evocadora. Supongamos que se dibujen algunas figuras en una superficie de hule y que luego ésta se estire sin rasgarse o sin pegarse. Algunas de las propiedades de las figuras cambiarán al deformarse la superficie: las líneas rectas se doblarán, haciéndose curvas; los puntos cercanos podrán alejarse unos de otros en virtud del estiramiento; un triángulo se podrá deformar suavemente, convirtiéndose en un círculo, etc. Pero algunas de las características de las figuras no cambiarán: la intersección de dos líneas antes de la deformación es la misma después; si un punto se encuentra en el interior de una figura cerrada y otro en el exterior antes de la deformación, permanecerán así después, etc. No se permite que las deformaciones “rasguen” o “peguen” la superficie del espacio, y la topología proporciona el nivel de estructura geométrica que define lo que sería “rasgar” y “pegar”. El efecto de rasgar separa algunas de las líneas continuas, haciéndolas discontinuas, y el efecto de pegar une líneas discontinuas, haciéndolas continuas. Si el espacio careciera de una topología, entonces no se podría distinguir entre la acción de dibujar una sola curva continua y la acción de dibujar varias curvas inconexas, por lo que las construcciones euclidianas ni siquiera podrían hacerse.[1]

El segundo instrumento de la geometría euclidiana es la regla recta (no la regla común; la regla recta no tiene escala de medición). Con una regla recta y un lápiz, no sólo podemos dibujar líneas continuas sino también líneas rectas. Los dos primeros axiomas de la geometría euclidiana involucran la utilización de la regla recta y por lo tanto se refieren implícitamente a la estructura de las líneas rectas en el espacio euclidiano. En particular, estas primeras dos premisas afirman:

1) Es posible dibujar una línea recta desde cualquier punto a cualquier otro punto.

2) Es posible extender continuamente en la forma de una línea recta cualquier línea recta finita.

Para que estos postulados sean aceptables, primero debe hacerse una distinción en el espacio entre las líneas rectas y las demás líneas. Esta distinción, que no se determina por la topología, se deriva por la *estructura afín* del espacio. En el espacio euclidiano, la estructura afín determina que todo par de puntos está formado por los puntos en los extremos de una sola línea recta y que toda línea recta finita puede hacerse continuar indefinidamente en cualquiera de las dos direcciones. Es posible describir los espacios que no tienen este tipo de estructura afín: un par de puntos podría no determinar una línea recta o podría determinar más de una, o podría tener límites la extensión continua de una línea recta. Por ende, en los dos primeros axiomas de Euclides —los cuales describen las posibles utilizaciones de una regla recta— ya se restringe la estructura afín del espacio que él describe.

La estructura afín de un espacio no determina las características de la *longitud* de las líneas o de la *distancia* entre los puntos. Para esto se requiere todavía otra categoría de forma geométrica, la llamada *estructura métrica* del espacio. El compás indica la estructura métrica de un espacio: un círculo es el lugar geométrico (*locus)* de puntos que se encuentran todos equidistantes de un centro dado. La tercera premisa de Euclides afirma que un círculo completo, continuo y cerrado puede dibujarse con cualquier centro y radio dados. También en este caso podríamos imaginar espacios en que tal afirmación no es válida.

Es posible ilustrar la forma jerárquica de estos tres niveles de estructura mediante tres diferentes tipos de transformaciones que pueden realizarse en las figuras del espacio euclidiano. Una *transformación topológica* convierte líneas continuas en líneas continuas. Una *transformación afín* debe asignar líneas rectas a las líneas rectas. Un "estiramiento uniforme" del espacio puede calificarse de transformación afín aun cuando modifique las distancias entre los puntos y deforme los círculos haciéndolos elipses. Una *isometría* es una asignación de un espacio a sí mismo que conserva las distancias, de manera que los círculos se transforman en círculos.[2] La figura I.1 ilustra los tres tipos de mapeos.

Todas las isometrías son transformaciones afines y todas las transformaciones afines son transformaciones topológicas, pero no inversamente.

La geometría moderna ha introducido otro nivel estructural. Es la *estructura diferenciable*, que distingue entre las curvas lisas continuas y las curvas con esquinas o recodos tajantes. A una asignación que conserva la estructura diferenciable se le denomina *difeomorfismo* y mapea curvas lisas en curvas lisas. Mientras que una transformación topológica puede asignar un triángulo en un círculo, un difeomorfismo no puede hacerlo, ya que un círculo es liso y un triángulo esquinado. La transformación topológica que muestra la figura I.1 es un difeomorfismo: obsérvese que las tres esquinas del triángulo todavía son reconocibles.

En casi todas las discusiones de la geometría euclidiana, el quinto axioma suele acaparar la atención. Esta premisa se relaciona con la existencia y las propiedades de las líneas paralelas. El descubrimiento original de las geometrías no euclidianas surgió de los intentos de demostrar el quinto axioma con base en los otros cuatro. A la larga se demostró que tanto el quinto axioma como la negación de su veracidad son consistentes con los demás axiomas de Euclides, de modo que puede haber espacios en los que las reglas rectas y los compases se utilicen tal como Euclides lo requiere, pero en los cuales las figuras geométricas no poseen las propiedades que Euclides describe. Por ejemplo, en algunos espacios no euclidianos la suma de los ángulos internos de un triángulo equivale a más de dos ángulos rectos, y en otros espacios a menos. El quinto axioma no juega un rol esencial en la formulación de la física de Newton. La mecánica newtoniana podría funcionar en un espacio que no contenga líneas paralelas. Sin embargo, la existencia de una estructura afín y una estructura métrica es absolutamente esencial para la comprensión de las leyes de Newton. Pero antes de que podamos adentrarnos en esas leyes, necesitamos hablar del tiempo.

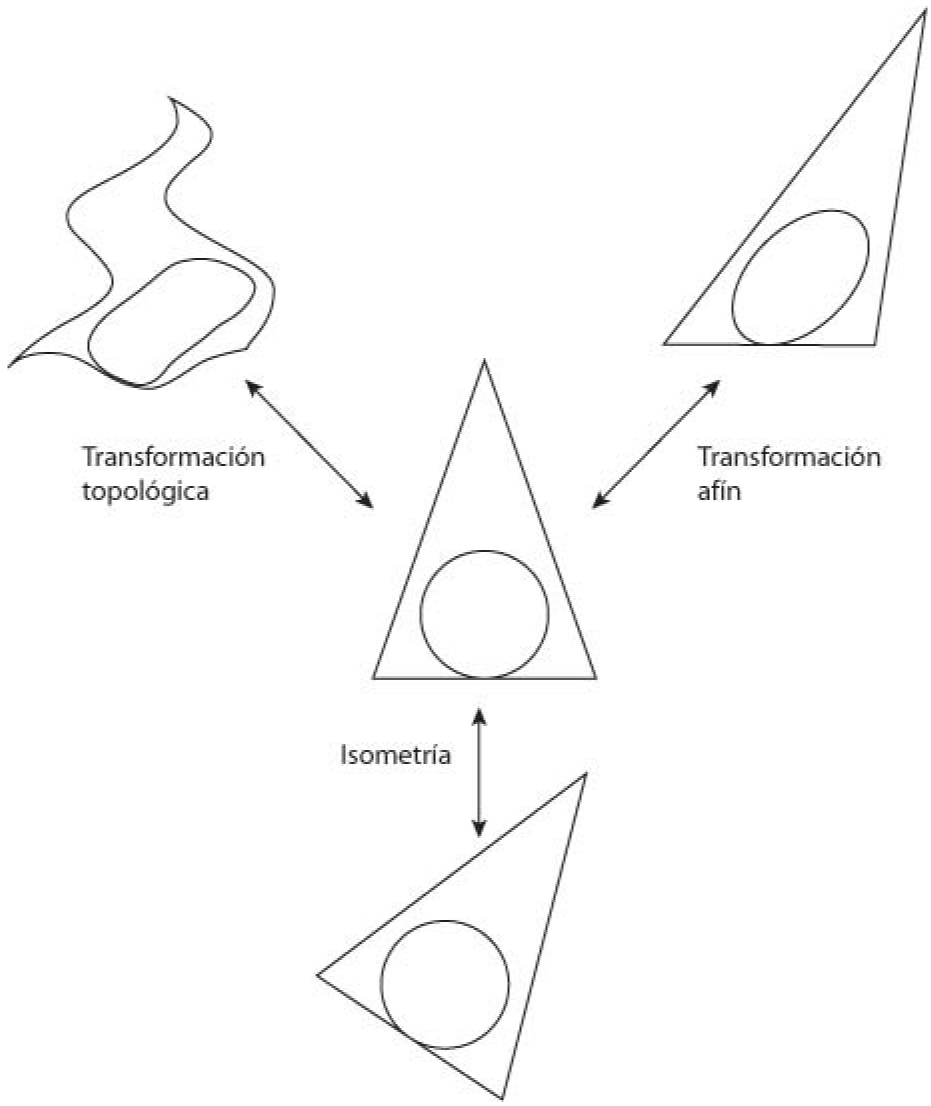

FIGURA I.1

EL TIEMPO ABSOLUTO Y LA PERSISTENCIA DEL ESPACIO ABSOLUTO

Newton creía en la existencia de un escenario espacial cuya estructura geométrica fuera la de E^3. Pensaba que este espacio tridimensional infinito existe en todo momento del tiempo. Y también creía algo mucho más sutil y controversial, a saber, que *exactamente los mismos puntos* del espacio persisten a través del tiempo.

Estamos tratando de entender qué es lo que hay que postular con el fin de que la primera ley tenga sentido, y la primera ley asevera que un cuerpo al que no se le aplique una fuerza permanece en reposo si se encuentra en reposo, y continúa moviéndose uniformemente en línea recta si está en movimiento. ¿Pero qué significa que un cuerpo *repose* o que *permanezca en reposo?* Si es verdad que los puntos individuales del espacio persisten a través del tiempo, entonces tenemos allí una explicación precisa: un cuerpo se encuentra en reposo absoluto cuando ocupa los mismos puntos del espacio absoluto a lo largo de un periodo de tiempo. La explicación del movimiento uniforme rectilíneo es similar, pero más compleja. Primero: si los puntos del espacio absoluto persisten a través del tiempo, entonces todo cuerpo en movimiento viaja a lo largo de una trayectoria en el espacio absoluto; es decir, a lo largo del conjunto de puntos del espacio absoluto que ocupa durante un periodo de tiempo dado. Y si el espacio absoluto tiene una estructura afín, entonces esta trayectoria espacial o forma una línea recta en el espacio, o no la forma. De manera que para que tenga sentido el "movimiento uniforme rectilíneo", los puntos en el espacio no sólo deben persistir a través del tiempo y tener una topología (para que tenga sentido caracterizar la trayectoria de un cuerpo como una línea continua), sino también una estructura afín (para que la trayectoria espacial pueda calificarse de recta o de curva).

No obstante, estas condiciones por sí mismas no definen el "movimiento uniforme rectilíneo", ya que no se ha explicado lo que significa "uniforme". Un automóvil en una carrera de aceleración, a diferencia de una de Fórmula Uno, compite en una pista recta, por lo que su movimiento es "rectilíneo". Aun así, el movimiento no es *uniforme:* el automóvil acelera incesantemente moviéndose con creciente velocidad. A este tipo de movimiento se le llama *aceleración lineal*. (En el SLAC National Accelerator Laboratory hay un tubo con una longitud de aproximadamente tres kilómetros donde se aceleran las partículas, a diferencia del Gran Colisionador de Hadrones, donde se aceleran las partículas en un tubo circular cerrado.) Para que un movimiento sea *uniforme*, deberá cubrir la misma distancia en el mismo tiempo. Así, la primera ley de Newton presupone que se sabe *cuál es la distancia* que un cuerpo recorre y *cuánto tiempo* le toma cubrirla. El primer hecho requiere una métrica del espacio para que sea posible asignarle una longitud a la trayectoria espacial de un cuerpo. Y el segundo hecho requiere algo totalmente nuevo: una estructura métrica del tiempo. La primera ley del movimiento de Newton no sólo presupone el espacio absoluto sino también el tiempo absoluto.

La estructura geométrica del tiempo, según Newton (y el sentido común), es más sencilla que la del espacio. El tiempo newtoniano es unidimensional: una sola secuencia ordenada de instantes que forma la totalidad de la historia. Ese conjunto de instantes posee una topología que se determina por su orden en el tiempo. No tiene sentido realmente preguntar si esta "línea del tiempo" es recta o curva, por lo que la noción de una estructura afín no se plantea. Pero hay una métrica temporal: entre cualesquiera dos instantes transcurre una cierta cantidad de tiempo absoluto, y esta cantidad se puede comparar con cualquiera otra en términos de su dimensión. Si una cierta cantidad de tiempo transcurre entre el instante 1 y el instante 2, y una cierta cantidad de tiempo transcurre entre el instante 2 y el instante 3, existe un hecho respecto a si tales intervalos tienen la misma dimensión o no, y un hecho respecto a la proporción exacta entre los intervalos.

Sobre la base de toda esta estructura, podemos definir ahora el “movimiento uniforme”: es un movimiento que recorre la misma cantidad de espacio en el mismo tiempo. Un movimiento uniforme no tiene que ser rectilíneo: el movimiento circular uniforme, por ejemplo, puede mantener una velocidad constante aun cuando cambie de dirección continuamente. Así que Newton requería todas estas características para que la primera ley pudiera enunciarse con absoluta precisión: si no se le aplica una fuerza a un objeto cuando se encuentra en reposo absoluto, el objeto habrá de permanecer en reposo absoluto, y si se encuentra en movimiento, habrá de seguir moviéndose en una trayectoria espacial rectilínea, recorriendo distancias iguales en tiempos iguales.

Para finalizar, hay otro aspecto que Newton le asigna al tiempo absoluto: éste, a diferencia del espacio, tiene una *dirección*. Newton no dice nada explícito al respecto y el asunto no es directamente pertinente en la comprensión de sus leyes del movimiento. Pero a todos nos resulta perfectamente natural decir que el tiempo transcurre del pasado al futuro, lo cual vale la pena señalar aquí porque habremos de regresar más adelante a las cuestiones en torno a la dirección del tiempo.

De hecho, Newton no discute en absoluto la estructura geométrica del espacio o del tiempo. Él siempre utiliza E^3 para describir el espacio y siempre presupone en sus pruebas que existe una métrica definitiva para el transcurso del tiempo. No se le hubiera ocurrido que podía haber una alternativa al respecto. En el análisis precedente hemos visto cuáles son los factores necesarios para poder enunciar la primera ley: el espacio debe tener una topología, una estructura afín y una métrica; el tiempo debe ser unidimensional y tener una topología y una métrica; y, sobre todo, las partes individuales del espacio deben persistir a través del tiempo. Dado todo esto, hay un hecho respecto a si un cuerpo permanece o no en la misma región del espacio a través del tiempo, un hecho sobre la trayectoria espacial de un objeto en movimiento y un hecho respecto a la velocidad con que un cuerpo en movimiento recorre diferentes partes de esa trayectoria. Sin el fundamento de toda esta estructura, no se podría entender con claridad la primera ley de Newton. Pero si negáramos que el espacio sea E^3, asignándole otro tipo de estructura afín y de métrica, la ley seguiría teniendo perfecto sentido.

La metafísica del espacio y el tiempo absolutos

Aunque Newton no describa explícitamente la estructura geométrica que él le asigna al espacio y al tiempo absolutos, sí discute muy claramente su estatus metafísico, tanto como las razones por las que debemos aceptar su existencia. La cuestión básica es obvia. Tomemos, por ejemplo, un objeto en reposo absoluto en algún momento del tiempo y que no está sujeto a ningún tipo de fuerza externa. Según la primera ley, el objeto habrá de permanecer en reposo absoluto, es decir, permanecerá ubicado en el mismo lugar del espacio absoluto. Pero Newton sabe muy bien que estas partes persistentes del espacio absoluto no pueden percibirse por los sentidos. Ningún tipo de observación puede revelar si un cuerpo permanece en la misma región del espacio absoluto o si constantemente se mueve de una parte a otra. Parecería, pues, que aunque la primera ley de Newton sea verdadera, y aunque pudiéramos comprobar que a un objeto no se le aplicaran fuerzas externas, ningún tipo de observación podría verificar la ley. Y lo que es más grave, si no podemos percibir el espacio absoluto y *a fortiori* no podemos percibir el movimiento absoluto, no resulta evidente cómo una teoría que explica tal movimiento absoluto podría hacer cualquier tipo de predicción respecto a los hechos observables.

Aquello que sí podemos observar, dice Newton, son las posiciones *relativas* de los cuerpos entre sí. De forma similar, no podemos observar directamente el transcurso del tiempo absoluto, pero sí podemos observar los cambios en las posiciones *relativas* de los cuerpos. En el "Escolio" que aparece tras sus definiciones de los nuevos términos, Newton cuidadosamente hace la distinción entre las cantidades observables y los entes absolutos que él propone:

> Nos ha parecido oportuno explicar hasta aquí los términos menos conocidos y el sentido en que se han de tomar en el futuro. En cuanto al tiempo, espacio, lugar y movimiento, son de sobra conocidos para todos. Hay que señalar, sin embargo, que el vulgo no concibe estas magnitudes si no es con respecto a lo sensible. De ello se originan ciertos prejuicios para cuya destrucción conviene que las distingamos en absolutas y relativas, verdaderas y aparentes, matemáticas y vulgares.
>
> I. El tiempo absoluto, verdadero y matemático, en sí mismo y por su naturaleza, fluye uniformemente sin relación con nada externo, y se le llama, con otro nombre, duración: el relativo, aparente y vulgar es cualquier medida (exacta o imprecisa) de la duración, realizada sensible y externamente por medio del movimiento, la cual es usada vulgarmente en vez del tiempo verdadero: como la hora, el día, el mes, el año.
>
> II. El espacio absoluto, por su naturaleza, sin relación con nada externo, siempre permanece igual e inmóvil; el relativo es cualquier cantidad o dimensión variable de este espacio, que se define por nuestros sentidos según su situación respecto a los cuerpos, espacio que el vulgo toma por el espacio inmóvil: así, una extensión espacial subterránea, aérea o celeste definida por su situación relativa a la Tierra. El espacio absoluto y el relativo son el mismo en especie y en magnitud, pero no permanecen siempre el mismo numéricamente. Pues si la Tierra, por ejemplo, se mueve, el espacio de nuestra atmósfera que relativamente y respecto a la Tierra siempre permanece el mismo, ahora será una parte del espacio absoluto por la que pasa el aire, después otra parte, y así, desde un punto de vista absoluto, siempre cambiará.
>
> III. El lugar es la parte del espacio que un cuerpo ocupa y es, en tanto que espacio, absoluto o relativo. […]
>
> IV. El movimiento absoluto es el paso de un cuerpo de un lugar absoluto a otro lugar absoluto; el movimiento relativo es el paso de un lugar relativo a otro lugar relativo. Así, en una nave empujada por las velas desplegadas, el lugar relativo de un cuerpo es aquella región de la nave en que está el cuerpo, o sea, la parte de la cavidad total que llena dicho cuerpo y que, por consiguiente, se mueve a la vez que la nave: mientras que el reposo relativo es la permanencia del cuerpo en la misma región de la nave o en la misma parte de su cavidad. Pero el reposo verdadero es la permanencia del cuerpo en la misma parte del espacio inmóvil en que se mueve la nave misma junto con su cavidad y todos sus contenidos. De donde si la Tierra verdaderamente está en reposo, el cuerpo que en la nave permanece relativamente en reposo se moverá verdadera y absolutamente con la misma velocidad con que la nave se mueve sobre la Tierra. Si la Tierra también se mueve, constará el verdadero y absoluto movimiento del cuerpo, parte del verdadero movimiento de la Tierra en el espacio inmóvil, parte de los movimientos relativos de la nave sobre la Tierra […][3]

Newton hace la distinción entre las nociones "absolutas, verdaderas y matemáticas" de espacio, tiempo, lugar y movimiento, y las nociones "relativas, aparentes y comunes".

El meollo del problema es que las leyes del movimiento de Newton se expresan en términos de nociones absolutas que nosotros no podemos observar directamente. Cuando intentamos observar el movimiento de un objeto, sólo podemos ver directamente su movimiento relativo: el cambio de su posición en relación con otros objetos visibles, midiéndose el cambio por el movimiento visible de los relojes u otros instrumentos que se utilizan para medir el tiempo. Pero si el movimiento absoluto de un objeto es imperceptible porque el espacio y el tiempo son imperceptibles, ¿cómo puede ser pertinente para la ciencia empírica la postulación de tales entes?

Newton dedica el resto del "Escolio" a elucidar esta cuestión:

> Mas como estas partes del espacio no pueden verse y distinguirse unas de otras por medio de nuestros sentidos, en su lugar utilizamos medidas sensibles. Por las posiciones y distancias de las cosas a un cierto cuerpo que consideramos inmóvil, definimos todos los lugares; posteriormente interpretamos todos los movimientos por respecto a los antedichos lugares, en tanto que los concebimos como pasos de los cuerpos por estos lugares. Así, usamos de los lugares y movimientos relativos en lugar de los absolutos y con toda tranquilidad en las cosas humanas: para la Filosofía, en cambio, es preciso abstraer de los sentidos. Pues es posible que en la realidad no exista ningún cuerpo que esté en total reposo, al que referir lugar y movimiento.
>
> Se distinguen el reposo y movimiento absolutos y relativos entre sí por sus propiedades, causas y efectos.[4]

Newton proporciona la evidencia *empírica* convincente de la existencia del movimiento absoluto (y por lo tanto del espacio y tiempo absolutos) utilizando consideraciones sobre las causas del movimiento. Para aquilatar este argumento, tenemos que examinar su segunda ley.

Pero antes de pasar a la segunda ley debemos hacer una pausa para reflexionar hasta qué grado es profundamente intuitiva la explicación newtoniana del espacio y tiempo absolutos, aun cuando el espacio y tiempo absolutos no sean directamente observables. Podría parecer que Newton propone entes extraños y fantasmales, pero la mayoría de la gente concibe el mundo físico en términos del espacio y tiempo absolutos. Por ejemplo, los artesanos y los científicos siempre están tratando de mejorar el diseño de los relojes, produciendo modelos cada vez más exactos. ¿Pero en qué consiste la "exactitud" de un reloj? Lo que se pretende es que los sucesivos tictacs del reloj ocurran *en intervalos iguales de tiempo*, o que la segunda manecilla del reloj recorra la carátula *a una velocidad constante*. ¿Pero "igual" y "constante" con relación a qué? Con relación al transcurso del tiempo mismo, es decir, con relación al tiempo absoluto. Nuestra percepción natural, intuitiva, nos dice que una cierta cantidad de tiempo absoluto transcurre entre los sucesivos tictacs de un reloj, y mientras mayor sea la precisión del reloj, más similares serán esos intervalos entre sí. Los relojeros suizos y los diseñadores de relojes atómicos intentan que sus instrumentos midan con precisión una cosa, y esa cosa no consiste en ningún tipo de tiempo relativo, observable. Para la física, cualquier movimiento físico observable —la rotación de la Tierra, el movimiento de la Tierra alrededor del Sol, etc.— está sujeto a perturbaciones y no es por lo tanto automáticamente uniforme. Pero esta falta de uniformidad no se define con relación a algún tipo de movimiento observable. El diseño de los relojes refleja el intento de eliminar o compensar los factores perturbadores. Esta práctica adopta implícitamente algún tipo de medida del tiempo mismo que pueda proporcionar un estándar de la uniformidad.

De manera similar, nuestra cotidiana comprensión del mundo lo concibe en términos del espacio absoluto. A nadie le sorprende oír, por ejemplo, que la órbita de la Tierra es una elipse cuyo punto focal es el Sol. Las ilustraciones del sistema solar en los libros muestran las órbitas de los planetas. ¿Pero qué se supone que muestran? En todo momento dado, la Tierra se encuentra en algún lugar. La "órbita" es de alguna manera un conjunto de todos los lugares que la Tierra ocupa en el curso de un año. Pero esto implica que los lugares que la Tierra ocupa en diferentes momentos son todos parte de un espacio común, tridimensional: el espacio absoluto. Mediante un profundo esfuerzo intelectual podríamos llegar a comprender cómo el mundo podría existir sin un espacio absoluto y sin un movimiento absoluto. Pero si lo lográramos no sólo rechazaríamos la teoría de Newton, también rechazaríamos el sentido común.

Newton no invocó el sentido común para justificar su idea de que el espacio y el tiempo son absolutos: invocó la experimentación. Newton intentó probar la existencia del movimiento absoluto en el laboratorio más que mediante un análisis conceptual. Será éste nuestro próximo tema.

[1] En las matemáticas modernas, la topología de un espacio se especifica en términos de la estructura de *conjuntos abiertos* del espacio, y la continuidad se define en términos de los conjuntos abiertos. Creo yo que esta explicación de la continuidad y, por ende, de la topología no es la forma más perspicaz de describir la estructura geométrica intrínseca del espacio-tiempo, así que he desarrollado una forma alternativa (véase T. Maudlin, "The Geometry of Space-Time", *Aristotelian Society Supplementary*, 84 (1): 63-78, 2010, para una visión de conjunto). No es éste el lugar para entablar esa batalla.

[2] Mientras que una transformación afín sólo debe asignar líneas rectas a las líneas rectas, una isometría debe hacer más que asignar círculos a los círculos: la dimensión de los círculos también debe cambiar. Una transformación de escala que uniformemente achica o amplía todas las figuras no es una isometría, aun cuando asigne círculos a los círculos.

[3] I. Newton, *Principia*, vol. 1, University of California Press, Berkeley, 1934, pp. 6 y 7. [Versión en español tomada de *Principios matemáticos de la filosofía natural*, t. 1, trad. de E. Rada García, Alianza, Madrid, 2011, pp. 127-128.]

[4] *Id*. [Versión en español: *ibid*., pp. 129-130.]

II. LA EVIDENCIA DE LA ESTRUCTURA ESPACIAL Y TEMPORAL

LA SEGUNDA LEY DE NEWTON Y EL EXPERIMENTO DE LA CUBETA

La primera ley de Newton, la llamada ley de la inercia, refiere sólo a cuerpos que no estén sujetos a fuerzas externas. No hay que caer en la tentación de decir que Newton postula que tales cuerpos "continúan en el mismo estado de movimiento", puesto que dicha formulación pasaría por alto el aspecto revolucionario de esta ley; la primera ley especifica exactamente *lo que en realidad significa* "el mismo estado de movimiento". Como hemos visto, para Aristóteles una porción de éter en movimiento circular en torno al centro del universo se encuentra siempre en "el mismo estado de movimiento", así que no habría razón para buscar causas externas en semejante situación. En la física de Aristóteles, las causas externas son responsables del movimiento no natural, tal como una roca moviéndose hacia arriba y no hacia abajo. Así que para Aristóteles, la caída de una piedra que se encuentre inicialmente en un estado de reposo, aunque sin apoyo, no requiere de una causa externa, y la continua rotación uniforme de la esfera de estrellas fijas no requiere de causa externa: es eso lo que estas especies de materia hacen por naturaleza.

Es interesante observar que Galileo, antes que Newton, quiso especificar el movimiento "inercial" de los objetos terrestres —es decir, el movimiento que éstos mostrarían si no estuvieran sujetos a fuerza ninguna— concluyendo que tal movimiento habría de ser *uniforme y circular*. Llegó a esta conclusión con base en sus experimentos con planos inclinados. Galileo observó que si una bola rueda hacia abajo en un plano inclinado y luego hacia arriba en otro, la bola casi asciende hasta la altura desde donde inició su movimiento. Concluyó —correctamente, por cierto— que cuando no había fricción ni resistencia de aire la bola rodaría hacia arriba en el segundo plano inclinado hasta alcanzar exactamente la altura desde donde había salido (figura II.1). Cuando el ángulo θ va disminuyendo, la bola tiene que rodar cada vez más en la segunda pendiente antes de alcanzar la altura original desde la cual empezó su movimiento. Si el ángulo se fijara en 0, de manera que la bola rodara sin ascender ningún plano inclinado, continuaría rodando para siempre (en ausencia de fricción) a una velocidad constante.[1]

¿Qué tipo de movimiento es éste, un movimiento con una velocidad constante que jamás gana ni pierde altitud? ¡Es un movimiento *circular* uniforme en torno al centro de la Tierra! Esa bola rodaría alrededor de la Tierra eternamente, sin perder ni ganar velocidad nunca. Ciertamente, la bola no podría continuar rodando así en una línea recta: una línea recta, tangente a la superficie de la Tierra, deja la Tierra y se eleva. Una bola que intentara seguir una ruta recta iría perdiendo velocidad y finalmente se detendría, rodando hacia atrás hasta su punto de partida. Galileo concluyó que el movimiento inercial, natural, de todos los objetos era como el movimiento natural del éter de Aristóteles: un movimiento circular uniforme. Para Galileo, esta *inercia circular* ayudaba a explicar cómo, por ejemplo, las aves podían volar sin ser rebasadas por la rotación de la Tierra. No tenían que esforzarse mucho, aunque tuvieran que volar a cientos de kilómetros por hora para no quedar rezagadas: su inercia circular las llevaría consigo sin necesidad de ninguna fuerza adicional.

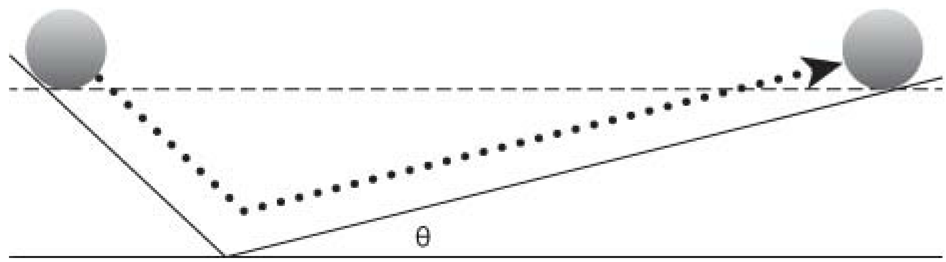

FIGURA II.1

La innovación de Newton no reside simplemente en la hipótesis de cierto movimiento interno que un cuerpo tiende a mantener en ausencia de una fuerza externa, sino en la aserción específica de que se trata de un movimiento uniforme en línea recta. Sin embargo, la primera ley es sin duda insuficiente para explicar el mundo que nos rodea: jamás vemos cuerpos que se mueven continua y uniformemente en línea recta. La primera ley tendría que complementarse con una explicación del efecto de las fuerzas externas en el movimiento de un cuerpo, y eso es exactamente lo que la segunda ley hace:

> Segunda ley: El cambio en el movimiento es proporcional a la fuerza motriz impresa y se realiza a lo largo de la línea recta en que la fuerza se imprime.

La segunda ley es conceptualmente parasítica respecto a la primera ley: sin la definición de un "estado de movimiento" en la primera ley, no podríamos entender el sentido del "cambio de movimiento" en la segunda ley. Según Aristóteles y Galileo, ciertos movimientos circulares uniformes son movimientos constantes: no muestran ningún cambio y por lo tanto no hay que explicarlos. Pero según Newton un cuerpo en movimiento circular uniforme constantemente cambia su estado de movimiento y por lo tanto debe estar sujeto a algún tipo de fuerza externa.

El concepto que Newton tiene de la "fuerza", tal como se utiliza en la segunda ley, no es exactamente el mismo que se usa en la física moderna. La noción newtoniana se corresponde más bien con lo que hoy se llamaría un "impulso", es decir, la acción de una fuerza a través de un cierto intervalo de tiempo. Para Newton, la misma "fuerza" puede aplicarse a un cuerpo cuando se ejerce con mayor ímpetu durante menos tiempo que cuando se ejerce con menor ímpetu durante más tiempo. Si se ejerce durante dos veces más tiempo pero con el mismo ímpetu, entonces el cuerpo queda sometido a dos veces más fuerza y su estado de movimiento cambiará dos veces.

Veamos cómo se relaciona la segunda ley de Newton con el movimiento circular uniforme. La figura II.2 indica la trayectoria —en el espacio absoluto— de una partícula en movimiento circular uniforme en sentido contrario a las manecillas del reloj en torno al punto *o*. Digamos que exactamente a mediodía la partícula se encuentra en el punto *a*. La velocidad es hacia la izquierda a lo largo de la línea tangente *T*. Según la primera ley del movimiento, si no se aplicara una fuerza en la partícula después de mediodía ésta continuaría a lo largo de *T* a una velocidad constante. De manera que para continuar en la ruta circular hacia *b*, una cierta fuerza debe aplicarse a la partícula para desviarla hacia abajo, hacia *o*. Además, si queremos que la partícula continúe su camino a una velocidad constante, sin ganar o perder velocidad, entonces ninguna porción de la fuerza debe aplicarse en la dirección de la velocidad actual (porque aceleraría a la partícula) o en la dirección contraria (porque desaceleraría a la partícula). De manera que según la segunda ley es necesario imprimir una fuerza constante con el fin de que la partícula se mantenga en movimiento circular uniforme, y esa fuerza siempre debe dirigirse exactamente hacia el centro del círculo. A esta fuerza se le denomina *centrípeta* (hacia el centro).

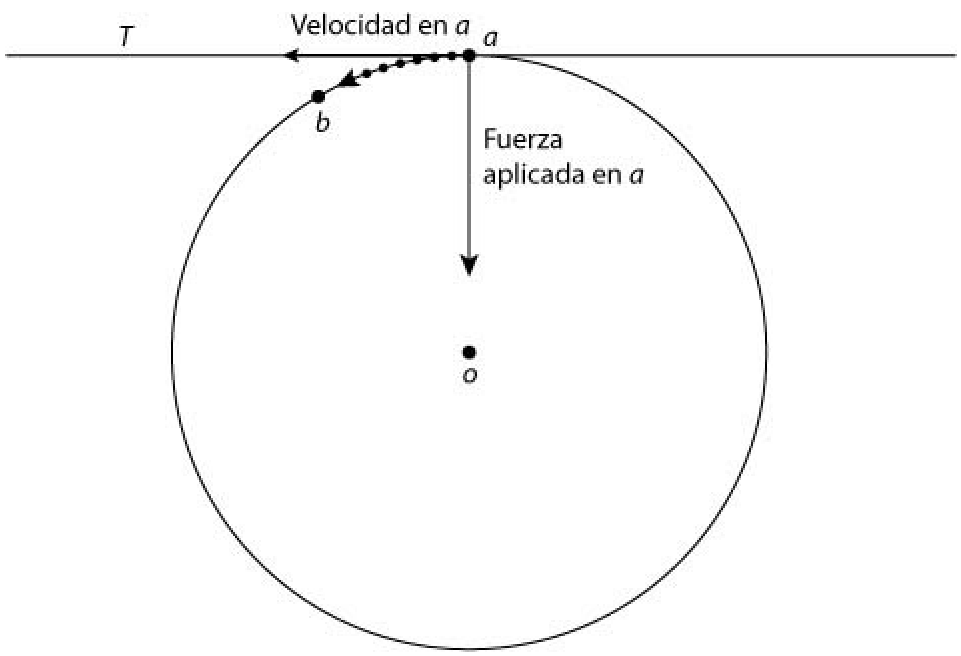

FIGURA II.2

Mientras que Aristóteles aseveraba que la Luna, estando compuesta de éter, naturalmente orbitaría en torno al centro del universo (y por lo tanto de la Tierra), la primera y la segunda ley del movimiento newtonianas implican que el comportamiento de la Luna requiere la existencia de una fuerza que se dirige hacia la Tierra y que impide que la Luna continúe en línea recta. Esa fuerza es la gravedad.

La asociación de las fuerzas con los cambios del movimiento absoluto proporciona la llave del argumento más fuerte de Newton en favor de los movimientos absolutos y, por extensión, en favor del espacio y el tiempo absolutos. Pues incluso si los movimientos absolutos de los cuerpos no son directamente observables, a veces sí lo son las *fuerzas* que se imprimen en esos cuerpos. Y donde se da una fuerza neta hay, según la segunda ley, un cambio en el movimiento absoluto.

Con el fin de demostrar esta conexión, Newton propone un experimento tan sencillo y familiar que no es necesario realizarlo en la práctica. Llene usted con agua una cubeta, cuélguela del techo con una cuerda, y enrosque la cuerda. Al soltarla, la cuerda empieza a desenrollarse, y el experimento avanza en cuatro etapas:

> […] la superficie del agua será plana al principio, al igual que antes del movimiento del recipiente, pero después, al transmitir éste su fuerza poco a poco al agua, hace que ésta también empiece a girar sensiblemente, se vaya apartando poco a poco del centro y ascienda hacia los bordes del recipiente, formando una figura cóncava (como yo mismo he experimentado) y con un movimiento siempre creciente sube más y más hasta que, efectuando sus revoluciones en tiempos iguales que el recipiente, repose relativamente en él. Muestra este ascenso del agua el intento de separarse del centro del movimiento, y por tal intento se manifiesta y se mide el movimiento circular verdadero y absoluto del agua, y aquí contrario totalmente al movimiento relativo […] este conato no depende de la traslación del agua respecto de los cuerpos circundantes y, por tanto, el movimiento circular verdadero no puede definirse por tales traslaciones. Único es el movimiento circular verdadero de cualquier cuerpo que gira, y responde a un conato único como un verdadero y adecuado efecto; los movimientos relativos, en cambio, por las múltiples relaciones externas, son innumerables, pero como las relaciones carecen por completo de efectos verdaderos, a no ser en tanto que participan de aquel único y verdadero movimiento.[2]

Un suceso tan familiar como el agua que tiende a subir por los costados de una cubeta giratoria le proporciona a Newton todas las municiones que necesita para atacar la idea de que el movimiento relativo de los cuerpos constituye todo el movimiento que existe. O, más precisamente, le proporciona las municiones para atacar la explicación relativista del movimiento que Descartes y Aristóteles propugnaron. Uno de los problemas más obvios de una teoría relativista del movimiento consiste en que todos los cuerpos pueden tener muchos movimientos relativos distintos, puesto que se les puede comparar con muchos otros cuerpos. Pero una física del movimiento exige que sólo haya un movimiento en un cuerpo, a saber, el movimiento que pueda explicarse por las leyes de la física. Aristóteles y Descartes se basaron en la misma idea: el movimiento físico importante de un cuerpo es su movimiento con relación al cuerpo que lo contiene directamente.[3] Intentemos aplicar esta explicación al experimento de la cubeta. En el inicio del experimento, antes de que la cubeta se suelte, el agua se encuentra en reposo respecto a la cubeta, y la superficie del agua es llana. Momentos después, cuando el agua y la cubeta giran al mismo tiempo, el agua de nuevo se encuentra en reposo en relación con la cubeta (y por lo tanto, según Descartes y Aristóteles, de verdad en reposo), pero la superficie cóncava indica que la situación es físicamente diferente. Podemos decir que esta situación se deriva del hecho de que el agua no gira en la etapa inicial, pero que sí gira en la etapa posterior; sin embargo, este movimiento no puede ser ni un movimiento *relacionado con la cubeta*, ni un movimiento relacionado con el *cuerpo que lo contiene*.

Un relativista podría intentar responder que el movimiento pertinente, el movimiento que explica el comportamiento del agua, no es un movimiento con relación a la cubeta, pero de todos modos sí es un movimiento relativo. La acción físicamente efectiva de girar es acaso una de girar *en relación con la habitación*, o *en relación con la Tierra*, o *en relación con las estrellas fijas*. Pero al final del "Escolio" Newton presenta el más puro ejemplo posible de los efectos observables del movimiento absoluto en ausencia de cualesquier movimientos relativos.

> Así, si a dos esferas, unidas entre sí por un hilo de determinada longitud, se las hace girar en torno al común centro de gravedad, aparecerá por la tensión del hilo el conato de las esferas de alejarse del eje de giro, y de ello se puede calcular la cantidad de sus movimientos circulares. […] De este modo se podría averiguar la cantidad y la determinación de este movimiento circular, incluso en un vacío inmenso, donde nada hubiese externo y sensible con lo que se pudiesen comparar las esferas.[4]

La tensión observable en la cuerda, la cual produce una fuerza centrípeta, da testimonio de la rotación absoluta de los globos, aun cuando no haya un movimiento relativo en todo el universo: todos los cuerpos materiales (los globos y la cuerda) mantienen una distancia constante entre sí.

El experimento de la cubeta de Newton, a pesar de su gran sencillez, ha sido uno de los experimentos más sólidos y atractivos en la historia de la física. El comportamiento del agua en la cubeta, o la tensión en la cuerda que conecta a los globos, son hechos observables para los cuales es necesaria una explicación. Se nos puede ocurrir la explicación natural: el agua sube por los costados de la cubeta y la cuerda se tensa porque el sistema se encuentra *girando*. Pero la acción de girar es un tipo de movimiento, de manera que debemos preguntar: girando *¿en relación con qué?* El movimiento pertinente e importante no es el movimiento con relación al entorno inmediato. Y si aceptamos el experimento especulativo de los globos en un espacio vacío, el movimiento relevante no puede ser el movimiento con respecto a cualquier tipo de cuerpo material. Newton concluye que el movimiento tiene que ser un movimiento con relación al espacio absoluto: los cuerpos giratorios sucesivamente ocupan diferentes ubicaciones en el espacio mismo. De esta manera, los movimientos absolutos se conectan a fuerzas y por lo tanto a efectos observables.[5]

Hoy en día, casi nadie cree que el espacio absoluto persiste a través del tiempo. Por lo tanto, nadie cree en el movimiento absoluto, tal como Newton lo concebía. Sin embargo, incluso en la teoría de la relatividad, el experimento de la cubeta de Newton sigue siendo una demostración de que el movimiento relativo de los cuerpos no explica el fenómeno que Newton expone. Aun en la relatividad, sigue siendo verdad que un par de globos conectados en un universo vacío (excepto por las cuerdas) podría mostrar una tensión en la cuerda que los conecta y que dicha tensión sería una indicación de que los globos no están siguiendo, en un sentido absoluto, una trayectoria natural, inalterada. Incluso en la relatividad, existe el hecho *absoluto* respecto a si los globos giran o no en torno a un centro común de gravedad.[6]

Sin el espacio y el tiempo absolutos de Newton, no es claro cómo es posible darse una idea de la rotación absoluta, la rotación "relacionada con el espacio mismo". Los recursos conceptuales que son necesarios para entender cómo separar el meollo de la dinámica newtoniana del espacio y tiempo absolutos sólo fueron desarrollados siglos después de la muerte de Newton. Aún hoy, pocos manuales de física explican la situación con claridad. De hecho, para comprender cómo puede existir la rotación absoluta sin el espacio y el tiempo absolutos, tendremos que reconceptualizar la geometría del espacio y del tiempo, amalgamándola como un solo ente. Pero antes de dar este paso, hagamos una pausa para examinar brevemente la forma en que se utilizan las matemáticas para exponer la teoría de Newton.

ARITMÉTICA, GEOMETRÍA Y COORDENADAS

He querido presentar la primera ley del movimiento y la segunda ley del movimiento sin recurrir a ninguna ecuación matemática.[7] Se suele decir que las presentaciones no técnicas de la física deben evitar las ecuaciones, puesto que éstas suelen intimidar a los lectores, pero no ha sido ésta la razón por la que no he querido utilizar las ecuaciones estándar, tal como **F** = m**A**. No he querido utilizar ecuaciones porque Newton no utilizó ecuaciones: él expuso su teoría de manera geométrica y demostró los teoremas en los *Principia* mediante las técnicas de la geometría euclidiana. No es ésta sólo una curiosidad histórica. Newton hubo de presentar su física geométricamente porque la problemática —el movimiento en el espacio— es en sí misma geométrica. Para decirlo *grosso modo*, el mundo físico no se compone de los *números* o de los entes para los cuales se hicieron las operaciones aritméticas estándar. El mundo físico contiene *dimensiones físicas* que poseen una estructura geométrica. La geometría está conectada más directamente con el mundo físico que la aritmética.

Por ejemplo, he dicho que Newton presupone que la estructura geométrica del espacio absoluto es E^3: la geometría tridimensional que los axiomas de Euclides describen. Si tal cosa es correcta, entonces la utilización de la geometría euclidiana constituye una ruta directa a la deducción de importantes hechos físicos. No obstante, la física moderna jamás se expone directamente en los términos de la geometría euclidiana, y los estudiantes contemporáneos podrían no captar con facilidad las comprobaciones de Newton o no serían capaces de construir comprobaciones similares. La física moderna se expone de manera *algebraica*, en función de los *números* y las *ecuaciones aritméticas*. Semejante asociación de la física con la aritmética se ha integrado tanto en la gnosis que incluso puede ser difícil notarla. Por ejemplo, un típico texto físico no se refiere al espacio tridimensional euclidiano como E^3; en cambio, se refiere a él como R^3. Y "R^3" tiene un significado matemático específico: R^3 es el conjunto de las tripletas ordenadas de los números reales. Un elemento de R^3 tiene la forma (x, y, z) en la que cada una de las variables representa un número real específico: $(1, 3.52, \sqrt{17})$ es un elemento de R^3. Pero $(1, 3.52, \sqrt{17})$ obviamente no es un elemento del espacio tridimensional euclidiano. Además, R^3 posee una estructura matemática que en gran parte no es análoga a la de E^3. Por ejemplo, dado cualquier elemento de R^3, podemos preguntar si los números en el primer y segundo lugares son los mismos o son diferentes. Y dados cualquier dos elementos de R^3, podemos preguntar si tienen los mismos números o números diferentes en sus primeros lugares. Dados cualquier par de elementos de R^3 (x_1, y_1, z_1) y (x_2, y_2, z_2), es fácil definir su *suma* como $(x_1 + x_2, y_1 + y_2, z_1 + z_2)$ y fácil definir lo que significa multiplicar un elemento por un escalar c: $c \times (x_1, y_1, z_1) = (cx_1, cy_1, cz_1)$. Pero ninguna de estas operaciones tiene análogos para los puntos en E^3. Los puntos en E^3 no tienen "primeros lugares" y "segundos lugares"; estos puntos no pueden "sumarse" de ninguna manera bien definida, y no tiene sentido preguntar cuál sería el resultado de multiplicar un punto por un valor escalar. R^3 y E^3 obviamente son objetos matemáticos diferentes. ¿Cómo ha sido, pues, que los dos se hayan llegado a confundir tan fácilmente?

La respuesta yace en el uso de los *sistemas de coordenadas*. Un sistema de coordenadas en E^3 establece una correspondencia de uno a uno entre los elementos de E^3 (puntos geométricos) y algún tipo de objeto aritmético. Es usual que los elementos de R^3 se utilicen como coordenadas para los puntos de E^3. Dado semejante sistema de coordenadas, la tripleta de los números reales $(1, 3.52, \sqrt{17})$ viene a *representar* un punto de E^3. La ventaja de utilizar semejantes representaciones aritméticas de los objetos geométricos está en que nos permite utilizar métodos *algebraicos* para solucionar problemas *geométricos*. Mediante la utilización de un sistema de coordenadas, los problemas de geometría pueden convertirse en cuestiones de aritmética, permitiendo que los métodos aritméticos se utilicen para solucionarlas.

No es posible exagerar el poder matemático de esta traducción de la geometría en aritmética. Pero este poder acarrea consigo un peligro: puede oscurecerse la distinción entre el objeto de estudio (digamos, un espacio geométrico) y una *representación* de este objeto (digamos, un conjunto ordenado de *n*-tuplas de números). No es éste un problema trivial: Einstein dice en sus notas autobiográficas, con relación a la prolongada incubación de la teoría de la relatividad general:

> ¿Por qué hicieron falta otros siete años para establecer la teoría general de la relatividad? El motivo principal radica en que no es tan fácil liberarse de la idea de que las coordenadas deben poseer un significado métrico inmediato.[8]

Incluso para Einstein fue difícil distinguir entre aquellos aspectos de las coordenadas que poseen cierto tipo de significación geométrica y aquellos que no la poseen. Con el fin de entender cómo las coordenadas pueden codificar una estructura geométrica, consideremos la aplicación de coordenadas del plano euclidiano bidimensional E^2.

Todo el mundo conoce las coordenadas cartesianas del plano euclidiano: cada punto en el plano recibe una coordenada *x* y una coordenada *y*, siendo cada una de éstas un número real. Así, cada punto llega a asociarse con un elemento (x, y) de R^2. Ahora bien, podríamos preguntar por qué se utilizan dos números reales en vez de sólo *uno*. Al fin y al cabo, es posible codificar la misma información que contienen un par de números reales en tan sólo un número. Por ejemplo, podríamos intercalar los dígitos en las expansiones decimales de los números reales: el par de números reales (.12341234…, .8888…) se asignaría al número real único .1828384818283848… Semejante sistema es, en cierto sentido, más eficaz que la práctica usual: en vez de que sea necesario un par de números para indicar un punto específico en el espacio, un solo número sería suficiente.

No obstante, la utilización de este sistema tiene serios inconvenientes. Si quisiéramos asociar de esta manera los puntos en el plano con números, *el movimiento continuo de un punto en el plano se representaría por cambios discontinuos en las coordenadas del punto*: los puntos cercanos en el plano no siempre tendrían valores de coordenadas cercanos. Veamos un ejemplo. Empecemos con las coordenadas cartesianas estándar en el plano, y luego convirtámoslas en una sola coordenada mediante la intercalación de los dígitos, tal como vimos arriba. Ahora consideremos un punto que empieza en el origen (0, 0) del sistema cartesiano de coordenadas y que se mueve continuamente a lo largo del eje x hasta (.9, 0). En este movimiento, la coordenada y siempre está en 0. Por lo tanto, en función de nuestra coordenada única, el movimiento se inicia en .00000… y termina en .90000, pero nunca pasa a través del punto marcado como .01000… El movimiento continuo en E^2 se representa por un cambio discontinuo en la coordenada. Pero si usamos un par de coordenadas cartesianas, el problema se soluciona: cuando un punto se mueve continuamente en el espacio, cada una de sus coordenadas cambia continuamente.

Este aspecto de un sistema de coordenadas bien adaptado es tan básico, y tan obvio, que a nadie se le ocurriría jamás algo semejante a nuestro procedimiento de la "intercalación". Pero, a pesar de su obviedad, merece cierto reconocimiento explícito. El requerimiento básico de cualquier sistema de coordenadas aceptable consiste en que la topología de las coordenadas se ajuste adecuadamente a la topología del espacio al que se pretende aplicar coordenadas. Las funciones de las coordenadas, cuyo dominio es el espacio y cuyo alcance está constituido (en este caso) por los números reales, tienen que ser necesariamente funciones continuas. En el caso de las coordenadas cartesianas en E^2, se dan dos funciones de este tipo: una función $X(p)$ de la coordenada x y una función $Y(p)$ de la coordenada y. Para cada punto p en E^2, estas funciones definen un par de números que son las coordenadas del punto dado $(X(p), Y(p))$. Si estas funciones de las coordenadas son continuas en todas partes, entonces podemos recuperar la topología del espacio mismo en la topología de las coordenadas. Las coordenadas llegan a representar, en la forma más directa posible, la estructura geométrica del espacio al que estamos aplicando coordenadas.

Una vez que comprobamos en qué forma un sistema de coordenadas bien elegido puede adaptarse a la topología del espacio al que se están aplicando coordenadas, resulta natural preguntarnos si es posible que los demás aspectos geométricos del espacio se reflejen en las coordenadas de un modo tan transparente. Podemos preguntarnos esto con relación a la estructura diferencial, la estructura afín y la estructura métrica del espacio.

Dada una estructura diferencial, es natural que exijamos que las curvas de las coordenadas no sólo sean continuas sino también suaves: no deben tener esquinas puntiagudas. Esto por lo general se presupone sin comentarios.

En el caso de las coordenadas cartesianas en E^2, la conexión entre la geometría del espacio y el álgebra de las coordenadas es particularmente clara. Consideremos primero la estructura afín. Dejemos que A, B, C y D sean números reales, evitando que A y C sean ambos 0. Entonces el conjunto de puntos con coordenadas cartesianas ($As + B$, $Cs + D$) forman una línea recta completa, con s tomando valores en los números reales. Por ejemplo, el eje x se corresponde con las opciones $A = 1$, $B = C = D = 0$; el eje y se corresponde con $C = 1$, $A = B = D = 0$; y la diagonal $x = y$ se corresponde con $A = C = 1$, $B = D = 0$. Inversamente, cada línea recta en E^2 se corresponde con por lo menos una (de hecho, con infinitas) opción para A, B, C y D. De manera que las líneas rectas en E^2 se corresponden con ciertas *ecuaciones lineales* para las coordenadas.

¿Y qué sucede respecto a la estructura métrica de E^2? La estructura métrica se relaciona con las distancias entre los puntos. Y la distancia entre dos puntos en E^2 es la longitud de la línea recta que los conecta. De modo y manera que la métrica determina, en primer lugar, las *proporciones de las longitudes de las líneas rectas*. Si se determinan todas las proporciones de este tipo, entonces el conjunto de círculos también se determina: un círculo en torno al punto p es el conjunto de todos los puntos que se encuentran equidistantes de p, es decir, cuya distancia de p está en proporción 1:1.

La estructura métrica de E^2 se puede expresar de manera particularmente sencilla en términos de las coordenadas cartesianas. Una *función métrica* es una función $D(u, v)$ de pares de puntos en un espacio a los números reales que satisface las siguientes tres condiciones:

1. $D(u, v) \geq 0$, con $D(u, v) = 0$, si y sólo si $u = v$ (definición positiva)
2. $D(u, v) = D(u, v)$ (simetría)
3. $D(u, v) + D(v, w) \geq D(u, w)$ (desigualdad triangular)

A $D(p, q)$ a veces se le llama la *distancia* desde p a q, pero esta terminología es algo confusa. $D(x, y)$ es una función desde los puntos en el espacio hasta los números reales. Pero si preguntáramos cuál es la distancia desde el punto p en E^2 hasta el punto q, la respuesta correcta no podría ser nunca un número real, tal como 4; ¿4 *qué?*. Cuando especificamos una distancia mediante la utilización de un número real, siempre tenemos que utilizar una unidad o una escala. Así que el valor particular de la función $D(u, v)$ no tiene sentido geométrico.

Pero sí tienen sentido geométrico las *correspondencias* entre los valores de las funciones métricas. Si $D(p, q) = 2$ y $D(r, s) = 4$, entonces la correspondencia entre la longitud de la línea recta pq y la longitud rs es 1:2, es decir, pq es dos veces más larga que rs. De manera que dos *funciones* métricas que difieren la una de la otra por un factor de escala constante representan la misma *estructura* métrica en el espacio. La libertad para elegir cualquiera de estas diferentes funciones con el fin de representar la misma geometría se denomina *libertad de norma*. Podemos encontrar muchos tipos de semejante libertad de norma cuando utilizamos los números para representar una estructura geométrica, y es importante desarrollar la comprensión de cuándo ocurre esto.

En el caso de E^2 con coordenadas cartesianas, una función métrica se especifica mediante las coordenadas de los puntos, utilizando la fórmula pitagórica:

$$D(p, q) = \sqrt{(X(p) - X(q))^2 + (Y(p) - Y(q))^2} \qquad \text{Ecuación I.1}$$

Los argumentos de esta función son los puntos p y q en E^2. Estos puntos no tienen propiedades aritméticas, de manera que no tiene sentido sustraerlos o elevar al cuadrado el resultado. Pero las coordenadas de estos puntos, $X(p)$, $Y(p)$, y $Y(q)$ son números reales: pueden someterse a las operaciones aritméticas usuales.

La estructura geométrica completa de E^2 —su topología, estructura afín y métrica— puede recuperarse fácilmente en un conjunto de coordenadas cartesianas. Y, lo que es más importante, cualquier espacio puede identificarse como E^2 mediante el criterio de que *admite* coordenadas cartesianas. Es decir, un espacio es E^2 si y sólo si existe un par de funciones X y Y desde el espacio hasta los números reales, tal que *1)* X y Y sean ambas continuas; *2*) las líneas rectas completas se correspondan con conjuntos de puntos cuyas coordenadas sean ($As + B$, $Cs + D$) para una opción de A, B, C y D, y *3*) las correspondencias entre las longitudes de las líneas rectas sean las correspondencias de los valores de la función dada por la ecuación I.1. Un espacio debe tener una cierta forma geométrica para que sea capaz de posibilitar semejante aplicación de coordenadas, de modo que podemos utilizar la existencia de tales sistemas de coordenadas como una forma indirecta de especificar la geometría de un espacio.

Una vez que se haya elegido un sistema de coordenadas para un espacio, las leyes dinámicas (es decir, las leyes que especifican las trayectorias de los objetos en el espacio) pueden enunciarse en una forma algebraica *con relación a esas coordenadas*. Esto es lo que usualmente se hace en los manuales contemporáneos de la física al exponer la mecánica newtoniana. Por ejemplo, el manual de Herbert Goldstein, *Classical Mechanics*, describe la teoría de Newton como sigue:

> Sea **r** el radio vector de una partícula desde un origen dado y **v** su vector velocidad:
>
> $$\mathbf{v} = \frac{d\mathbf{r}}{dt}.$$
>
> El *momentum lineal* **p** de la partícula se define como el producto de la masa y de la velocidad de la partícula:
>
> $$\mathbf{p} = m\mathbf{v}.$$

En consecuencia de sus interacciones con objetos y campos externos, la partícula puede experimentar fuerzas de varios tipos, por ejemplo, gravitacionales o electrodinámicas; la suma de los vectores de estas fuerzas que se imprimen en la partícula equivale a la fuerza total **F**. La mecánica de la partícula se encuentra en la segunda ley del movimiento de Newton, la cual asevera que existen marcos referenciales en los que el movimiento de la partícula se describe por la ecuación diferencial[9]

$$\boldsymbol{F} = \frac{d\boldsymbol{p}}{dt}.$$

La descripción de la teoría de Newton que presenta Goldstein poco tiene en común con la de Newton. Ya desde la primera frase Goldstein invoca un vector "desde un origen dado", aunque Newton nunca mencionara un "origen". Ya desde el inicio Goldstein presupone (sin decirlo) cierta aplicación de coordenadas del espacio, siendo el "origen" el punto con las coordenadas (0, 0, 0). Además, todos los símbolos que Goldstein utiliza — a saber, **r, v, p, F**, *m* y *t*— supuestamente son números o tienen propiedades algebraicas: **v** puede multiplicarse por *m*, y las derivadas (que se definen por medio de límites de operaciones aritméticas) se supone que existen. Más notable aún, según Goldstein la segunda ley del movimiento de Newton menciona la existencia de "marcos referenciales". Sin embargo, como hemos visto, Newton expone su segunda ley sin mencionar cosa semejante: de hecho, él no tenía idea de lo que *era* un "marco referencial". Es también notable que Goldstein de alguna manera salta hasta la segunda ley sin hacer mención siquiera de la primera; pero, como hemos visto, la segunda ley tiene que *presuponer* la primera ley incluso para hacerse de un contenido.

En suma, la física moderna se ha aritmetizado hasta tal punto, se ha vuelto tan dependiente de los números que han sido introducidos en ella mediante los sistemas de coordenadas, que de buenas a primeras nos resulta un poco difícil reconocer la teoría original de Newton en las presentaciones modernas. Éstas son mucho más eficaces cuando se trata de solucionar los problemas que allí se exponen: la utilización del cálculo elimina la necesidad de los diagramas y las construcciones geométricas de Newton. Pero al mismo tiempo se ha vuelto más arduo dilucidar cuál es exactamente la estructura de la teoría y, en particular, lo que la teoría afirma con relación al mundo. Por ejemplo, a veces se ha dicho que las leyes del movimiento de Newton sólo son válidas en *marcos de referencia inerciales*. Y si insistimos, preguntándonos qué es exactamente un marco de referencia inercial, podemos caer en la tentación de decir que es un marco de referencia en el que las leyes de Newton son válidas. Un círculo definicional tan estrecho tendría que incomodarnos, pero el procedimiento es en cierto modo comprensible. De la misma manera que la estructura geométrica intrínseca de E^3 implica que los sistemas de coordenadas cartesianos existen, la estructura geométrica intrínseca del espacio y el tiempo según Newton implica la existencia de conjuntos de coordenadas especiales. En las más convenientes aplicaciones de coordenadas del espacio y el tiempo newtonianos, la aceleración de una trayectoria a través del tiempo es proporcional a la segunda derivada de las coordenadas espaciales con relación a la coordenada del tiempo. Por ejemplo, el cambio en el estado de movimiento en una dirección dada (llamémosla la dirección *x*), que para Newton es una cantidad que existe independientemente de todas las coordenadas, se vuelve proporcional a d^2x/dt^2, donde *x* y *t* son las coordenadas en este sistema especial. En semejante planteamiento, la existencia de tales coordenadas convenientes es un asunto derivado: se derivan de la estructura misma del espacio-tiempo. Las leyes de Newton (tal como Newton las formuló) describen el movimiento de los cuerpos en un espacio con la estructura geométrica de E^3 que persiste a través del tiempo. No es necesaria la mención de un "marco de referencia inercial" y tampoco de ningún tipo de "marco de referencia".[10]

Los sistemas de coordenadas mismos pueden caracterizarse con referencia a la estructura geométrica del espacio que se pretende coordenar. Por ejemplo, se dice que son *rectilíneas* y *ortogonales* las coordenadas cartesianas en un espacio euclidiano. Esto significa que las *curvas coordenados* son líneas rectas que se cruzan en ángulos rectos. Una curva coordenada consiste en el conjunto de puntos en el espacio que obtenemos si todas, excepto una, de las coordenadas se fijan y a la coordenada restante se le permite variar. Así, por ejemplo, veamos el caso de ciertas coordenadas cartesianas (x, y) en E^2. Si fijamos la coordenada x en 2 y dejamos variar la coordenada y, obtenemos todos los puntos con $x = 2$. Esto forma una línea "vertical" en el plano (figura II.3). Las curvas coordenadas x en un sistema de coordenadas cartesiano son las líneas rectas "horizontales", y las curvas coordenadas y son las líneas rectas "verticales". Cualquier par de curvas coordenadas que se cruzan forman un ángulo recto en el punto en que se encuentran.

Es obvio que la caracterización de un sistema de coordenadas como "rectilíneo" u "ortogonal" depende del hecho de que el espacio mismo, independientemente de todas las coordenadas, tenga una cierta estructura geométrica. Si el espacio no tiene una estructura afín, entonces no es posible clasificar las curvas coordenadas ni como rectas ni como curvas; si el espacio no tiene una estructura métrica, entonces no es posible clasificar el ángulo en que las curvas se encuentran ni como recto, ni como agudo, ni como obtuso. Por lo tanto, este tipo de taxonomía de los sistemas de coordenadas presupone una cierta estructura geométrica en el espacio mismo.

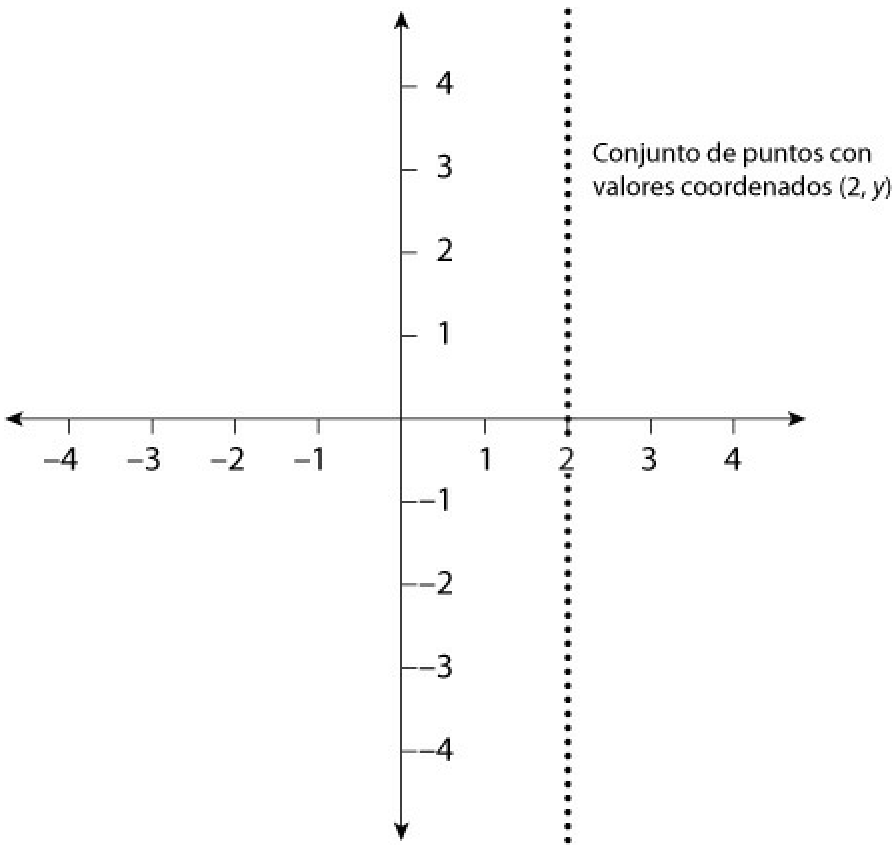

FIGURA II.3

De forma similar, el hecho de caracterizar un sistema de coordenadas que se utilice en la física como "inercial" o "no inercial" también presupone que el espacio y el tiempo físicos poseen algún tipo de estructura geométrica concreta. Pero la naturaleza exacta de semejante estructura geométrica no es tan evidente. Ya hemos visto los rasgos que el mismo Newton le atribuye al espacio absoluto: éste se caracteriza por una geometría tridimensional euclidiana, y los puntos individuales del espacio absoluto persisten a través del tiempo, conservando sus relaciones geométricas. Resulta que la dinámica de Newton no requiere de tanta estructura: es posible formular sus leyes del movimiento sin referirse al espacio y el tiempo absolutos. No obstante, toma tiempo y trabajo y paciencia llegar a hacer esto.

Tras haber incursionado así en los sistemas de coordenadas y en las representaciones aritméticas de la estructura geométrica, intentaré, hasta donde sea posible, presentar las ideas físicas sin utilizar ni coordenadas ni números. La forma más efectiva para hacer esto es mediante gráficas: los diagramas se pueden usar como representaciones geométricas de hechos geométricos, ya que son capaces de proporcionar una rápida captación intuitiva de las ideas fundamentales. La interpretación de estos diagramas requiere bastante precaución porque los diagramas, siendo objetos espaciales, frecuentemente poseen rasgos geométricos que no se corresponden con ninguna cosa en el espacio representado. Así que el lector deberá estudiar los diagramas cuidadosamente, aunque teniendo en cuenta las observaciones respecto a la manera en que se deben interpretar.

Ahora podemos continuar con nuestra revisión histórica retomándola en el punto donde la dejamos.

Las simetrías del espacio y el debate Leibniz-Clarke

Hemos dado una explicación de algunas de las estructuras geométricas fundamentales que un espacio puede tener —topológicas, afines y métricas— y hemos visto que la explicación newtoniana del movimiento absoluto presupone un espacio que posee las tres y cuyos puntos individuales persisten a través del tiempo absoluto. El tiempo absoluto mismo posee una topología, una estructura métrica, y también, a diferencia del espacio, una direccionalidad. La explicación de las leyes del movimiento de Newton, sin embargo, no utiliza de manera esencial la estructura euclidiana del espacio: aunque Newton supone que el espacio es E^3, ninguna parte de la dinámica newtoniana lo requiere. Pero la propuesta específica de E^3 dio origen a un conjunto de objeciones a la teoría de Newton.

El espacio euclidiano tridimensional muestra un alto grado de simetría. Tales simetrías se manifiestan en las isometrías del espacio: las asignaciones del espacio al espacio mismo que preservan las distancias entre los puntos. La isometría que se muestra en la figura I.1 muestra dos de estas simetrías: la simetría traslacional y la simetría rotacional. Es decir, para generar la imagen de abajo en la figura I.1 a partir de la imagen de en medio, podemos pensar primero en mover todos los puntos la misma distancia hacia abajo en la página (traslación) y luego en girar todo ello en torno a un eje (rotación). Todas las partes del diagrama tienen las mismas relaciones geométricas entre sí, una vez que la traslación y la rotación se hayan realizado.

A un espacio que posee simetría traslacional se le denomina *homogéneo*, y a uno que muestra simetría rotacional en torno a un punto se le denomina *isotrópico*. En un espacio homogéneo, todas las ubicaciones "parecen ser las mismas": es posible mover todas las cosas a lo largo y ancho de ese espacio conservando las mismas relaciones geométricas que tienen entre sí. Si un espacio es isotrópico en un punto, entonces todas las direcciones en ese punto "parecen ser las mismas": todas las cosas se pueden reorientar mediante la rotación en torno al punto, conservando, aun así, todas sus relaciones geométricas.

El universo esférico de Aristóteles no es homogéneo: es imposible trasladar cualquier cosa en cualquier dirección porque los límites del espacio no lo permiten. La física de Aristóteles refleja esta ausencia de homogeneidad. Una piedra que no se encontrara en el centro del universo no se comportaría de la misma forma como lo hiciera en el centro: desde su nueva ubicación regresaría espontáneamente al centro. El universo de Aristóteles es isotrópico en el centro, pero no lo es en cualesquier otro punto: si una cosa no se encuentra en el centro del universo, entonces se produce una sola dirección "hacia abajo” que se dirige hacia el centro. Pero E^3 posee una simetría completa, tanto traslacional como rotacional. Todo punto puede trasladarse a lo largo de una distancia dada sin que se modifiquen las relaciones geométricas entre los puntos, y cualquier punto puede hacerse girar en torno a cualquier eje sin que se produzca un cambio en las relaciones geométricas entre los puntos. Estas simetrías se asumen de manera implícita en la geometría euclidiana: debería ser posible dibujar cualquier diagrama euclidiano y obtener el mismo resultado sin que tenga importancia el sitio en el mundo donde se encuentre el papel o la forma en que se oriente. Si no fuera así, las instrucciones para dibujar un diagrama euclidiano tendrían que especificar en qué lugar se debería dibujar.

Como consecuencia de estas simetrías de E^3, es posible que las cosas se encuentren distribuidas de diferentes maneras en E^3 y sin embargo conserven las mismas ubicaciones *relativas*. Tómese cualquier distribución de cosas y muévase toda ella una distancia dada, o hágase girar toda ella la misma distancia en torno a un eje, y se comprobará que las posiciones relativas de los objetos no habrán de cambiar. Si la distribución de las cosas misma no tiene la misma simetría, entonces la nueva distribución será diferente de la anterior, pero todas las posiciones relativas serán iguales.

En un famoso intercambio epistolar indirecto, Gottfried Leibniz discutió con Samuel Clarke sobre la aceptabilidad de la física de Newton. Inicialmente, el debate giró en torno al tema de si la física de Newton socavaría las creencias religiosas. En el curso de la discusión, Leibniz introdujo un anuncio publicitario respecto a su propia filosofía, afirmando que él había elevado la metafísica hasta convertirla en una ciencia demostrativa mediante un único principio: el *principio de la razón suficiente* (PRS).

> Luego por ese solo principio, a saber, que es necesario que haya una razón suficiente por la que las cosas sean más bien así que de otra manera, se demuestra la Divinidad, y todo el resto de la metafísica o de la teología natural, e incluso, de alguna manera, los principios físicos independientes de la matemática, es decir, los principios dinámicos o de la fuerza.[11]

Clarke, que aparentemente estaba de acuerdo con Leibniz pero que pretendía demostrar a su vez en qué forma la voluntad de Dios podía ser la causa última de todas las cosas, propuso un fatídico ejemplo:

> Es verdad que nada existe sin que haya una razón suficiente de por qué existe, y de por qué es así antes que de otro modo; por esta razón, donde no hay causa no puede haber efecto. Pero esta razón suficiente con frecuencia no es otra que la mera voluntad de Dios. Por ejemplo, ¿por qué este sistema particular de materia habría de ser creado en un lugar determinado y aquél en otro, cuando siendo todo lugar indiferente a toda materia podría haber sido exactamente al revés? Suponiendo que los dos sistemas (o partículas) de materia son iguales, no puede haber otra razón que la mera voluntad de Dios.[12]

De esta manera tuvo origen uno de los debates más significativos, a la vez que más intrincados y confusos, en la historia de la filosofía.

Puesto que Clarke en este pasaje aparentemente acepta el PRS de Leibniz, parece extraño que pudiera constituirse en el origen de una muy tormentosa discusión. Pero Leibniz se dio cuenta de que el ejemplo dado por Clarke no apoyaba realmente el concepto del PRS. Clarke dice que al crear la materia, Dios podía haber situado a la materia en cualquier lugar en absoluto del espacio absoluto, "siendo todo lugar absolutamente indiferente a toda materia". Supongamos que Dios decide crear la materia en un espacio euclidiano de dos dimensiones completamente vacío y en la forma de un círculo dentro de un triángulo isósceles. Como se ve en la figura I.1. Dios podría haberlo hecho de diferentes formas: la simetría del espacio euclidiano garantiza que semejante sistema de materia podría caber en cualquier lugar y con cualquier orientación. Entonces, según Clarke, Dios debió elegir entre todas las distintas formas en que el espacio absoluto podía dar cabida a la materia. Clarke insiste en que "la mera voluntad de Dios" debió ser el factor determinante, porque no había nada que pudiera hacer que cualquiera de estas posibles distribuciones de la materia fuera superior a cualquier otra.

Leibniz señala el problema de inmediato:

> Se me acepta ese principio importante, que nada ocurre sin que exista una razón suficiente por la que sea así más bien que de otro modo. Pero se me acepta sólo de palabra, y se me niega de hecho. Lo que hace pensar que no han comprendido bien toda su fuerza. Y para eso se valen de un ejemplo que cae justamente bajo una de mis demostraciones contra el espacio real absoluto, ídolo de algunos ingleses modernos.
>
> […]
>
> Digo entonces que si el espacio es un ser absoluto, entonces se daría alguna cosa de la cual sería imposible que hubiera una razón suficiente, lo que va contra nuestro axioma. He aquí como lo pruebo. El espacio es una cosa absolutamente uniforme y, sin las cosas en él colocadas, un punto del espacio no difiere absolutamente en nada de otro punto del espacio. De lo que se sigue, suponiendo que el espacio en sí mismo sea algo distinto del orden de los cuerpos entre sí, que es imposible que haya una razón por la que Dios, conservando las mismas situaciones de los cuerpos entre ellos, haya colocado los cuerpos en el espacio así y no de otra manera, y por la que no haya sido puesto al revés (por ejemplo) por un cambio de oriente y de occidente.[13]

Llamémosle a éste el *argumento del PRS*, puesto que descansa en el principio de la razón suficiente. El argumento también utiliza algunas otras premisas. La premisa que más ha llamado la atención afirma que tanto la existencia como la estructura geométrica del espacio absoluto son independientes de toda materia. Por ende, es posible que el espacio exista sin poseer en absoluto objetos materiales, como un perfecto vacío. Un vacío no es la nada: es *espacio* vacío y, según Newton, semejante espacio vacío en efecto existió por un periodo de tiempo infinito antes de que Dios creara la materia en él.

Esto nos introduce en la segunda premisa del argumento leibniziano: el acto creativo de Dios. Leibniz argumenta que incluso si Dios hubiera querido crear un mundo material en un espacio absolutamente vacío, Él se hubiera visto frustrado: no siendo capaz de encontrar una razón verdadera para ubicar el mundo en un lugar en vez de en otro, Dios se hubiera visto impedido por el PRS para actuar en cualquier sentido. A Clarke esto le parece una impía negación de la omnipotencia de Dios, pero Leibniz insiste en su argumento: "Cuando dos cosas que no pueden estar juntas son igualmente buenas y ninguna de las dos goza, ni en sí misma ni en su conjugación con otras cosas, de una ventaja sobre la otra, Dios no habrá de producir ninguna de las dos".[14] Puesto que carece de un motivo real para preferir una de las dos opciones incompatibles, Dios *no puede* crear ninguna de las dos. Siendo las diversas ubicaciones en el espacio absoluto todas de igual forma adecuadas, Dios no podría decidir voluntariamente la existencia de cualquiera de ellas, a pesar de lo adecuadas que pudieran ser.

El hecho de involucrar a Dios en el argumento del PRS puede volverlo irrelevante para cualquier persona que no acepte una explicación teológica de la creación. ¿Cuál sería la situación, por ejemplo, si *nadie* hubiera creado el mundo material, si éste siempre hubiera existido? Y existen razones suficientes perfectamente buenas para su ubicación en cualquier momento del tiempo, a saber, tanto su ubicación en una época anterior, como las leyes de la física. No resulta totalmente claro si este tipo de explicación solucionaría el problema de Leibniz: aun cuando la ubicación de un objeto en cualquier momento del tiempo se explica por su ubicación en momentos del tiempo anteriores, no se explica por qué existe esta particular distribución de la materia a través de todos los momentos del tiempo en vez de que exista otra (por ejemplo, una en que, durante todos los momentos del tiempo, la materia "mirara hacia el otro lado" en el espacio absoluto).

Otra premisa del argumento PRS es la simetría del espacio absoluto. La aseveración de que "un punto en el espacio no difiere en absoluto de cualquier otro punto del espacio" se basa en la homogeneidad del espacio absoluto. No existe semejante homogeneidad en un espacio esférico aristotélico, de manera que habría una razón para ubicar un objeto en el centro del espacio en vez de en la periferia. Claro, el espacio aristotélico es isotrópico en torno al centro, por lo que no habría motivo para *orientar* la materia ni hacia un lado ni hacia el otro, suponiendo que eso tuviera algún tipo de importancia. (Si la materia misma formara esferas perfectas, todas ubicadas en el centro del universo, entonces el problema de la orientación no surgiría para nada.) La invocación de semejantes simetrías en el argumento del PRS jamás se impugna porque tanto Leibniz como Clarke suponen que el espacio es E^3. Pero si abandonáramos esa suposición y propusiéramos un espacio que careciera de tales simetrías, el argumento del PRS ni siquiera podría despegar. Como veremos cuando incursionemos en la relatividad general, no es ésta una especulación ociosa.

Y, claro es, la suposición principal del argumento del PRS es el PRS mismo. Podemos simplemente rechazar el PRS, asumiendo que existen hechos contingentes en el mundo, cosas que pudieron haber sido diferentes y para las cuales no hay una predeterminada causa o explicación. Puesto que Clarke confiesa, como Leibniz señala, que acepta el PRS, esta respuesta a Leibniz no se basa en los argumentos de aquél.

Es necesario observar hasta qué punto es sólido el PRS en su alcance general. Una de las aplicaciones del PRS se corresponde con las leyes de la dinámica: por lo general el principio se considera como incompatible con el indeterminismo dinámico, es decir, con la posibilidad de que dos sistemas físicos en el mismo estado inicial y el mismo medio ambiente puedan evolucionar de manera distinta. La historia de la física cuántica muestra que en este sentido los físicos actuales no aceptan totalmente el PRS: consideran que las leyes de la dinámica, fundamentalmente indeterminadas, son propuestas válidas en la física moderna. Pero el PRS tiene consecuencias todavía más radicales, las cuales para Leibniz eran evidentes. Si tiene que haber una razón suficiente para el hecho de que las cosas sean de una manera en vez de otra, entonces debería haber una razón que explique por qué este universo exactamente, en todos sus detalles específicos, existe en vez de cualquier otra posibilidad física. Es decir, que tiene que haber una razón suficiente para el *estado inicial* del universo, no sólo una razón suficiente para todas y cada una de las alteraciones físicas en el universo. Leibniz desarrolló su principio hasta su conclusión lógica: tiene que haber un cierto rasgo del universo que explique por qué éste precisamente existe, en vez de cualquier otro. Y para Leibniz esta razón se explica por el hecho de que este universo es el mejor de los mundos posibles. Pocos entre nosotros estaríamos dispuestos a aceptar esta última consecuencia del PRS, pero si rechazamos esta posibilidad, el argumento del PRS de Leibniz ni siquiera puede empezar a desarrollarse.

Si Leibniz en su disputa con Clarke tan sólo hubiera expuesto el argumento del PRS, no sería tan difícil para nosotros remontarnos hasta las premisas del argumento con el fin de examinarlas minuciosamente. Por desgracia, para atacar a Newton, Leibniz también utiliza otro argumento, y éste se basa en premisas totalmente distintas. Es necesario aplicar en este punto un análisis disciplinado para distinguir entre ambos argumentos. No nos ayuda en esto el hecho de que Leibniz introduzca este segundo argumento al mismo tiempo que resume el primero, pero sin declarar que ha introducido consideraciones totalmente nuevas. El pasaje que hemos citado arriba continúa así:

> Pero si el espacio no es otra cosa que ese orden o producto, y no es nada sin los cuerpos más que la posibilidad de colocar en él esos dos estados, uno tal como es, el otro supuesto al revés, éstos no diferirían entre sí: su diferencia no se encuentra más que en nuestra suposición quimérica de la realidad del espacio en sí mismo. Pero en la realidad el uno sería justamente la misma cosa que el otro, ya que son absolutamente indiscernibles y, por consecuencia, no hay lugar para preguntar la razón de la preferencia del uno sobre el otro.[15]

La primera parte de este pasaje hace referencia a la explicación relacionista leibniziana del espacio, a la cual regresaremos más adelante. Pero la última frase introduce una nueva consideración que se basa, no en el PRS, sino en el *principio de la identidad de los indiscernibles* (PII) del mismo Leibniz. Este principio es diferente del PRS, de manera que los argumentos que se basan en uno de ellos deben ser cuidadosamente diferenciados de los argumentos basados en el otro.

El argumento del PII, tal como Leibniz lo formula, es el siguiente: si como Newton afirma el espacio absoluto existe, entonces la situación que resulta cuando la materia se despliega según cierta configuración en el espacio es diferente de la situación que se despliega en algún otro lugar del espacio, incluso si las dos situaciones se relacionan mediante una isometría. Podemos imaginar de manera extravagante que tomamos toda la materia y la arrojamos en otro lugar del universo. Pero estas dos distintas configuraciones de la materia en el espacio absoluto tendrían el aspecto "de ser exactamente las mismas" en un sentido obvio. Ninguna descripción *cualitativa* podría distinguir entre ambas. En este sentido, las dos situaciones serían (cualitativamente) indiscernibles la una de la otra. Pero de acuerdo con el PII no puede haber dos cosas distintas (incluso dos situaciones *meramente posibles*) que sean indiscernibles la una de la otra.[16] Podemos decir sin pensarlo demasiado que dos cosas indiscernibles son idénticas, pero lo que el principio afirma es que simplemente no pueden existir dos cosas indiscernibles. Puesto que la teoría de Newton propone la posibilidad de dos situaciones distintas pero indiscernibles, debe contener un error metafísico. Este error es el planteamiento del espacio absoluto.

Hay que señalar primero que la utilización que Leibniz hace de las situaciones donde la materia se desplaza a diferentes lugares del universo, produciendo diferentes posibilidades, es innecesariamente complicada. Imaginemos un espacio absoluto newtoniano por completo carente de materia, es decir, imaginemos un vacío newtoniano. Se supone que el E^3 vacío contiene infinitas ubicaciones espaciales distintas (o puntos), todas las cuales son exactamente iguales, desde el punto de vista cualitativo. La homogeneidad y la isotropía de E^3 aseguran que ninguna descripción cualitativa podría aplicarse a sólo un punto del espacio absoluto. Puesto que estos puntos distintos del espacio absoluto supuestamente son cualitativamente idénticos, el PII implica que no puede haber en realidad más que uno de ellos, pero el E^3 contiene más de un punto. Por lo tanto, el planteamiento de un espacio absoluto vacío que es el E^3 viola el PII, con independencia de cualesquiera otras consideraciones sobre la materia.

El PII, si se aceptara, también descartaría la mayoría de las explicaciones de la materia, independientemente de las consideraciones sobre el espacio. El mismo Leibniz pensaba que el PII no se correspondía con la teoría atómica de la materia, puesto que se supone que los átomos individuales eran cualitativamente idénticos. Ésta es una de las motivaciones de su teoría de las mónadas, según la cual los constituyentes últimos del mundo son entes pensantes sin extensión, cada uno de los cuales se representa a sí mismo el universo entero desde una perspectiva única. De esto se seguiría que toda mónada es cualitativamente diferente de cualquier otra mónada, desde el punto de vista del contenido de sus representaciones. La física moderna, claro, rechaza este tipo de limitación: piensa que dos átomos de hidrógeno en el estado fundamental son completamente idénticos cualitativamente.

Por lo tanto, la cuestión fundamental en torno al argumento del PII es por qué debemos aceptar el PII, para empezar. Leibniz ofrece esta breve justificación en su cuarta carta: “El hecho de suponer que dos cosas son indiscernibles, significa la suposición de la misma cosa bajo dos nombres distintos”.[17]

La explicación más razonable de esta afirmación es que Leibniz presupone una versión *naif* de la teoría empirista de las ideas. Según ésta, un concepto es una cosa que se parece a una imagen en la mente y que se refiere a lo que la imagen representa. Pero dos *imágenes* con idénticas cualidades representan la misma cosa exactamente. De manera que cuando creemos que podemos *imaginar* dos situaciones distintas pero cualitativamente indiscernibles, nos equivocamos: lo que existe en la mente es la misma idea exactamente, con exactamente la misma referencia que se utiliza dos veces.

Es justo decir que esta explicación de las ideas y de cómo se refieren a las cosas en la actualidad se considera insostenible, de manera que la justificación leibniziana del PII hoy carece de solidez. Pero hay otros tipos de preocupaciones que merecen nuestra atención.

Una objeción al planteamiento de los objetos indiscernibles es más epistemológica que metafísica o semántica: si hay dos objetos indiscernibles, entonces (casi por definición) no podríamos decir cuál es cuál. Y luego podría haber otras cuestiones que ningún tipo de observación sería capaz de responder. El planteamiento de los indiscernibles conllevaría una cierta limitación absoluta sobre lo que podemos saber, incluso en principio.

Aunque parezca muy sencilla esta idea, en realidad es bastante difícil entenderla bien. Cuando intentamos aplicarla a las ubicaciones en el espacio absoluto, todas las cuales supuestamente son cualitativamente idénticas, la conclusión es que "no podríamos saber mediante la observación dónde nos encontramos en el espacio absoluto". Pero al reflexionar en esto, no parece haber una manera de articular con exactitud qué es lo que no podríamos saber. ¿Cuál es exactamente la pregunta cuya respuesta no podemos proporcionar? *¿Dónde nos encontramos?* Y bien, ¿cuáles son las posibles respuestas a esta pregunta, esas respuestas entre las cuales no podemos decidirnos por una? Como veremos en la próxima sección, la teoría de Newton se compromete a que haya preguntas bien formuladas con diferentes respuestas posibles que no puedan resolverse mediante la observación. Pero "¿dónde nos encontramos en el espacio absoluto?" no es una de esas preguntas.

Incluso si lo fuera, ¿cuál es el principio que nos permite sacar conclusiones metafísicas de premisas epistemológicas? El hecho de que no podamos determinar algo empíricamente no significa que no hay nada que determinar. No es necesario que el mundo y nuestros sentidos hayan sido construidos de tal forma que todos los hechos físicos puedan resolverse mediante la observación. Resulta más milagroso el hecho de que podamos descubrir *cualquier cosa* sobre el mundo mediante nuestra interacción con él, que el hecho de que no podamos descubrir *todas las cosas* respecto a él.

Claro es, la incapacidad para determinar mediante la observación un hecho que haya sido propuesto por una teoría física nos proporciona el fundamento para un cierto tipo de sospecha. Si el supuesto hecho físico no es importante para el comportamiento observable de las cosas, entonces caemos en la sospecha de que realmente no tiene gran importancia en la teoría física. Acaso una nueva teoría física que haya sido purgada del hecho inaccesible nos pueda otorgar las mismas consecuencias observables a un menor costo ontológico. El recurso de la navaja de Ockham podría entonces militar en favor de la teoría renovada y en contra de la original.

Con frecuencia tal cosa es más fácil de decir que de hacer. Los hechos inobservables pueden encontrarse tan inter-conectados dentro de la producción del comportamiento físico observable, de acuerdo con una teoría, que no es posible separarlos. Los propios esfuerzos de Leibniz para purgar a la física del espacio absoluto proporcionan un ejemplo de esta situación.

Newton identificó los hechos físicos *observables* con los hechos *relativos, perceptibles*, como los de la posición y la velocidad de una porción observable de materia en comparación con otra. Las versiones absolutas de estos hechos, los movimientos absolutos, no son observables del mismo modo. Y la relación entre los hechos observables y los inobservables es perfectamente clara: el movimiento relativo de dos objetos se determina por la diferencia en su velocidad absoluta. Aun cuando la velocidad absoluta de la partícula 1 sea inobservable y la velocidad absoluta de la partícula 2 sea inobservable, no podemos asignarle a cada una algún valor que nos agrade: su movimiento relativo es observable.

La idea de Leibniz consistió en enmarcar la física directamente en términos de las cantidades relativas de Newton: las ubicaciones y velocidades relativas de las cosas materiales. Leibniz hubiera querido desnatar las porciones observables de la ontología newtoniana, y dejar lo demás intacto. Pero ya conocemos el argumento de Newton contra la factibilidad de tal pretensión: el argumento de la cubeta o, más directamente, el argumento de los globos giratorios. Si aceptamos que los globos, en un universo vacío (excepto por la presencia de ellos mismos), podían estar girando o no estar girando y en consecuencia podían tensar o no tensar la cuerda que los unía, no es posible explicar la diferencia física entre estas situaciones en función de los movimientos relativos de los cuerpos. Todos los cuerpos materiales en ambos casos se encuentran en relativo reposo.

Es extraño que Clarke no hubiera mencionado el argumento de Newton hasta un momento tardío de la correspondencia (cuarta respuesta, § 13); en su respuesta, Leibniz aparentemente concede la afirmación fundamental de Newton:

> Sin embargo, estoy de acuerdo en que hay diferencia entre un verdadero movimiento absoluto de un cuerpo y un simple cambio relativo de su situación por referencia a otro cuerpo. Pues cuando la causa inmediata del cambio está en el cuerpo, éste está verdaderamente en movimiento y entonces la situación de los otros en relación con él estará, en consecuencia, cambiada, aunque la causa de este cambio no esté en ellos. Es verdad que, hablando con exactitud, no hay cuerpo que esté perfecta y enteramente en reposo; pero es de esto de lo que se hace abstracción, al considerar la cosa matemáticamente.[18]

No queda totalmente claro en qué sentido Leibniz se imaginaba que ésta fuera una respuesta al argumento de Newton. Si la cubeta giratoria o los globos giratorios se encuentran en "verdadero movimiento absoluto" gracias a la presencia (observable) de las fuerzas que en ellos influyen, entonces este verdadero movimiento absoluto no puede ser una especie de movimiento relativo: no existe en absoluto en el caso de los globos. Pero si hay un movimiento que no es un movimiento relativo de los cuerpos, y si el movimiento es el cambio de ubicación, entonces hay ubicaciones que no son relativas a los cuerpos, todo lo cual es justamente lo que Newton argüía. De modo y manera que Leibniz jamás enfrenta el reto que Newton presenta y que consiste en explicar el comportamiento observable de los cuerpos sin invocar el espacio, el tiempo y el movimiento absolutos.

En la era moderna se ha intentado varias veces enfrentar el reto de Newton. La tentativa más famosa ha sido la de Mach, aunque no alcance a ser una teoría física acabada. Es claro que para un relativista de corte leibniziano la pretensión de superar el ejemplo de los globos es nada menos que una tarea insuperable: si no existe un movimiento relativo de los globos con relación a ningún otro cuerpo material, el relativista no puede invocar especie alguna de "movimiento absoluto" para explicar las diversas tensiones que experimenta la cuerda. Por lo tanto, Mach simplemente *rechaza* el ejemplo, diciendo que no podemos darnos una idea de lo que le podría suceder a un par de globos en un universo vacío: esta situación se encuentra demasiado distante de la realidad para que nuestra experiencia real de las cosas pueda servirnos de guía.

Con relación al experimento real de la cubeta, Mach asevera que los cambios de forma en la superficie del agua se explican con referencia al movimiento relativo del agua *respecto a la esfera de las estrellas fijas*. Es claro que Mach no puede mencionar ningún movimiento relativo local: la rotación de la Tierra como un todo, por ejemplo, se evidencia en el engrosamiento del ecuador y en el movimiento giratorio de los huracanes. Por ende, para encontrar un movimiento relativo que pueda explicar los fenómenos, Mach se vio obligado a buscarlo más allá de nuestro sistema solar. Pero él jamás pudo proponer una teoría verdadera: una dinámica que se expresara en función de los movimientos relativos y que hiciera el tipo de predicciones sobre los fenómenos observables que la teoría de Newton hace. La investigación más extensa de la forma en que alguien podría tratar de enmarcar a la física sobre la base de las cantidades relativas se debe a Julian Barbour y Bruno Bertotti.[19] Este planteamiento también tiene que rechazar el ejemplo de los globos y establecer como una verdad necesaria la suposición de que la masa total del universo no se encuentra en estado de rotación. Puesto que la teoría de Newton sí permite tal rotación, esta teoría no logra recuperar por completo la teoría newtoniana, incluso en relación con movimientos relativos físicamente posibles de los cuerpos.

Cualquier evaluación de la teoría de Newton por sus impugnadores tiene que empezar con una exposición clara de los defectos de la mecánica newtoniana que la nueva teoría pretenda subsanar. Para Mach, el defecto era de cariz filosófico: pensaba que puesto que la evidencia en que descansa la física se relaciona con las propiedades observables de los cuerpos observables, la física misma debería formularse totalmente en esos términos. Esto le condujo a rechazar la teoría atómica de la materia, tanto como el espacio y el tiempo absolutos. Pero el negocio de la física obviamente consiste en proponer entes inobservables para explicar el comportamiento observable de las cosas. Estas propuestas son siempre riesgosas, pero es evidente, gracias a la hipótesis atómica, que el riesgo a veces aporta grandes beneficios. Newton bien sabía que el espacio y el tiempo absolutos no son, en sí mismos, observables; sin embargo, él también sabía que el hecho de teorizarlos podía ayudar a explicar los hechos observables. ¿En qué sentido es esto peor que proponer átomos?

Aun así, el carácter inobservable del movimiento absoluto newtoniano constituye un problema preocupante. Ya hemos visto que no es fácil precisar la supuesta inobservabilidad de la ubicación en el espacio absoluto, ya que no queda claro qué tipo de información podría ser capaz de transmitirnos la observación. Si Clarke está en lo cierto, el universo material *hubiera podido* ubicarse en otro lugar del espacio absoluto; es decir, ubicarse en un lugar diferente del lugar en que está, manteniéndose iguales todas las posiciones relativas. Pero nosotros no necesitamos realizar ningún tipo de observaciones para saber que esto *en realidad* no sucedió: por hipótesis, la otra ubicación es contrafáctica. No obstante, hay una cierta inobservabilidad inherente en la teoría newtoniana y podemos aprender muchas cosas mediante la consideración de cómo esos hechos inobservables pueden eliminarse con la modificación de la ontología del espacio y el tiempo.

[1] G. Galilei, *Dialogue Concerning the Two Chief World Systems*, University of California Press, Berkeley, 1967, pp. 145-149.

[2] I. Newton, *op. cit.*, p. 11. [Versión en español: *op. cit.*, pp. 131-132.]

[3] El pasaje donde Aristóteles expone esta doctrina del lugar —y por lo tanto del cambio de lugar— se encuentra en *Física*, IV.4.212.20. Descartes afirma lo mismo en *Sobre los principios de la filosofía*, parte 2, sec. 25.

[4] I. Newton, *op. cit.*, p. 12.

[5] Un análisis más detallado y sutil del "Escolio" de Newton se encuentra en R. Rynasiewicz, "By Their Properties, Causes and Effects: Newton's Scholium on Time, Space, Place and Motion. Part I: The Text", *Studies in History and Philosophy of Science*, 26: 133-153, 1995; así como en "By Their Properties, Causes and Effects: Newton's Scholium on Time, Space, Place and Motion. Part II: The Context", *Studies in History and Philosophy of Science*, 26: 295-321, 1995.

[6] Acaso valga la pena señalar que la descripción del argumento newtoniano del espacio absoluto que se expone en algunas obras de la literatura filosófica no es totalmente rigurosa. L. Sklar dice en su clásica obra, *Space, Time, and Spacetime*, University of California Press, Berkeley, 1977, pp. 183 y 184, que "las aceleraciones provocan fuerzas *observables*", las cuales él llama "fuerzas inerciales", añadiendo que Newton propone el movimiento absoluto con el fin de explicar tales fuerzas. Entre estas fuerzas "inerciales" se encuentran, según Sklar, las "fuerzas centrífugas a través del universo astronómico (esas mismas fuerzas 'impiden que los planetas caigan en el Sol')". Pero Newton no postula tal cosa como las fuerzas "inerciales": sólo existen las fuerzas físicas, como la fuerza que se debe a la gravedad. Y las aceleraciones no causan fuerzas: las fuerzas causan aceleraciones (cambios en el movimiento). No existe una fuerza centrífuga (que huye del centro) que "impide que los planetas caigan en el Sol": únicamente existe una fuerza centrípeta que impide que los planetas *se alejen* del Sol. Ésta es la fuerza de la gravedad.

En los textos de física, las "fuerzas inerciales" se describen correctamente como fuerzas "ficticias"; es decir, no existen tales fuerzas. Uno puede engañarse y pensar que tales fuerzas son reales cuando utiliza inadecuadamente un "marco referencial no inercial". Lo que *esto* significa con exactitud será examinado enseguida.

[7] He dejado de lado la tercera ley del movimiento, la cual asevera que por cada acción de un cuerpo sobre otro cuerpo se da una acción igual y opuesta del segundo sobre el primero, porque no es esencial en nuestro discurso. Uno podría argumentar que esta tercera ley se sigue de las dos primeras: si se violara, un sistema compuesto de dos cuerpos podría acelerarse espontáneamente.

[8] A. Einstein, "Autobiographical Notes", en P. A. Schilpp (comp.), *Albert Einstein: Philosopher-Scientist*, vol. 1, Open Court, LaSalle, 1982, p. 67. [Versión en español: *Notas autobiográficas*, trad. de M. Paredes, Alianza, Madrid, 2003, p. 68.]

[9] H. Goldstein, *Classical Mechanics*, 2ª ed., Addison-Wesley, Reading (Mass.), 1981, pp. 1-2.

[10] Para discusiones más detalladas de la historia del concepto del marco inercial, véanse R. DiSalle, "Conventionalism and the Origins of the Inertial Frame Concept", en A. Fine, M. Forbes y L. Wessels (comps.), *PAS 1990*, vol. 2, Philosophy of Science Association, East Lansing, 1991; J. Barbour, *The Discovery of Dynamics*, Oxford University Press, Nueva York, 2001, cap. 12, y H. Brown, *Physical Relativity*, Oxford University Press, Oxford, 2006, cap. 2.

[11] Segunda carta, § 1, *apud* H. G. Alexander (comp.), *The Leibniz-Clarke Correspondence*, Barnes and Noble, Nueva York, 1956. [Versión en español: G. W. Leibniz, *La polémica Leibniz-Clarke*, trad. de E. Rada, Taurus, Madrid, 1980, p. 57.]

[12] Segunda respuesta, § 1. [Versión en español: *ibid.*, p. 62.]

[13] Tercera carta, § 5. [Versión en español: *ibid.*, p. 68.]

[14] Cuarta carta, § 19.

[15] Tercera carta, § 5. [Versión en español: *ibid.*, p. 68.]

[16] Leibniz no es coherente respecto al estatus exacto del PII. En el pasaje citado arriba, es claro que supuestamente se aplica a las dos (supuestamente distintas) posibilidades de ubicar la materia en el universo, y su consecuencia, el hecho de que no puedan ser posibilidades distintas. En la quinta carta de Leibniz, el principio es sin duda menos sólido: allí Leibniz dice que los objetos distintos indiscernibles, incluso en un mismo mundo, son

posibles, pero que Dios no los produce en virtud de la existencia del PRS. Si Leibniz deriva el PII del PRS, entonces el primero falla cuando rechazamos el segundo, de manera que sólo las justificaciones independientes del PII nos deben preocupar. Un análisis de los argumentos de Leibniz en la *Correspondencia* se encuentra en Gonzalo Rodriguez-Pereyra, "Leibniz's Argument for the Identity of Indiscernibles in His Correspondence with Clarke", *Australasian Journal of Philosophy*, 77 (4): 429-438, 1999.

[17] Cuarta carta, § 6.

[18] Quinta carta, § 53. [Versión en español: *ibid*., p. 116.]

[19] J. Barbour y B. Bertotti, "Mach's Principle and the Structure of Dynamical Theories", *Proceedings of the Royal Society A*, 382 (1783): 295-306, 1982. Véanse también las obras de J. Barbour, *The End of Time*, Oxford University Press, Nueva York, 2000, y *The Discovery of Dynamics*, Oxford University Press, Nueva York, 2001.

III. ELIMINACIÓN DE LA ESTRUCTURA INOBSERVABLE

VELOCIDAD ABSOLUTA Y RELATIVIDAD GALILEANA

En el capítulo precedente nos enfocamos en el primer argumento de Clarke y su hábil refutación por Leibniz. Clarke decía que los hechos relacionados con el espacio tenían que transcender las solas relaciones espaciales entre los cuerpos, puesto que hay cuerpos con las mismas relaciones espaciales que pueden situarse de una manera diferente en el espacio absoluto; Leibniz respondió que no es posible que el espacio absoluto exista, puesto que *1*) Dios no podía optar entre las diferentes posibilidades mencionadas por Clarke, si es que existieran (el argumento del principio de la razón suficiente), y *2*) las diferentes posibilidades, siendo cualitativamente idénticas, no pueden ser verdaderamente diferentes (el argumento del principio de la identidad de los indiscernibles). En cada uno de los casos, las dos supuestas posibilidades únicamente difieren con respecto a la *ubicación* de los cuerpos en E^3. Pero Clarke no había terminado con sus argumentos. Cuando Leibniz negó que hubiera cualquier tipo de verdadera diferencia física entre ambas situaciones, Clarke propuso otro ejemplo donde las dos distintas situaciones físicas generan las mismas ubicaciones relativas de los cuerpos materiales:

> Si el espacio no fuera sino el orden de los coexistentes, se seguiría que, si Dios desviara todo el mundo material con una velocidad cualquiera en línea recta, permanecería sin embargo, siempre inmóvil en el mismo lugar. Y también se seguiría que nada colisionaría con el parón en seco de ese movimiento.[1]

Mientras que el primero de los ejemplos de Clarke concierne a la *ubicación* de la materia en el espacio absoluto, el segundo ejemplo concierne, por su parte, a la *velocidad absoluta* de la materia en el espacio. La idea es la misma desde una perspectiva: si las dos situaciones pueden ser físicamente diferentes, aunque concuerden en la cuestión de las relaciones espaciales de todos los cuerpos entre sí, entonces la física tiene que consistir en algo más que sólo en esas relaciones. Pero si se examina con detalle, la situación cambia drásticamente.

La respuesta de Leibniz a este argumento se parece mucho a la primera, puesto que amalgama el PRS y el PII, aun cuando ambos principios se aplican de modo diferente:

> Decir que Dios hiciera avanzar todo el universo, en línea recta, o de otra forma, sin cambiar nada, es también una suposición quimérica. Pues dos estados indiscernibles son un mismo estado y, en consecuencia, es un cambio que no cambia nada. Es más, no hay ni rima ni razón, pues Dios no hace nada sin razón, y es imposible que la haya aquí. Por otro lado, sería *agendo nihil agere* [actuar sin hacer nada], como acabo de decir, a causa de la indiscernibilidad.[2]

Las dos líneas defensivas contra Clarke son *1*) que realmente no hay dos posibilidades distintas en el caso, una en que el mundo material se encuentre en reposo y la otra en que se mueva uniformemente en línea recta (argumento del PII), y *2*) que si hubiera dos posibilidades distintas, *per impossibile* [por imposible que fuera], Dios no tendría motivos para optar por una de ellas (argumento del PRS). Pero, al reflexionar en el problema, el argumento del PRS no puede despegar en el caso de que hubiera una velocidad absoluta. Supongamos que Dios quisiera crear el mundo material en un espacio absoluto que hasta entonces hubiera estado vacío. Dios le podría dar al mundo material, como un todo, cualquier velocidad absoluta dentro de ese espacio sin que tal cosa afectara las posiciones y los movimientos relativos de los cuerpos. Entre todas estas velocidades absolutas posibles, una se destaca en especial: el reposo absoluto. Pues si Dios le diera al mundo cualquier tipo de velocidad absoluta que no fuera nula, tendría que escoger una *dirección* en el espacio absoluto a donde apuntar esa velocidad. Y la isotropía de E^3 implica que todas las posibles direcciones distintas son cualitativamente iguales: no podría haber una razón para optar por una en particular. La *única* velocidad absoluta que no requiere que se opte por una de las direcciones es la velocidad cero: el reposo absoluto. De modo y manera que el PRS, lejos de prohibir cualquier opción de una velocidad absoluta para el mundo material, exige que se opte por precisamente una de ellas.

En todo caso, como hemos visto, el argumento del PRS se apoya en una especie de teología dudosa, y por lo tanto no lo tendremos en cuenta.

Por otro lado, del argumento del PII pueden derivarse algunas reflexiones interesantes. En primer lugar, tanto Leibniz como Clarke están de acuerdo en que no sería posible detectar mediante la observación el movimiento uniforme en línea recta de un sistema material a través del espacio absoluto. Clarke (y Newton) negarían que esto fuere así en el caso de la *rotación* absoluta de un sistema; es éste el sentido del argumento de la cubeta. Pero las partes de un sistema en movimiento absoluto uniforme a lo largo de una línea recta muestran los mismos movimientos relativos, y por ende el mismo comportamiento observable, que mostrarían en el estado de reposo absoluto. Galileo defendió con brillantez esta idea, la cual no es intuitivamente obvia.

Al defender el sistema de Copérnico, Galileo tuvo que explicar por qué el movimiento tremendamente veloz de la superficie de la Tierra al girar en torno a su eje no sería evidente para la observación casual. Aquí es necesario tener en cuenta hasta qué grado es veloz este movimiento, según Copérnico. La circunferencia de la Tierra tiene más de 38 000 kilómetros, de manera que en el ecuador la superficie de la Tierra se mueve a más de 1 600 kilómetros por hora debido a la rotación. Y la órbita de la Tierra en torno al Sol es de cerca de 941 millones de kilómetros, de manera que viaja casi 2.5 millones de kilómetros al día en torno a él. Éstas son velocidades casi incomprensibles. ¿Cómo es posible que no *notemos* que nos movemos tan velozmente?

Galileo señaló que en ciertas circunstancias el movimiento absoluto de un sistema no afecta el comportamiento de las cosas que es posible observar dentro del sistema. En su ejemplo, utiliza los barcos:

Y aquí, para dejar clara definitivamente la nulidad de todas las experiencias aducidas, me parece que es el

> lugar y el momento adecuado para mostrar el modo de experimentarlas todas facilísimamente. Encerraos con algún amigo en la mayor estancia que esté bajo cubierta de algún gran navío, y meted en ella moscas, mariposas y animalillos voladores parecidos. Haya también un recipiente grande de agua con pececillos dentro. Además manténgase en alto un cubo, que gota a gota vaya dejando caer el agua en otro recipiente de boca estrecha, situado debajo. Cuando la nave está quieta, observad atentamente que los animalillos volantes se mueven en todas las direcciones de la cabina con igual velocidad. Veréis que los peces nadan indistintamente hacia todos los lados. Las gotas que caen entrarán todas en la vasija situada debajo. Y al lanzar algo a un amigo, no necesitas arrojarlo con mayor fuerza en una dirección que en otra si las distancias son iguales. Y si saltáis, como suele decirse, con los pies juntos, os desplazaréis igual espacio con independencia de la dirección. Una vez que hayáis observado diligentemente todas estas cosas (aunque no haya ninguna duda de que mientras el bajel está parado tienen que suceder así), haced mover la nave con la velocidad que sea. Veréis que (con tal que el movimiento sea uniforme y no fluctuante hacia aquí y hacia allá) no observaréis el más mínimo cambio en ninguno de los efectos mencionados y que, a partir de ellos, no podréis determinar si la nave avanza o está quieta. Al saltar, os desplazaréis en el entablado los mismos espacios que antes y no se dará el caso de que, porque la nave se mueva velocísimamente, daréis mayores saltos hacia popa que hacia proa aunque en el tiempo que estáis en el aire el entablado que está debajo de vos se desplace hacia la parte contraria a vuestro salto. Y al lanzar alguna cosa al compañero, no necesitaréis tirarla con más fuerza para que llegue, si él está hacia la proa y vos hacia la popa, que si estuvieran al revés. Las gotas caerán como antes en el vaso inferior, sin que ni siquiera una caiga hacia popa, por más que, mientras la gota está en el aire, la nave se desplace muchos palmos. Los peces en su agua no requerirán mayor esfuerzo para nadar hacia la parte delantera del recipiente que hacia la parte posterior, sino que llegarán con igual facilidad a la comida puesta sobre cualquier lugar del borde del recipiente. Finalmente, las mariposas y las moscas continuarán su vuelo indistintamente hacia cualquier lado, y en ningún caso sucederá que se queden hacia la pared que mira a popa como si estuvieran fatigadas de seguir la veloz carrera de la nave, de la que habrán estado separadas por mucho tiempo al mantenerse en el aire.[3]

La afirmación de que todos los experimentos que se llevaran a cabo en ambos estados de movimiento del barco tendrían el mismo resultado observable se denomina *relatividad galileana*.

El análisis físico completo del experimento de Galileo revelaría una tremenda sutileza, como se verá cuando lleguemos a la teoría de la relatividad. Pero por ahora solamente tenemos que conceder que los experimentos desarrollados en esas dos circunstancias son en un sentido obvio indiscernibles desde el punto de vista observacional: no sería posible demostrar experimentalmente dentro de la cabina debajo de la cubierta en cuál estado de movimiento se encuentra el barco. Galileo, al igual que Newton y que Clarke, considera como un hecho obvio que la nave se encuentra en un *estado diferente de movimiento* cada vez que los experimentos se llevan a cabo, pero los resultados no son discernibles porque las posiciones *relativas* y los movimientos *relativos* de todos los cuerpos son las mismas y los mismos en ambos casos. Basándose en el planteamiento de que los estados de movimiento son diferentes, Clarke utiliza el ejemplo para socavar la afirmación leibniziana de que todo movimiento se constituye por el movimiento relativo de los cuerpos.

Es claro que Leibniz invoca el argumento del PII para afirmar que en realidad no constituyen dos estados diferentes. O más bien, Leibniz concedería que en el ejemplo de Galileo la nave se encuentra en diferentes estados de movimiento *en relación con la costa*; pero Clarke exige teóricamente algo por completo distinto: que toda la materia del universo sea puesta en movimiento uniforme, con el fin de mostrar que no existe el movimiento relativo en absoluto. Esta idea, según Leibniz, es una tontería.

Newton conocía la relatividad galileana. De hecho, lo demuestra casi de inmediato en los *Principia* en el corolario v de las leyes del movimiento, al suponer que todas las fuerzas son fuerzas de impulso debido a las colisiones entre las fuerzas y que la magnitud de estos impulsos únicamente depende de la velocidad relativa de los cuerpos:

> *Los movimientos entre sí de los cuerpos incluidos en un determinado espacio son los mismos, ya esté dicho espacio en reposo, ya se mueva recta y uniformemente sin movimiento circular.*
>
> Pues las diferencias de los movimientos tendentes a un lado y las sumas de las tendentes al lado contrario, son las mismas desde el principio en ambos casos (por hipótesis) y de tales sumas y diferencias proceden los choques y fuerzas con los que los cuerpos mutuamente se afectan. Por tanto, en virtud de la segunda ley, serán iguales los efectos de los choques en ambos casos; y por tanto los movimientos de los cuerpos entre sí en un caso permanecerán iguales a los movimientos entre sí en el otro. Esto se comprueba con un experimento clarísimo: todos los movimientos se comportan de modo igual entre sí en una nave tanto si se halla en reposo como si se halla en movimiento uniforme y rectilíneo.[4]

Ya hemos impugnado el argumento del PII en cuanto principio fundamental ontológico o metafisico, aunque concediendo que la postulación de hechos físicos inobservables debe hacernos reflexionar. En el primer argumento de Clarke, no había al parecer ninguna forma de argumentar que existiera cualquier tipo de pregunta física respecto al estado del mundo que no fuera posible determinar mediante la observación. Si preguntáramos si el universo material se orienta o no en el espacio absoluto tal como se encuentra, es decir, con la parte oriental en el este y la parte occidental en el oeste, o más bien “de manera contraria”, entonces la respuesta es obvia: el universo se encuentra ubicado en donde está ubicado. Newton y Clarke admiten que hay una posibilidad *contrafáctica*, en la cual todo el comportamiento observable fuera el mismo, pero su formulación misma nos da a entender que la situación planteada es contrafáctica y no real.

Pero este segundo argumento recurre al movimiento absoluto, más que a la ubicación en el espacio absoluto, y es bastante diferente. Pues ahora podemos formular preguntas físicas respecto al universo material para las cuales, según Newton, existen determinadas respuestas que no podríamos descubrir experimentalmente. Por ejemplo, según Newton, ahora mismo usted se encuentra en una determinada velocidad absoluta y en una determinada dirección. Usted podría encontrarse en reposo absoluto (aunque por la rotación de la Tierra tal cosa sólo podría durar un instante) o podría estar moviéndose con una velocidad de millones de kilómetros por hora a través del espacio absoluto en dirección de Alfa Centauro. A diferencia del caso de la ubicación, no podemos saber cuál de estas dos descripciones del movimiento absoluto de usted es la correcta y, según el corolario v, ningún experimento nos podría ayudar a saberlo.

Así, la teoría de Newton del espacio y tiempo absolutos, junto con sus leyes del movimiento y el planteamiento de que todas las fuerzas son fuerzas de impulso causadas por colisiones, lo obligan a proponer la existencia de hechos físicos experimentalmente indeterminables. Esto nos debería hacer reflexionar. ¿Hay alguna forma de purgar, por así decirlo, la teoría de Newton para limpiarla de esas velocidades absolutas inobservables, pero sin despojarla de todo su gran poder explicativo?

No se ve de buenas a primeras cómo realizar la empresa depuradora. Las observaciones de Galileo, y el corolario de Newton, no son en nada parecidas a la propuesta leibniziana en la que todos los movimientos son relativos. De hecho, tanto en el ejemplo de Galileo como en el de Newton se plantea que las naves se hallan ya sea "en reposo" o ya sea "moviéndose" hacia adelante en línea recta sin un "movimiento circular" y "sin fluctuar de un lado a otro". Según Newton, éstas son caracterizaciones del movimiento *absoluto* de las naves. Y la relatividad galileana únicamente se aplica a semejantes naves. No es cierto, por ejemplo, que un experimento que se realice en un laboratorio por lo general produzca el mismo resultado cuando tenga lugar en otro laboratorio que se encuentre en movimiento uniforme con relación al primero. De hecho, ni siquiera es verdad que un experimento por lo general producirá el mismo resultado que otro que se ha realizado en un laboratorio *en reposo* con relación al primero. Por ejemplo, si yo estoy llevando a cabo un experimento en una rueda de la fortuna que está girando, observando si unas gotas de agua caen o no de una botella a un recipiente que se halla justamente debajo, comprobaré que ése es el caso cuando la botella se encuentra en el eje de rotación, pero no cuando se encuentra hacia el borde de la rueda de la fortuna. Por lo tanto, el principio no se puede afirmar simplemente en función de los movimientos relativos de los laboratorios: cada uno de los laboratorios debe estar en movimiento *inercial* o debe constituir un *marco de referencia inercial*.

Para cualquier estudiante de física resulta familiar la frase "marco de referencia inercial". ¿Pero qué significa exactamente esa frase? Newton podría dar la explicación: un marco de referencia inercial es un cuerpo perceptible —en relación con el cual se observan las posiciones y los movimientos de otros cuerpos— *que se encuentra en reposo absoluto o en movimiento uniforme sin rotación a lo largo de una línea recta a través del espacio absoluto*. Esta explicación obviamente en nada nos ayudaría a eliminar el espacio absoluto en la física. De manera que si explicáramos la relatividad galileana como el planteamiento de que "todos los marcos de referencia son iguales" y luego definiéramos un marco inercial de este modo, nos basaríamos en el espacio y el movimiento absolutos al igual que Newton.

En este punto nos encontramos ante un pequeño problema. La explicación de Newton del espacio y el tiempo lo obliga a basarse en hechos sobre el movimiento de los cuerpos que no pueden determinarse experimentalmente. Pero también el hecho de que un cuerpo se encuentre o no en rotación —en cierto sentido absoluto— al parecer tiene consecuencias observables. Nos gustaría hallar algún modo de eliminar los movimientos absolutos y conservar las rotaciones absolutas, lo cual parece ser una contradicción. Pero resulta que es posible realizar el truco. Sin embargo, tenemos que modificar radicalmente nuestra manera de entender el espacio y el tiempo.

EL ESPACIO-TIEMPO GALILEANO

Hagamos un rápido repaso. Newton teoriza que el espacio tiene la estructura de E^3 y que el tiempo posee una estructura métrica unidimensional. Y, lo que es más importante, plantea que las partes individuales del espacio absoluto persisten a través del tiempo, de manera que pueda haber un hecho sobre si un cuerpo se encuentra o no en un momento dado en *el mismo lugar precisamente* en que estaba en otro momento dado. Si es así, el cuerpo se encuentra en reposo absoluto. Si un cuerpo no se encuentra en reposo absoluto, tiene que haber un hecho en cada momento tanto sobre la velocidad con que viaja como sobre la dirección en que viaja. Todos los cuerpos tienen una *trayectoria a través del espacio absoluto*: los puntos en el E^3 persistente que ocupan a lo largo de un intervalo de tiempo. Esta trayectoria espacial tiene una forma: podría ser una línea recta, por ejemplo, o una elipse. Además hay un hecho sobre la rapidez con que un cuerpo pasa a través de una parte dada de su trayectoria espacial.

Fue en función de esta ontología que Newton pudo definir el reposo absoluto y el movimiento absoluto uniforme en línea recta, y por lo tanto formular su primera ley del movimiento. Si pretendemos conservar de alguna manera la dinámica de Newton, aunque rechazando su explicación del espacio y el tiempo, necesitamos algún otro tipo de estructura sobre cuya base pueda formularse la primera ley. Nuestra ruta hasta esta estructura es a través del espaciotiempo.

Empezaremos con una *representación* de la ontología newtoniana. Será más conveniente desde el punto de vista gráfico describir un espacio absoluto newtoniano que sea de dos dimensiones en vez de ser tridimensional. Supongamos que sólo hay dos cuerpos, A y B, en este espacio. En cualquier momento dado del tiempo, los dos cuerpos se encuentran ubicados en ciertos puntos en particular del espacio absoluto. Al transcurrir el tiempo, los cuerpos pueden cambiar su ubicación en el espacio absoluto. Por ejemplo, podrían chocar uno con el otro y rebotar. La figura III.1 representa los cuerpos en cuatro momentos del tiempo distintos. Además, se indican en las cuatro fases dos puntos del espacio absoluto, marcados como p y q.

La primera cosa que haremos es apilar estas representaciones de nuestro espacio bidimensional. Puesto que hay infinitos momentos de tiempo absoluto, también debería haber infinitas "hojas" en la pila, cada una de ellas la representación del estado del mundo en ese momento. Seguiremos la pista a nuestras cuatro imágenes. Además, utilizaremos dos obvios, pero esenciales, métodos al hacer la pila de nuestros "retratos" de los momentos en el tiempo. Primero, haremos la pila de manera que cada punto en el espacio absoluto forme una línea vertical a través del tiempo. Las trayectorias del punto p y del punto q se indican. Segundo, hacemos la pila de manera tal que la distancia vertical entre cualquier par de hojas sea proporcional al tiempo absoluto que ha transcurrido. De modo que si el tiempo que ha transcurrido entre $t = 0$, $t = 1$, $t = 2$ y $t = 3$ es siempre el mismo, las cuatro hojas deberán estar ubicadas en intervalos de espacio iguales. El resultado es la representación *espacial* tridimensional de nuestro espacio de dos dimensiones, tal como persiste a través del tiempo. Ya que cada uno de nuestros cuatro cuerpos también existe en cada momento, podemos indicar sus posiciones, formando las *líneas de mundo* de estos cuerpos (figura III.2).

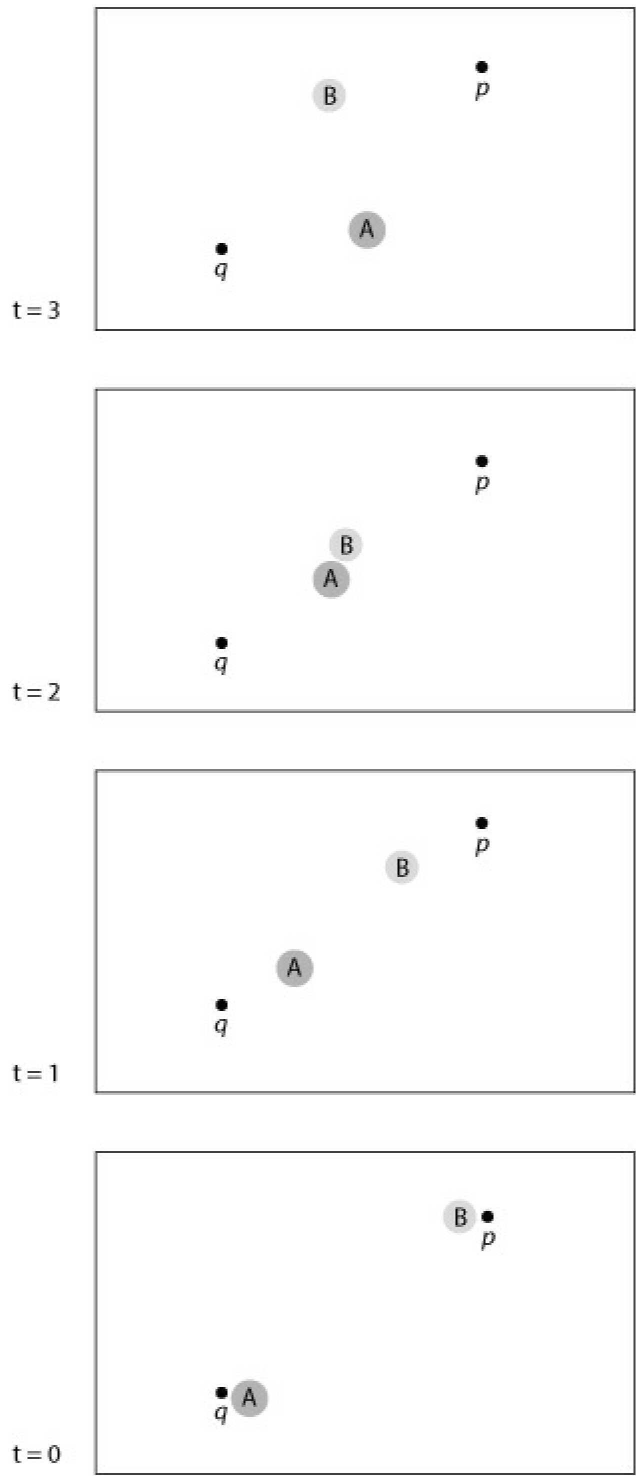

FIGURA III.1

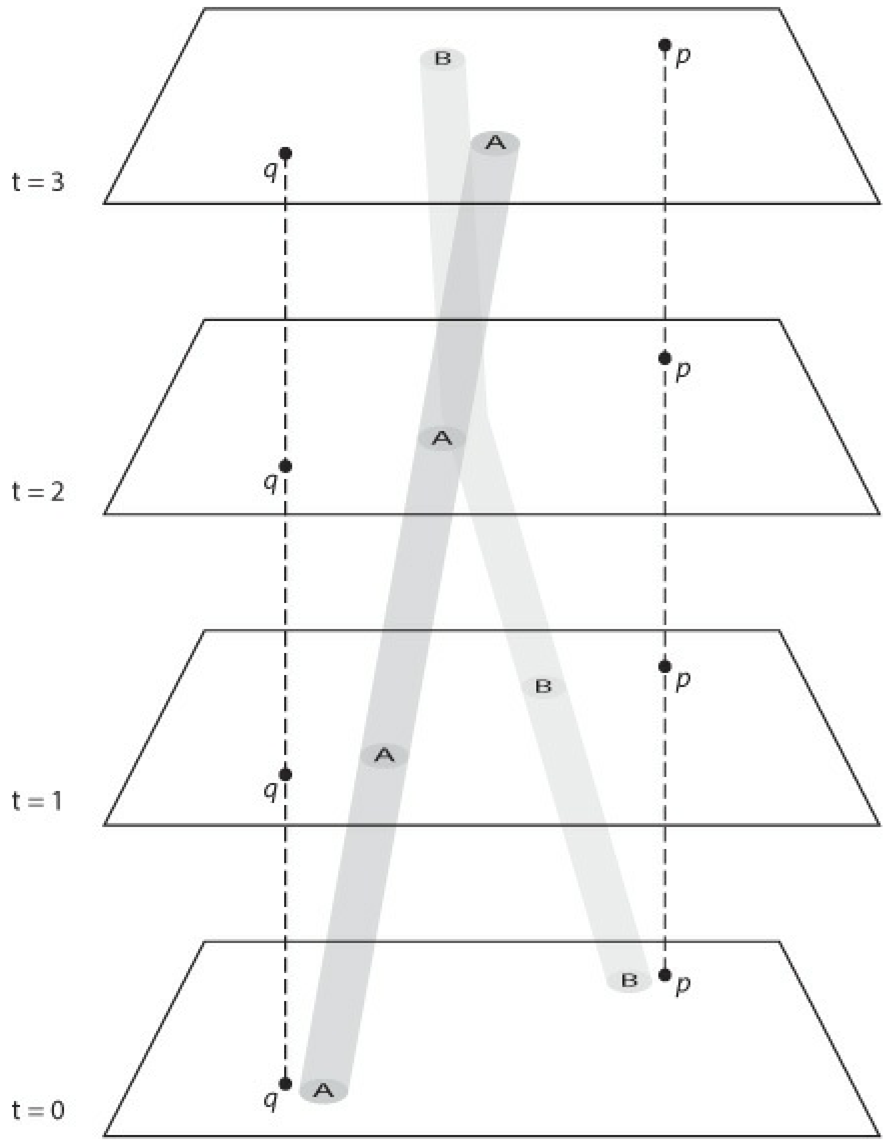

FIGURA III.2

Hasta aquí no hemos modificado en nada la explicación newtoniana del espacio y el tiempo absolutos: simplemente hemos producido una mera *representación* espacial de la estructura espacial y temporal que Newton teoriza. Este tipo de representación del espacio y el tiempo juntos se denomina *diagrama de espacio-tiempo*. Es esencial observar con atención el rol que tanto el espacio absoluto como el tiempo absoluto han jugado en la producción del diagrama: únicamente porque los puntos del espacio absoluto persisten a través del tiempo pudimos especificar la manera de apilar las hojas del tiempo, y sólo porque el tiempo absoluto tiene una estructura métrica pudimos especificar las distancias espaciales entre las diversas hojas. En este diagrama los objetos en reposo absoluto se representan por líneas verticales, puesto que permanecen en el mismo punto del espacio absoluto. Pero lo que es mucho más importante, *las trayectorias de los cuerpos en movimiento uniforme absoluto a lo largo de una línea recta se representan por líneas rectas en el diagrama de espacio-tiempo*.

Hagamos otro diagrama. En un espacio completamente vacío colocamos dos pares de globos unidos cada uno por una cuerda. Un par se encuentra en reposo absoluto, por lo que no hay tensión en la cuerda. El otro par se encuentra en rotación absoluta en torno al centro de su masa, y la cuerda está tensa. La figura III.3 es el diagrama de espacio-tiempo de la situación.

Ya que los globos en la parte derecha de la figura III.3 se encuentran en reposo absoluto, sus trayectorias se representan por líneas verticales en el diagrama. Los globos giratorios, al contrario, cambian constantemente su estado de movimiento absoluto. Esto se indica en el diagrama como una curvatura de sus líneas de mundo. La velocidad absoluta de un globo en cualquier momento dado se corresponde con la tangente a la hélice en ese punto. Las velocidades absolutas de uno de los globos en los puntos *r* y *s* están representadas por las dos tangentes que se muestran en el diagrama. La velocidad absoluta de un objeto se indica por el ángulo de esta tangente: una tangente vertical se corresponde con el reposo absoluto y mientras más inclinada esté la tangente respecto a la vertical, mayor será la velocidad absoluta del cuerpo. (La velocidad absoluta se representa por el ángulo y la orientación de la tangente; la “longitud” de la tangente no es significativa.) La aceleración de los globos en la parte izquierda se representa por el cambio constante en estas tangentes al pasar el tiempo. La modificación continua del movimiento absoluto de los globos se debe a la tensión de la cuerda, en concordancia con la segunda ley de Newton.

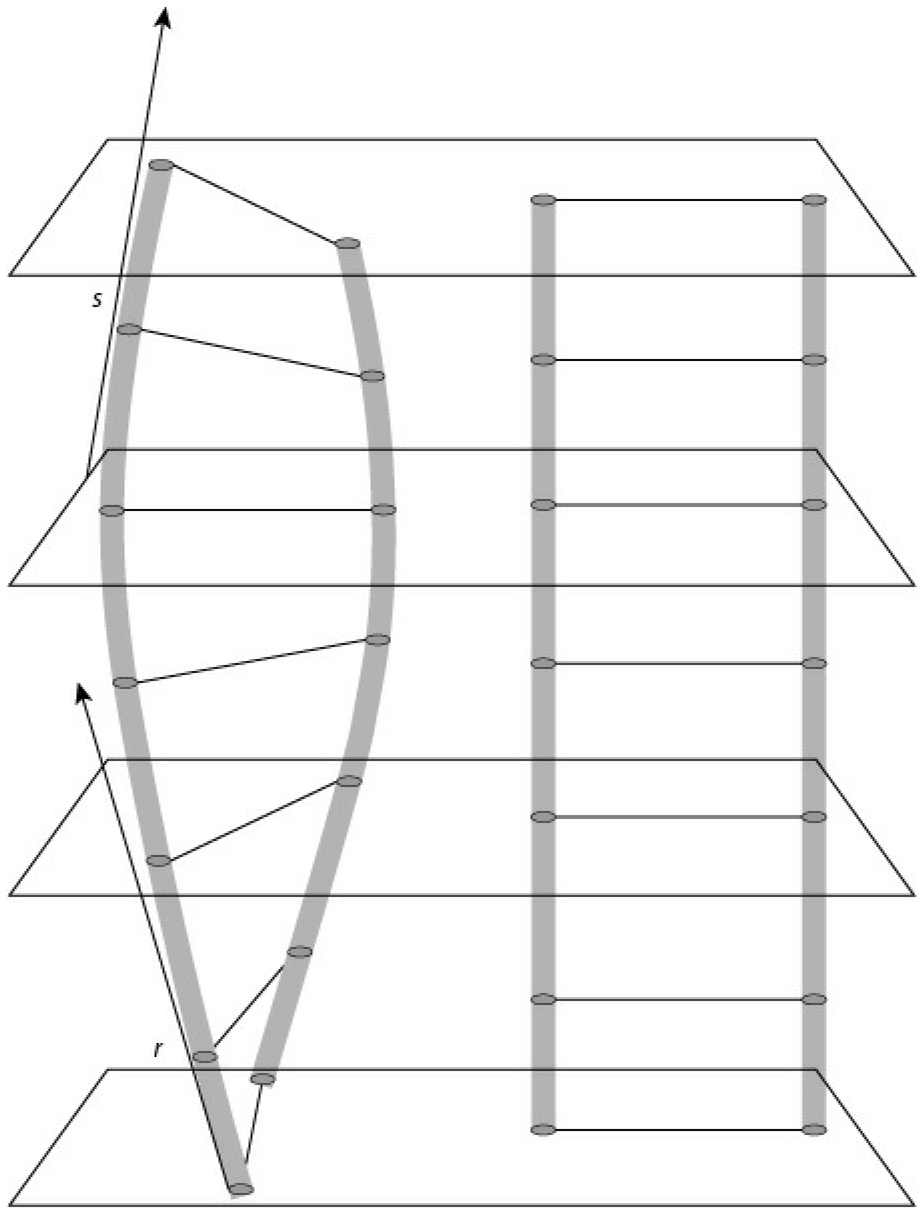

FIGURA III.3

En este diagrama de espacio-tiempo, *los movimientos no acelerados se representan por líneas de mundo rectas*. Esto es válido tanto para los cuerpos en reposo absoluto como para los cuerpos que se mueven uniformemente en línea recta. Así, en función de este tipo de diagrama, podemos reformular la primera ley de Newton en la siguiente forma:

> Primera ley (versión del diagrama de espacio-tiempo): La trayectoria de un cuerpo se representa por una línea recta en el diagrama de espacio-tiempo, excepto en la medida en que al cuerpo se le obligue a cambiar su estado mediante fuerzas impresas.

Es de notar que en nuestra nueva versión no se distingue entre los cuerpos en reposo absoluto y los cuerpos en movimiento uniforme. Ambos son movimiento*s inerciales*, es decir, los movimientos de un cuerpo que no experimente fuerzas externas. En nuestra nueva versión de la primera ley los movimientos inerciales se especifican como precisamente los movimientos representados por líneas rectas en el diagrama de espacio-tiempo.

La segunda ley de Newton, en esta reformulación, afirmaría que el efecto de una fuerza impresa se representa por una *curvatura o doblamiento* de la línea mundial en el diagrama. La línea mundial se hará curva en la dirección de la fuerza impresa, y la dimensión de la curvatura será proporcional tanto a la magnitud de la fuerza impresa como a la masa del cuerpo. La masa de un cuerpo, en particular la masa inercial de un cuerpo, no se menciona en la primera ley porque no juega ningún rol en el comportamiento de un objeto libre de fuerzas externas. En esencia, la masa inercial de un objeto es una medida de la resistencia al doblamiento de sus líneas de mundo por las fuerzas impresas: mientras más grande sea su masa, menos se curvará la trayectoria. Así, esquemáticamente, en vez de representar las leyes de Newton utilizando

$$\mathbf{F} = \mathrm{m}\mathbf{A},$$

podemos expresar la dinámica como

$$\mathbf{F} = \mathrm{m}\mathbf{Curva},$$

donde **Curva** es una medida de la curvatura de la línea mundial. Al imprimir una fuerza en un objeto, su línea mundial se curva en la dirección de la fuerza, en proporción directa a la fuerza y en proporción inversa a la masa inercial.[5]

El punto esencial en estas reformulaciones de las leyes de Newton es que en ellas ha desaparecido toda mención de la velocidad absoluta de un cuerpo cualquiera. Al aplicar la dinámica de Newton, no es necesario que hagamos ninguna distinción entre el reposo absoluto y el movimiento absoluto uniforme: es sólo necesario que distingamos entre las trayectorias *rectas* y las trayectorias *curvadas* en el espacio-tiempo. En la medida en que podamos identificar las líneas rectas en nuestro diagrama de espacio-tiempo y cuantificar la curvatura de las líneas curvadas, podremos determinar lo que las leyes de Newton predicen en cualquier caso dado. *No necesitamos saber cuáles son las trayectorias de los puntos individuales del espacio absoluto, como se indica en la figura* III.2.

Y lo que no necesitamos saber no lo necesitamos para nada. Es decir, aunque Newton afirma que los puntos individuales del espacio absoluto persisten a través del tiempo, hemos sido capaces de reformular sus leyes sin referirnos a tal persistencia o a la velocidad absoluta de ningún objeto. Tenemos la libertad de rechazar la explicación newtoniana del espacio y el tiempo en favor de una nueva: el espaciotiempo galileano. Claro es que no podemos rechazar la persistencia de los puntos del espacio a través del tiempo sin pagar el precio en otro lugar: tenemos que hacer algo más que eliminar esa pequeña porción de metafísica. Para introducir la estructura geométrica que requerimos, nuestra comprensión total de la ontología del espacio y el tiempo tiene que cambiar.

Newton propone dos tipos de cosas distintos: el espacio absoluto, que posee una cantidad de puntos infinita, y el tiempo absoluto, que posee una cantidad infinita de momentos o instantes. Puesto que los puntos del espacio absoluto persisten a través del tiempo, el mismo conjunto de puntos constituye el espacio en todo momento. Nuestra nueva ontología rechaza esta propuesta en favor de los *puntos del espacio-tiempo* o *eventos*. Un evento es esencialmente un lugar-en-el-espacio-en-un-momento-dado: la explosión de un cohete en particular tiene lugar en un punto del espacio-tiempo que ocurre una sola vez y que, idealmente, no tiene una extensión espacial y no toma tiempo. Aprender a pensar sobre el espacio y el tiempo en función de los eventos requiere de una cierta práctica, y para esto nuestros diagramas de espacio-tiempo pueden ser de gran ayuda. Un evento se corresponde con un solo punto del diagrama de espaciotiempo, ya sea que cualquier cosa notable esté sucediendo o no allí.

Una vez que hayamos aceptado los eventos como los elementos básicos de nuestra ontología espacio-temporal, nuestra tarea consistirá en especificar cuál es la estructura física de estos eventos: la geometría del espacio-tiempo. El espacio-tiempo galileano propone diferentes tipos de estructura.

Primero, hay un lapso de tiempo absoluto, objetivo, entre cualquier par de eventos. En este sentido, el espaciotiempo galileano adopta el tiempo absoluto newtoniano sin ningún cambio. Hay una métrica para el paso del tiempo, de manera que es posible definir adecuadamente, por ejemplo, si el tiempo que ha pasado entre un par de eventos es el mismo, o mayor, o menor que el tiempo que ha pasado entre cualquier otro par.

Dada esta estructura temporal absoluta, el espaciotiempo galileano puede seccionarse en conjuntos de eventos que ocurren todos en el mismo momento, es decir, eventos entre los cuales el paso del tiempo es nulo. Cada uno de estos conjuntos completos de eventos forma una *rebanada de simultaneidad* en el espacio-tiempo. Las figuras III.2 y III.3 indican estas rebanadas de simultaneidad en una forma obvia. El tiempo que ha pasado entre cualquier evento en la rebanada marcada $t = 0$ y cualquier evento en la rebanada marcada $t = 1$ en la figura II.1 es el mismo. Llamamos a este seccionamiento del espacio-tiempo en rebanadas de simultaneidad una *foliación* del espacio-tiempo. Así, la estructura del tiempo folia el espacio-tiempo galileano en la forma de rebanadas de simultaneidad y de una manera totalmente objetiva. En cualquier par de eventos, hay un hecho sobre cuál ocurre (o no) primero y cuánto tiempo transcurre entre ambos eventos.

Hemos explicado, pues, la estructura temporal. ¿Qué sucede respecto al espacio? El espacio-tiempo galileano plantea una geometría espacial particular *en cada una de las rebanadas de simultaneidad*. Específicamente, plantea que los eventos en cada una deben tener la estructura espacial de E^3. La geometría de cada rebanada tiene que especificarse independientemente de la de cualquier otra, porque los puntos del espacio no persisten a través del tiempo. Por lo tanto, la indicación de que los eventos en la rebanada $t = 0$ poseen la estructura de E^3 nada nos dice por sí sola sobre la geometría de los eventos en $t = 1$. Por lo tanto, la segunda parte de la estructura geométrica de nuestro espacio-tiempo es la geometría espacial de E^3 en cada rebanada individual.

Pero la simple propuesta del tiempo absoluto y E^3 en cada rebanada de simultaneidad no nos da la estructura suficiente como para formular las leyes del movimiento de Newton. Como hemos visto, las leyes newtonianas requieren de una porción de geometría *transtemporal:* una estructura geométrica entre eventos que ocurren en momentos *diferentes*. Para Newton mismo, la estructura transtemporal se aseguraba por la persistencia de los puntos del espacio absoluto a través del tiempo, pero nosotros estamos tratando de eliminar ese planteamiento. Para suplantarlo, necesitamos una topología y una estructura afín en nuestro nuevo espacio-tiempo. La estructura topológica determina cuáles conjuntos de eventos integran trayectorias continuas, ininterrumpidas, a través del tiempo. Y la estructura afín determina cuál de estas trayectorias continuas espacio-temporales son rectas. (Por una "trayectoria continua espacio-temporal" entendemos un conjunto continuo de eventos que no contenga más de un evento en cada momento.)

La idea de que haya una topología en el espacio-tiempo (es decir, el conjunto de todos los eventos) es tan natural e intuitiva que apenas nos percatamos de que integre un planteamiento teórico. Cuando diseñamos diagramas de espacio-tiempo, tenemos que dibujarlos en un espacio (la página) que tiene una topología, una estructura afín, y una métrica. Pero ahora estamos proponiendo una estructura para el espacio-tiempo mismo y no tenemos el derecho de suponer que tenga la misma geometría del diagrama. De hecho, pronto veremos que la estructura de los diagramas de espaciotiempo es *diferente* de la estructura del espacio-tiempo que supuestamente representan, por lo que tenemos que aprender a *ignorar* de forma sistemática un cierto aspecto del dibujo al interpretar los diagramas. Las trayectorias de los cuerpos a través del espacio-tiempo, sus líneas de mundo, sólo pueden caracterizarse como continuas o como discontinuas si el espacio-tiempo tiene una topología, de manera que teorizamos que en efecto la tiene. De hecho, la topología del espacio-tiempo se corresponde con la topología del *diagrama* de espacio-tiempo, de manera que las trayectorias continuas se representan con curvas continuas en el dibujo.

Pero además de la topología, el espacio-tiempo galileano contiene una estructura afín, por lo que es posible caracterizar las líneas de mundo ya sea como rectas o como curvadas. A esta estructura afín se le denomina *estructura inercial* del espacio-tiempo, por obvios motivos. Un cuerpo que no se encuentra sujeto a fuerzas externas está en *movimiento inercial* y por lo tanto tiene una *trayectoria inercial* a través del espacio-tiempo. En el espacio-tiempo galileano, las trayectorias inerciales únicamente están constituidas por las trayectorias rectas. Así, dada esta estructura del espaciotiempo es posible formular las leyes de Newton sin proponer la persistencia de los puntos a través del tiempo y por ende sin postular velocidades absolutas. Se soluciona el enigma de cómo puede existir la *aceleración* absoluta (por ejemplo, la rotación de los globos en la figura III.3) sin una *velocidad* absoluta: la aceleración absoluta es la curvatura de una línea mundial en el espacio-tiempo.

Indiqué anteriormente que la estructura geométrica de un *diagrama* de espacio-tiempo no es la misma que la estructura del espacio-tiempo que se pretende representar. Cuando dibujamos un diagrama, lo que hacemos es utilizar un tipo de objeto geométrico para representar otro objeto geométrico. Puesto que las geometrías de ambos objetos no son las mismas, debemos fijarnos con cuidado en cuáles aspectos del diagrama corresponden a los hechos físicos reales y cuáles son meramente convencionales. En el caso de la explicación del espacio y el tiempo que da el mismo Newton, hemos visto que el *ángulo* de una trayectoria en el diagrama es físicamente significativo: representa la *velocidad absoluta* de un cuerpo, mientras que los objetos en reposo absoluto ocupan trayectorias verticales. Se elige arbitrariamente en el diagrama cuáles trayectorias rectas se representarán en forma vertical y cuáles en forma "inclinada". Así, por ejemplo, la situación física real en el espacio-tiempo galileano que se representa en la figura III.3 podía haber sido representada con la misma precisión tal como está en la figura III.4.

La figura III.4 se relaciona con la III.3 gracias a una transformación afín, tal como la que se muestra en la figura I.1. En esta transformación del *diagrama* toda la estructura galileana permanece intacta: los tiempos entre los eventos son los mismos, las rebanadas de simultaneidad son las mismas, la estructura euclidiana de cada una de las rebanadas es la misma en todas, y el conjunto de trayectorias rectas en el espacio-tiempo es la misma. El "estado de movimiento" de un cuerpo en un momento aún se representa por la tangente a su trayectoria en ese momento, así que el estado de movimiento de la esfera en *r* todavía difiere de su estado de movimiento en *s*. Esta diferencia es atribuible, por la misma razón, a la tensión en la cuerda. De manera que si teorizamos que el espacio-tiempo físico tiene una estructura galileana, es posible utilizar las figuras III.3 y III.4 para ilustrar *exactamente la misma situación física*.

Al utilizar nuestros métodos convencionales para ilustrar el espacio-tiempo total newtoniano —la estructura que Newton propone—, las figuras III.3 y III.4 tendrían que ilustrar *diferentes* situaciones físicas: en la figura III.3 los globos del lado derecho se encuentran en reposo absoluto y en la figura III.4 se encuentran en movimiento, a una velocidad absoluta constante. Las *aceleraciones* de los cuerpos en ambas situaciones son las mismas, de modo que las *fuerzas* presentes son las mismas (existe la misma tensión en la cuerda), pero son situaciones físicamente diferentes para Newton. No lo son en el espacio-tiempo galileano.

Se presupone la estructura del espacio-tiempo galileana en todas las explicaciones estándar de la mecánica newtoniana, aun cuando esta presuposición usualmente se sobrentienda y no se señale. Por ejemplo, en la explicación de la mecánica newtoniana que Goldstein dio anteriormente, la presuposición aparece casi de inmediato, aunque jamás nos daríamos cuenta de ello. Goldstein empieza por decir: "Sea **r** el radio vector de una partícula desde un origen dado y **v** su vector velocidad… " ¿Pero *cuál* "origen dado"? Si elegimos *al azar* un nuevo origen para cada momento del tiempo, entonces el radio vector habrá de cambiar aleatoriamente, sin que importe cómo se comporte la partícula. Si lo que se quiere dar a entender es que siempre hay que utilizar *el mismo* origen, entonces la explicación presupone el espacio absoluto newtoniano que persiste a través del tiempo. Pero lo que realmente se supone es que la trayectoria del origen del sistema de coordenadas debe ser una trayectoria *inercial:* el origen espacial del sistema de coordenadas deberá trazar una línea recta a través del espacio-tiempo.

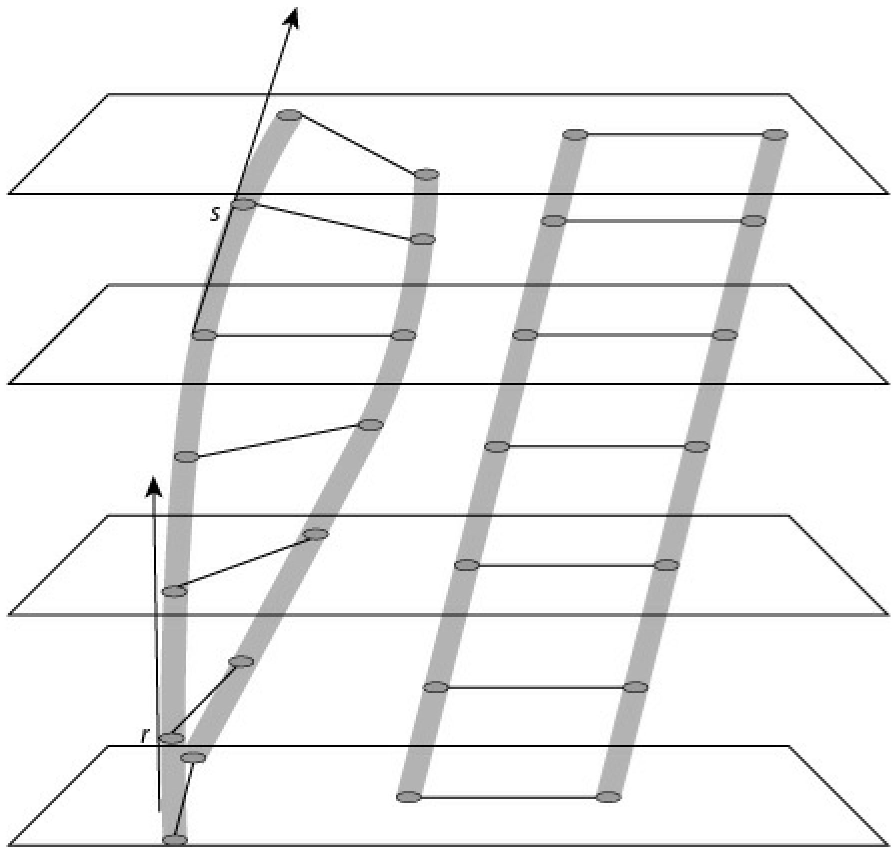

FIGURA III.4

De hecho, ya tenemos los recursos para definir, de manera general, lo que un *sistema de coordenadas inercial* es en la mecánica newtoniana. En el espacio-tiempo galileano, siendo el espacio E^3, un sistema de coordenadas asignará cuatro números a cada evento: tres "coordenadas espaciales" y una "coordenada temporal". Las diferencias en los valores de la coordenada temporal deberán ser proporcionales al tiempo que ha transcurrido entre los eventos. Por ende, a todos los eventos que tienen lugar en una sola rebanada de simultaneidad deberá asignarse la misma coordenada del tiempo. Y en *cada rebanada temporal*, las tres coordenadas espaciales deberán formar un sistema de coordenadas cartesiano en E^3. La distancia espacial, así como cualquier otra geometría puramente espacial, sólo se da entre los eventos que ocurran al mismo tiempo. Estos requerimientos son bastante obvios en un sistema de coordenadas. Pero se requiere de algo más para que un sistema sea un sistema de coordenadas *inercial:* las curvas coordenadas que se asocian con la variación temporal tienen que ser trayectorias inerciales. Supongamos que estamos utilizando las cuatro coordenadas (*t, x, y, z*). Fijemos los valores de las tres coordenadas espaciales y dejemos que la coordenada *t* varíe. Por ejemplo, consideremos todos los eventos cuyas coordenadas sean (*t*, 2, 3, 5), al variar *t*. Este conjunto de puntos formará una línea continua en el espacio-tiempo (suponiendo que las funciones de las coordenadas sean continuas). En un sistema de coordenadas inercial, todas estas curvas coordenadas serán rectas.

Es por esta razón que “ligar un sistema de coordenadas” a un cuerpo en rotación da como resultado coordenadas no inerciales. Por ejemplo, si utilizáramos uno de los globos en el lado izquierdo de la figura III.3 como el origen de un sistema de coordenadas, las coordenadas resultantes no serían inerciales. La cuestión crítica consiste en darse cuenta de que *la distinción entre los sistemas de coordenadas inerciales y los sistemas de coordenadas no inerciales se basa en una distinción previa entre las trayectorias inerciales y las trayectorias no inerciales en el espacio-tiempo mismo*. Es sólo porque el espacio-tiempo físico posee un tipo de estructura afín correcto que se hace posible una diferenciación significativa de los sistemas de coordenadas.

El espacio-tiempo galileano proporciona el escenario ideal para la mecánica newtoniana. La estructura del espacio-tiempo permite la exposición de las leyes del movimiento de Newton; permite la diferenciación de los sistemas giratorios y los no giratorios y por lo tanto la explicación de los fenómenos en el experimento de la cubeta; permite la definición de los sistemas de coordenadas inerciales y de los no inerciales. Pero *no* permite definir el reposo absoluto y tampoco la velocidad absoluta en general. Así que la pregunta embarazosa que Newton no podía responder empíricamente —¿cuál es mi velocidad absoluta actual?— no puede plantearse en el espacio-tiempo galileano. Si la mecánica newtoniana hubiera resultado ser adecuada para explicar todos los fenómenos observables, los físicos casi seguramente hubieran adoptado el espacio-tiempo galileano como la explicación correcta de la estructura espacio-temporal. Claro es que podríamos insistir en la totalidad de la propuesta newtoniana mencionando los puntos del espacio que persisten a través del tiempo, pero esa estructura adicional no nos proporcionaría un mayor poder explicativo. Y como Leibniz y Mach, podríamos tratar de eliminar completamente el espacio-tiempo en cuanto entidad, pero no es claro cuáles podrían ser tanto la motivación como la factibilidad de semejante empresa. La estructura del espacio-tiempo no es directamente observable, pero aun así juega un rol esencial en la formulación de la teoría física. No hay más necesidad de que se intente eliminar en la física la estructura del espacio-tiempo, que de que se intente eliminar la existencia de los átomos, tan sólo porque no es posible observarlos directamente.

Sin embargo, la teoría de Newton tenía sus defectos. Diversos fenómenos, especialmente aquellos en que participan la luz y el electromagnetismo, se resistieron a todos los esfuerzos por explicarlos en términos newtonianos. Como resultado, la mecánica newtoniana fue derrocada en favor de la teoría especial de la relatividad, y tanto el espacio y tiempo absolutos de Newton como el espacio-tiempo de Galileo fueron derrocados por ella.

[1] Tercera respuesta, § 4. [Versión en español: G. W. Leibniz, *op. cit.*, p. 74.]

[2] Cuarta carta, § 13. [Versión en español: *ibid.*, p. 80.]

[3] G. Galilei, *op. cit.*, pp. 186-187. [Versión en español: *Diálogo sobre los dos máximos sistemas del mundo, ptolemaico y copemicano*, trad. de A. Beltrán, Alianza, Madrid, 1994, pp. 162-163.]

[4] I. Newton, *op. cit.*, pp. 20-21. [Versión en español: *op. cit.*, pp. 144-145.]

[5] Esta ley obviamente exige que uno pueda ser capaz de cuantificar la curvatura de una línea y determinar la dirección de la curvatura. Es posible hacerlo, pero la explicación de las operaciones matemáticas nos detendrían innecesariamente.

IV. LA RELATIVIDAD ESPECIAL

LA RELATIVIDAD ESPECIAL Y EL ESPACIO-TIEMPO DE MINKOWSKI

La relatividad especial es una teoría muy sencilla que comúnmente se expone de manera compleja y confusa. Algunos de los motivos de esta situación son históricos. El mismo Einstein exponía la teoría como la consecuencia de dos principios: *1*) la equivalencia de todos los marcos inerciales y *2*) la constancia de la velocidad de la luz. Sobre la base de estos dos principios es posible derivar las *transformaciones de Lorentz*, un conjunto de ecuaciones que relacionan un conjunto de coordenadas con otro conjunto de coordenadas. Pero ya desde ahora podemos ver que este acercamiento a la comprensión de la teoría se ha descarrilado seriamente.

Primero, la noción de un sistema inercial, o de un conjunto inercial de coordenadas, o de un marco de referencia inercial, más que fundamental es derivativa. Sólo es posible definir cada uno de estos conceptos mediante la referencia a una cierta estructura geométrica objetiva del espacio-tiempo mismo, para entender el sentido del calificativo “inercial”. De manera que se debería empezar con la geometría intrínseca, no con los sistemas de coordenadas ni con los marcos de referencia. Segundo, en el capítulo precedente nos empeñamos en eliminar las velocidades absolutas de la mecánica newtoniana. Fue grato concluir que en el espaciotiempo galileano nada tiene una velocidad absoluta. Así que nos debería parecer una especie de retroceso el hecho de fundar la relatividad especial en propuestas sobre la velocidad de cualquier cosa. Basar una explicación de la relatividad en un discurso sobre velocidades inevitablemente sugiere que de nuevo regresamos al espacio-tiempo absoluto newtoniano y al movimiento absoluto newtoniano. Es lo que suelen hacer de manera tácita las presentaciones populares de la relatividad, diciendo cosas como “En la relatividad, el tiempo se hace más lento cuando un objeto viaja rápidamente” o “En la relatividad, un objeto se encoge al acercarse a la velocidad de la luz” o “No nos percatamos de los efectos relativistas en la vida cotidiana porque distamos mucho de viajar a la velocidad de la luz”. Todas estas afirmaciones sugieren que un cuerpo, en cualquier momento dado, *tiene* una velocidad que puede estar más cerca o más lejos de la velocidad de la luz. Pero en la relatividad, tanto como en el espacio-tiempo galileano, tales velocidades simplemente no existen. No hay ningún hecho físico respecto a la velocidad de la Tierra en este momento; no es más correcto decir que dista mucho de viajar a la velocidad de la luz que decir que viaja a 99% de la velocidad de la luz. Para entender la relatividad, tenemos que suprimir toda idea de que las cosas, incluyendo la luz, tienen velocidades.

Para empezar, nos enfocaremos en únicamente un fenómeno físico que puede comprobarse de forma experimental. El fenómeno consiste en que *la trayectoria de la luz en un vacío es independiente del estado físico de su fuente*. En específico, supongamos que dos focos de *flash* se acercan uno al otro velozmente. Al pasar uno junto al otro, ambos se encienden. Es un hecho empírico verificable el hecho de que la luz de los dos focos llegará hasta un solo observador en cualquier parte del universo exactamente en el mismo instante. Así, las trayectorias de las luces, las rutas que siguen, no dependen de lo que la fuente estaba haciendo al emitir la luz.

Pueden ser diferentes otros aspectos de la luz proveniente de los dos focos de *flash*. Por ejemplo, los observadores suelen ver una diferencia en el *color* o la *longitud de onda* o la *frecuencia* de una de las dos luces, pero de todos modos los dos rayos de luz le llegan al mismo tiempo a un observador. Ésta es obviamente una condición necesaria si queremos asentar que “la velocidad de la luz es constante”: si un rayo de luz puede rebasar o alcanzar al otro, entonces no se podría considerar que sus velocidades son las mismas. Pero en el fenómeno que hemos descrito no se menciona la velocidad de ninguna cosa ni se dice que debamos observar o calcular la velocidad de cualquier cosa. Esta cuestión física será para nosotros como una piedra de toque.

Si aceptamos que en un vacío no existe una estructura física, salvo la estructura misma del espacio-tiempo, entonces el comportamiento de la luz en un vacío implica que únicamente *la geometría del espacio-tiempo puede determinar la trayectoria de los rayos de luz*. Es decir, dado cualquier punto *p* en el espacio-tiempo, la estructura del espaciotiempo debe determinar hacia dónde la luz emitida desde ese punto *p* (en cualquier dirección posible) habrá de ir. Este conjunto de puntos, los lugares a donde la luz emitida desde *p* (en un vacío) podría llegar, tiene el nombre de *cono de luz futuro de p*. Y de manera similar, cualquier luz que logre llegar a *p* tuvo que haber venido a lo largo de una de las trayectorias de un conjunto particular de trayectorias que forman el *cono de luz pasado de p*. Así, nuestra simple observación sobre el comportamiento de la luz en un vacío implica que la geometría del espacio-tiempo, sea como fuere, debe asociar con cada evento un cono de luz pasado y un cono de luz futuro.

En este punto, en vez de pretender derivar la relatividad especial a partir de algún fenómeno o de algunos principios generales, simplemente expondremos, en la forma más clara posible, qué especie de estructura espacio-temporal propone la relatividad especial. Dada esta propuesta teórica, nosotros pasaremos a examinar algunos planteamientos físicos que es posible enmarcar con base en esa geometría. Como veremos, la afirmación de que “la velocidad de la luz es constante” es físicamente muy compleja, y sólo podremos entenderla con claridad una vez que hayamos estudiado a fondo otros aspectos del tema.

Con el fin de especificar la estructura geométrica del espacio-tiempo en la relatividad especial, seguiremos la misma estrategia que anteriormente empleamos en el caso de E^3. Ya vimos que es posible describir la geometría de E^3 indirectamente mediante la referencia a los sistemas de coordenadas cartesianos: un espacio es E^3 si y sólo si *admite* coordenadas que se relacionan con la estructura geométrica en una forma peculiar. El espacio-tiempo de la relatividad especial es el *espacio-tiempo de Minkowski*, y su geometría también puede especificarse indirectamente mediante la referencia a ciertos sistemas de coordenadas especiales, las *coordenadas de Lorentz*. De hecho, la forma en que las coordenadas de Lorentz se relacionan con la geometría del espacio-tiempo de

Minkowski es casi la misma en que las coordenadas cartesianas se relacionan con el espacio euclidiano. Sólo hay una mínima diferencia entre ambos casos. Pero no debemos olvidar la advertencia de Einstein: las coordenadas no necesariamente tienen una importancia *física* directa. En particular, una de las coordenadas será la llamada coordenada *t*, lo cual sugiere que tiene algo que ver con el *tiempo* o con los *relojes*. Pero nosotros no estamos presuponiendo semejante vínculo: por ahora, las coordenadas no son más que números que se asignan a los eventos.

El espacio-tiempo de Minkowski es tetradimensional, es decir, necesitamos cuatro coordenadas para cubrirlo con funciones coordenadas continuas. Las cuatro tradicionalmente se denominan *t, x, y* y *z*. Puesto que las funciones coordenadas son siempre continuas, este hecho por sí solo determina la topología del espacio-tiempo de Minkowski: un punto se traslada continuamente en el espacio-tiempo de Minkowski tan sólo si sus coordenadas de Lorentz cambian todas de forma continua al mismo tiempo que se mueve. Ya que esta situación es con exactitud la misma que la de las coordenadas cartesianas en E^4, el espacio-tiempo de Minkowski y el espacio euclidiano de cuatro dimensiones son topológicamente idénticos.[1]

La estructura afín del espacio-tiempo de Minkowski tiene exactamente la misma relación con las coordenadas de Lorentz que la estructura afín del espacio euclidiano tiene con las coordenadas cartesianas. Es decir, un conjunto de eventos en el espacio-tiempo de Minkowski forma una línea recta completa si y sólo si, para los números elegidos *A, B, C, D, E, F, G* y *H*, las coordenadas de los puntos en el conjunto tienen los valores

$$(As + B,\ Cs + D,\ Es + F\ Gs + H),$$

donde *s* corre sobre los números reales. Como en el caso del espacio euclidiano, por lo menos uno de los números *A, C, E* y *G* debe ser no nulo.

En resumen, tanto la estructura topológica como la estructura de línea recta del espacio de Minkowski son iguales a E^4. Esto significa que los *diagramas* del espacio-tiempo euclidiano pueden representar esos aspectos del espaciotiempo de Minkowski de una manera particularmente sencilla. Las líneas continuas en el diagrama pueden corresponder a líneas continuas en el espacio-tiempo, y las líneas rectas en el diagrama pueden corresponder a líneas rectas en el espacio-tiempo. Claro, no podemos producir diagramas tetradimensionales reales, pero podemos describir de manera usual el espacio euclidiano tridimensional. Así que si eliminamos una de las dimensiones del espacio-tiempo de Minkowski, las demás pueden describirse sencilla y precisamente. Sólo nos queda por examinar la estructura métrica del espacio-tiempo de Minkowski.

Dadas las coordenadas cartesianas en E^3, la estructura métrica se representa, en términos de las coordenadas, por la función

$$D(p, q) = \sqrt{(X(p) - X(q))^2 + (Y(p) - Y(q))^2 + (Z(p) - Z(q))^2},$$

donde $X(p)$, $Y(p)$ y $Z(p)$ son las tres funciones coordenadas de las coordenadas cartesianas. En el espacio-tiempo de Minkowski, la estructura geométrica análoga es el llamado *intervalo relativista invariante*, y se puede representar, en términos de las coordenadas de Lorentz, como

$$I(p, q) = \sqrt{(T(p) - T(q))^2 - (X(p) - X(q))^2 - (Y(p) - Y(q))^2 - (Z(p) - Z(q))^2}$$

ECUACIÓN IV.1

Aquí hemos terminado nuestra explicación del espaciotiempo de Minkowski. Nos queda por hacer el análisis de esta geometría.

El meollo de la relatividad especial se encuentra en la ecuación IV.1, e invertiremos bastante tiempo en el devela-miento de su significación. Pero antes unas cuantas advertencias, especialmente para los lectores que han estudiado física.

En muchos manuales de física, el intervalo se define como el cuadrado de la cantidad que aparece arriba. Pero es más apropiada la raíz cuadrada, puesto que las proporciones de las dimensiones en el espacio-tiempo de Minkowski son proporcionales a las proporciones de la cantidad definida en la ecuación IV.1. La opción de la raíz cuadrada al parecer tiene consecuencias muy extrañas: por ejemplo, el intervalo que así se define resulta ser a veces un número *imaginario*. Pero puesto que lo que nos importa en realidad son las proporciones entre estos números, eso no representa una dificultad. También, en algunos manuales se utiliza un método convencional diferente, donde el cuadrado de la diferencia en la coordenada t se sustrae de la suma de los cuadrados de las diferencias de las otras coordenadas. Cambiar la ecuación IV.1 en esta forma daría como resultado que todos los valores reales del intervalo cambiaran, convirtiéndose en valores imaginarios, y viceversa. Pero nuevamente, puesto que lo que importa son las proporciones entre estos números, esto no significa una diferencia real.

Finalmente, en muchos manuales de física se da una definición para una versión diferencial del intervalo. Es decir, si los puntos p y q están suficientemente cercanos el uno del otro, sus valores coordenados estarán cercanos el uno del otro. Así, por ejemplo, $T(p) - T(q)$ será un número muy pequeño, que podemos designar como dT. En la forma diferencial, la ecuación IV.1 se convierte en

$$dI = \sqrt{dT^2 - dX^2 - dY^2 - dZ^2}.$$

Matemáticamente, la forma diferencial es mucho más fuerte y debe usarse en la teoría general de la relatividad. Pero en nuestra presentación usaremos la ecuación IV.1, porque con ella podremos hacer los cálculos con mayor facilidad.

A la cantidad $I(p, q)$ que se define en la ecuación IV.1 a veces se le llama la *métrica de Minkowski*, pero esto puede resultar confuso. La función no cumple con los requerimientos definicionales de una función métrica: no es definida-positiva y no satisface la desigualdad del triángulo. Esto debería servir como una advertencia de que la geometría del espacio-tiempo de Minkowski es fundamentalmente muy diferente de la geometría del espacio euclidiano, a pesar de la simili-tud de la estructura topológica y la estructura afín. Así, mientras que las estructuras topológicas y las estructuras de línea recta en un diagrama de espacio-tiempo representan exactamente lo que parecen representar, *las distancias entre los puntos* en un diagrama de espacio-tiempo no se corresponden de forma directa con *los intervalos entre los eventos* representados. Esto siempre se debe recordar al interpretar un diagrama de espacio-tiempo relativista.

Es un hecho matemático que si un espacio admite un conjunto de coordenadas de Lorentz, también admite muchos sistemas de coordenadas similares, del mismo modo que el espacio euclidiano admite una infinita cantidad de sistemas de coordenadas cartesianos. Estos sistemas de coordenadas varían de acuerdo con la ubicación del origen, la orientación de los ejes de las coordenadas y la escala. Dado cualquier par de sistemas de coordenadas de Lorentz, habrá un conjunto de ecuaciones que transforman a las coordenadas de un evento en un sistema en las coordenadas de otro sistema. El conjunto de transformaciones que concuerdan en el origen y la escala de las coordenadas son las llamadas *transformaciones de Lorentz*; hay un conjunto más amplio que sólo tiene que concordar en la escala, las llamadas *transformaciones de Poincaré*. La opción de los ejes de coordinadas en el sistema de coordenadas de Lorentz no se puede hacer con tanta libertad como en el caso de los ejes de las coordenadas cartesianas en el espacio euclidiano: cualquier línea recta en el espacio euclidiano puede servir como un eje de coordenadas en el sistema cartesiano, pero algunas líneas rectas en el espacio-tiempo de Minkowski no pueden servir como curvas coordenadas en un marco lorentziano. El espacio-tiempo de Minkowski no es isotrópico en la misma forma en que lo es el espacio euclidiano. Las diferentes direcciones en el espacio-tiempo de Minkowski pueden tener un tono geométrico diferente. Esto le permite al espacio-tiempo de Minkowski explicar el comportamiento de la luz.

La función que define el intervalo entre los eventos no es definido-positivo, y la función toma el valor cero para algunos pares de puntos distintos. Sólo por esta razón no es posible considerar con precisión el intervalo como una *distancia:* cualquier par de puntos distintos debe tener una cierta distancia no nula entre ellos. Para entender la estructura de estos intervalos de valor cero, dibujaremos un diagrama de espacio-tiempo de Minkowski con una dimensión suprimida. Ya que la topología y la estructura afín del espacio-tiempo de Minkowski son iguales a la topología y la estructura afín del espacio euclidiano, nuestro diagrama tendrá el aspecto de E^3. Además, utilizaremos las coordenadas cartesianas en este E^3 para representar las coordenadas de Lorentz en el espacio-tiempo de Minkowski. Por lo tanto, en vez de nombrar las coordenadas x, y y z, las designaremos como x, y y t. Según el método convencional, la coordenada t corre verticalmente en la página, y las coordenadas x y y horizontalmente. Consideremos el origen o del sistema de coordenadas, como el evento con las coordenadas (0, 0, 0, 0). Para que cualquier otro evento p tenga un intervalo cero desde el origen, debemos tener

$$I(p, 0) =$$
$$\sqrt{(T(p) - T(o))^2 - (X(p) - X(o))^2 - (Y(p) - Y(o))^2 - (Z(p) - Z(o))^2} =$$
$$\sqrt{(T(p) - 0)^2 - (X(p) - 0)^2 - (Y(p) - 0)^2 - (Z(p) - 0)^2} =$$
$$\sqrt{(T(p))^2 - (X(p))^2 - (Y(p))^2 - (Z(p))^2} = 0.$$

Es decir, el conjunto de eventos en el intervalo cero desde el origen son los puntos p cuyas coordenadas satisfacen la ecuación

$$t_p{}^2 - x_p{}^2 - y_p{}^2 - z_p{}^2 = 0.$$

Si dibujamos este conjunto de eventos en un diagrama de espacio-tiempo (suprimiendo la dimensión *z)*, veremos que forman un doble cono cuyo punto de cono es el origen. La geometría intrínseca del espacio-tiempo de Minkowski asocia un doble cono de esta especie con cualquier evento. Dos conos de este tipo aparecen en la figura IV.1.

En la teoría de la relatividad, el comportamiento de la luz en un vacío se asocia con tales conos, así que se les denomina *conos de luz*. En específico, la relatividad (o, más exactamente, la teoría relativista del electromagnetismo) implica la

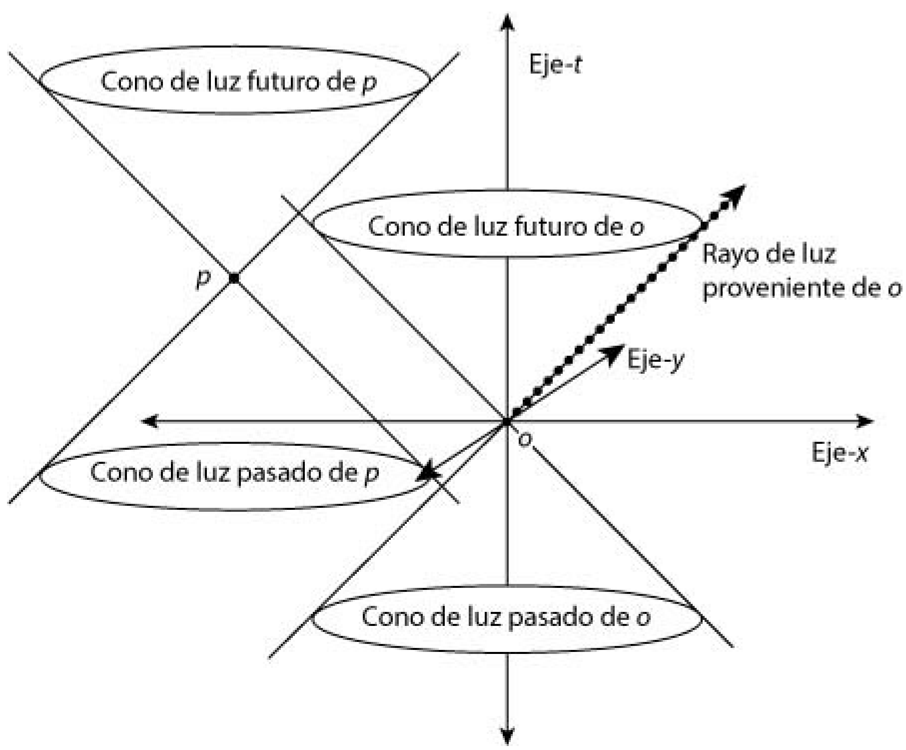

FIGURA IV.1

Ley de la luz: La trayectoria de un rayo de luz proveniente de un evento (en un vacío) es una línea recta en el cono de luz futuro de ese evento.

Si un foco de *flash* se enciende en *o* (en un vacío), la luz, extendiéndose en todas las direcciones, habrá de ocupar la superficie de todo su futuro cono de luz. En un espacio tridimensional, un foco de *flash* produce una esfera de luz que se expande. Con una dimensión espacial suprimida, esto se convierte en un círculo de luz que se expande, apareciendo como un cono en nuestro diagrama.

La ley de la luz sólo puede formularse en un espacio-tiempo que asocie con cada evento un cono de luz. Es de notar que la ley de la luz nada menciona sobre la fuente de la luz, salvo que la fuente la emite en un evento específico. Así, la ley de la luz implica los fenómenos arriba mencionados: dos rayos de luz emitidos desde el mismo punto en un vacío llegarán juntos hasta un observador distante. Hemos explicado este fenómeno sin definir, o siquiera mencionar, la "velocidad de la luz".

El cono de luz de *o* divide al resto del universo en cinco conjuntos topológicamente conectados: los eventos dentro del futuro cono de luz los eventos en el futuro cono de luz, los eventos dentro del pasado cono de luz, los eventos en el pasado cono de luz y los eventos en el exterior del cono de luz. A los eventos dentro del pasado o futuro cono de luz se les denomina eventos *separados de tipo tiempo* de *o*, a los eventos en el pasado o futuro cono de luz se les califica de eventos *separados de tipo luz* de *o*, y a los eventos fuera del cono de luz se les califica de eventos *separados de tipo espacio* de *o*. Dados nuestros métodos convencionales, en cualquier sistema de coordenadas de Lorentz y en el caso de todos los pares de eventos separados de tipo luz, el intervalo *I(p, q*) es positivo y real: para los eventos separados de tipo luz, *I(p, q*) es cero, y para los eventos separados de tipo espacio, *I*(*p, q*) es imaginario. Todos los marcos lorentzianos concuerdan en la proporción entre estos intervalos.

La afirmación de que *nada se puede mover más velozmente que la luz* se expresa en términos de esta geometría del siguiente modo:

> El rol limitador del cono de luz: La trayectoria de cualquier ente físico que pasa a través de un evento jamás pasa por fuera del cono de luz de ese evento.

Nuevamente, no tuvimos que definir la velocidad de ninguna cosa con el fin de aseverar el principio físico.

Puesto que el espacio-tiempo de Minkowski tiene una estructura afín, la primera ley de Newton puede expresarse en la misma forma que en el espacio-tiempo galileano:

> Ley relativista de la inercia: La trayectoria de cualquier ente físico que no se encuentre sujeto a ningún tipo de influencias externas es una línea recta en el espacio-tiempo de Minkowski.

Dado que un rayo de luz en un vacío no está sujeto a ningún tipo de influencias externas, se deduce que la trayectoria de luz en un vacío es recta. Pero la ley relativista de la inercia también gobierna los cuerpos masivos cuyas trayectorias no yacen en el cono de luz. De hecho, según una premisa de la relatividad, la trayectoria de cualquier cuerpo masivo es en todos los casos de tipo tiempo, y por lo tanto siempre yace dentro del cono de luz de cualquier evento que atraviese.

La ley relativista de la inercia, junto con el hecho de que la estructura afín del espacio-tiempo de Minkowski es idéntica a la de E^4, implica que en un diagrama de espacio-tiempo la imagen de la trayectoria de un objeto libre de fuerzas será una línea recta. Nosotros utilizaremos ampliamente este hecho al resolver los problemas relativistas.

Dada la estructura geométrica del espacio-tiempo de Minkowski, la ley relativista de la inercia nos permite desde ahora hacer algunas predicciones. Por ejemplo, si las trayectorias de dos entes físicos libres (ya sean cuerpos masivos o rayos de luz) se intersecan, sólo lo hacen una sola vez. Esto se deduce del hecho de que en el espacio-tiempo de Minkowski ningún par de líneas rectas se intersecan más de una vez. Ésta pudiera parecer una predicción algo trivial, pero más adelante veremos que no resulta verdadera en la relatividad general.

Si queremos obtener predicciones físicas más impresionantes de la teoría, necesitamos más conexiones entre la geometría del espacio-tiempo de Minkowski y el comportamiento observable de los objetos materiales. La más importante de semejantes conexiones es una *hipótesis* cuyo estatus escrutaremos posteriormente:

> Hipótesis del reloj: La cantidad de tiempo que un reloj de precisión muestra que ha transcurrido entre dos eventos es proporcional al intervalo a lo largo de la trayectoria del reloj entre esos eventos o, en resumen, los relojes miden el intervalo a lo largo de sus trayectorias.[2]

En la hipótesis del reloj está la clave para comprender toda la relatividad.

La noción misma de un “reloj” en la relatividad podría parecer un poco extraña. Al fin y al cabo, jamás propusimos algo parecido al tiempo absoluto newtoniano al construir el espacio-tiempo de Minkowski. A diferencia del espacio-tiempo galileano, el espacio-tiempo de Minkowski no se encuentra foliado en rebanadas de simultaneidad; de hecho, la noción misma de los “eventos simultáneos” carece totalmente de contenido. La estructura del cono de luz de algún modo reemplaza la foliación. Puesto que no existe el tiempo absoluto en la relatividad, no es posible que los relojes lo midan. Pero los relojes deben medir *algo*: dos relojes de precisión, colocados el uno junto al otro, medirán los segundos al unísono, de manera que debe haber algún tipo de estructura geométrica en el espacio-tiempo que ambos relojes reflejen. Según la hipótesis del reloj, esa estructura es el intervalo. Los relojes de precisión en la relatividad, como los contadores de kilómetros en los automóviles, miden la longitud de su trayectoria a través del espacio-tiempo. Si aceptamos esto, entonces dos relojes que se encuentren lado a lado deberán hacer tic-tac al unísono, puesto que sus trayectorias a través del espacio-tiempo tendrán la misma longitud. Pero los relojes que sigan una trayectoria diferente habrán de registrar tiempos transcurridos diferentes entre el mismo par de eventos, de la misma manera que los autos que sigan rutas diferentes entre los mismos lugares habrán de registrar kilometrajes diferentes.

LA PARADOJA DE LOS GEMELOS

Utilicemos la hipótesis del reloj para explicar el fenómeno icónico de la relatividad: la llamada paradoja de los gemelos. De forma cualitativa, la situación es ésta: dos gemelos, con relojes idénticos, inicialmente están en una situación en que se encuentran lado a lado en sendas naves espaciales y sin estar sujetos a ningún tipo de fuerzas. El gemelo A enciende el motor de su nave y tras un momento lo apaga. Las naves de los gemelos se separan. Un poco después, el gemelo A nuevamente enciende su motor, pero en dirección opuesta, acercándose al gemelo B, quien nunca ha encendido el motor de su nave. El gemelo A enciende el motor por tercera vez, regresando para posarse junto al gemelo B. Cuando los gemelos comparan sus relojes, descubren que en el reloj del gemelo B ha transcurrido más tiempo que en el del gemelo A. Además, el gemelo B parece haber envejecido biológicamente más que el gemelo A. ¿Cómo predice la relatividad espacial este caso?

Con el fin de obtener una explicación cuantitativa del fenómeno, tenemos que especificar exactamente las trayectorias de ambos gemelos, y para hacerlo utilizaremos un sistema de coordenadas de Lorentz conveniente. La conveniencia del sistema elegido no juega un rol en la explicación: cualquier otro marco lorentziano daría el mismo resultado, pero los cálculos serían más complejos. En este sistema de coordenadas conveniente, la trayectoria del gemelo B siempre se da a lo largo del eje t: sus coordenadas x, y y z siempre son 0. El gemelo A enciende el motor de su nave por vez primera en el origen o. Luego se mueve inercialmente hasta llegar al evento p, cuyas coordenadas son (5, 4, 0, 0). De nuevo enciende el motor de su nave y se mueve inercialmente hasta llegar junto a su hermano en el evento q, cuyas coordenadas son (10, 0, 0, 0). Cuando se reúnen, el reloj del gemelo B indica que han transcurrido 100 días. ¿Cuánto tiempo ha transcurrido según el reloj del gemelo A?

Utilizaremos un diagrama de espacio-tiempo para resolver este problema. Los movimientos inerciales de los gemelos se representarán mediante líneas rectas en el diagrama, de acuerdo con la ley relativista de la inercia. Puesto que los valores coordenados de y y z de ambos gemelos siempre son cero, ignoraremos estas dimensiones y usaremos un diagrama de dos dimensiones en la figura IV.2.

Para determinar cuánto tiempo habrá transcurrido según el reloj del gemelo A, necesitamos saber la proporción de la longitud de la trayectoria del gemelo B en comparación con la trayectoria del gemelo A. Para hacerlo, emplearemos la cantidad I calculada en estas coordenadas de Lorentz. La longitud de la línea recta $\overline{oq}$ es

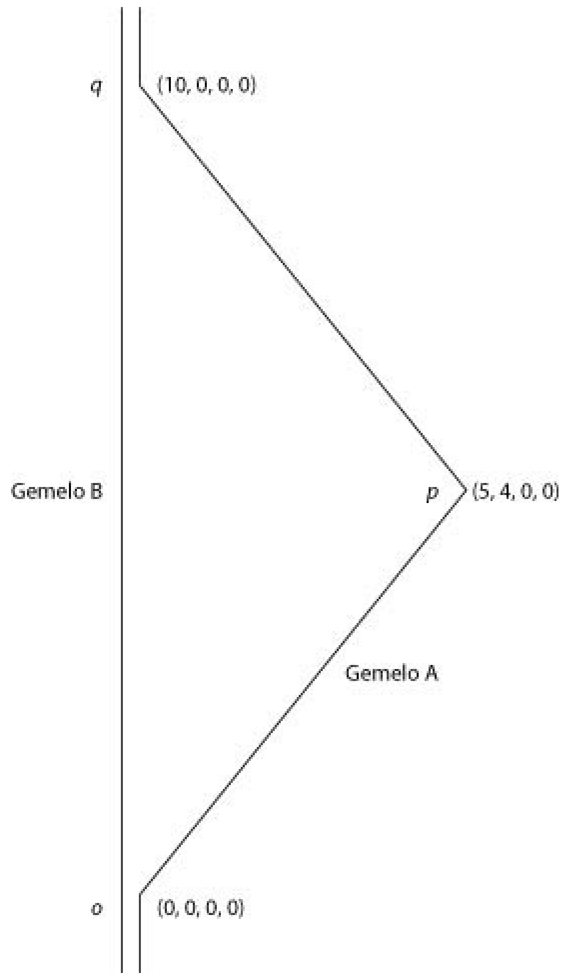

FIGURA IV.2

$$\sqrt{\left(T(q) - T(o)\right)^2 - \left(X(q) - X(o)\right)^2 - \left(Y(q) - Y(o)\right)^2 - \left(Z(q) - Z(o)\right)^2} =$$

$$\sqrt{(10 - 10)^2 - (0 - 0)^2 - (0 - 0)^2) - (0 - 0)^2} =$$

$$\sqrt{(100} = 10.$$

La longitud de $\overline{oq}$ es

$$\sqrt{\left(T(p) - T(o)\right)^2 - \left(X(p) - X(o)\right)^2 - \left(Y(p) - Y(o)\right)^2 - \left(Z(p) - Z(o)\right)^2} =$$

$$\sqrt{(5 - 0)^2 - (4 - 0)^2 - (0 - 0)^2) - (0 - 0)^2} =$$

$$\sqrt{(25 - 16} = \sqrt{9} = 3.$$

La longitud de $\overline{pq}$ es

$$\sqrt{(10-5)^2-(0-4)^2-(0-0)^2)-(0-0)^2} =$$

$$\sqrt{(25-16} = \sqrt{9} = 3.$$

Así, la longitud total de la trayectoria del gemelo A, en estas unidades, es 3 + 3 = 6, mientras que la longitud total de la trayectoria del gemelo B es 10. Si el reloj de B muestra 100 días de tiempo transcurrido, el de A muestra 60. Cuando los gemelos se reúnan, A será 40 días más joven que B.

Una razón por la cual pudiera parecer "paradójico" el caso de los gemelos, es que cualquiera que vea la figura IV.2 habría de concluir que la ruta de A desde *o* a *q* es *más larga* que la de B. Y en efecto, tal cosa es verdad en las representaciones *del diagrama*. Pero éste se ha dibujado en el espacio euclidiano, por lo que tenemos que ser muy cautos al interpretarlo. En el espacio-tiempo de Minkowski, la desigualdad del triángulo no es válida para el intervalo.[3] Más bien, el intervalo obedece a una especie de anti-desigualdad del triángulo: la suma de las longitudes de los dos lados de A en el triángulo en la figura IV.2 siempre será *menor* que la longitud del lado restante.

La "paradoja" de los gemelos ha inspirado más confusión respecto a la relatividad que cualquier otro fenómeno. La explicación del fenómeno en función de la geometría intrínseca del espacio-tiempo de Minkowski es bella en su sencillez: los relojes miden el intervalo a lo largo de sus líneas de mundo, y la línea de mundo de B entre *o* y *q* es más larga que la de A. Punto. No hay nada más que decir.

Aclaremos algunas de las confusiones al respecto.

Confusión 1: En la relatividad, todo el movimiento es el movimiento relativo de los cuerpos. Pero el movimiento relativo de A con respecto a B es exactamente igual al movimiento relativo de B con respecto a A. Así, sus situaciones físicas son exactamente las mismas. Por lo tanto, cuando vuelvan a reunirse, sus relojes deberán mostrar el mismo tiempo transcurrido, en virtud de la simetría.

Todos los planteamientos en la confusión 1 son incorrectos. Primero, no es verdad en cualquier sentido interesante que "en la relatividad todo los movimientos son relativos". Por ejemplo, la explicación de los globos giratorios newtonianos en la relatividad especial es esencialmente *idéntica* a la explicación que el mismo Newton da o a la explicación que se propone en el espacio-tiempo galileano: hay una estructura geométrica objetiva en el espacio-tiempo que define las trayectorias inerciales de todos los cuerpos, independientemente de otros cuerpos. La tensión en la cuerda en un caso indica que los globos no siguen trayectorias inerciales: sus líneas de mundo son curvas. El diagrama de espacio-tiempo de los globos giratorios en la relatividad especial sería como se muestra en la figura III.3, excepto que en lugar de las rebanadas de simultaneidad veríamos una estructura de cono de luz. La relatividad no es una teoría "relacionista" del espacio-tiempo del tipo que Leibniz buscaba.

Confusión 2: Como "solución" de la confusión 1, muchos manuales sugieren que la paradoja de los gemelos se resuelve mediante la *aceleración*: la situación del gemelo A no es realmente simétrica respecto de la situación del gemelo B porque el primero ha tenido que encender el motor de su nave. El motor (de acuerdo con la versión relativista de la segunda ley de Newton) produce una fuerza que el gemelo A puede sentir y que dobla su línea mundial. El gemelo B no enciende el motor de su nave y no siente ninguna fuerza. Por lo tanto, el gemelo B sabe que es él quien está "realmente en reposo", mientras que su gemelo, sintiendo la fuerza, sabe que él está "realmente viajando".

La confusión 2 se halla tan extendida que sería imposible citar todos los ejemplos al respecto, pero dos de ellos pueden darnos una idea del tono general. Wolfgang Rindler dice en su *Essential Relativity:*

> Ahora bien, la paradoja es ésta: ¿no podría A […] decir con igual derecho que había sido *él* quien había permanecido donde estaba, mientras B hacía un viaje redondo y que, en consecuencia, B debería ser el más joven cuando se reunieran de nuevo? La respuesta, claro, es no, y lo siguiente resuelve la paradoja: B ha permanecido en reposo en un marco inercial único, mientras que A, en el caso más sencillo de un movimiento de ida y vuelta uniforme —digamos, de la Tierra a una estrella cercana y de ésta a la Tierra—, debió haber por lo menos acelerado brevemente, saliendo del marco de B y entrando en otro, para luego desacelerar brevemente con el fin de girar, y al final para desacelerar de nuevo y detenerse junto a B. Estas aceleraciones (positivas y negativas) las *siente* A, y por lo tanto él no puede haberse imaginado que había sido él quien había permanecido en reposo.[4]

El gran Richard Feynman, en sus *Lecciones de física*, hace una exposición similar:

> Solamente la gente que cree que el principio de la relatividad significa que *todo movimiento* es relativo lo califica de "paradoja"; dicen: "¿Bueno, no podemos decir que desde el punto de vista de Pablo era *Pedro* el que se estaba moviendo y por lo tanto debería parecer que había envejecido más lentamente? Con base en la simetría, el único resultado posible es que ambos necesariamente tengan la misma edad cuando se reúnan nuevamente". Pero para que ambos se reúnan de nuevo y que la comparación pueda hacerse, Pablo debe detenerse al final del viaje y comparar su reloj con el de Pedro, o sencillamente tiene que regresar, y el que regresa tiene que ser el individuo que estaba viajando, cosa que éste sabe porque tiene que dar la vuelta. Al dar la vuelta, todo tipo de cosas inusuales sucedieron en su nave espacial —los cohetes se encendieron, hubo cosas que se apilaron contra una pared, y así por el estilo— mientras que Pedro nada sintió.
>
> Así que la forma de expresar la regla es decir que el individuo que ha sentido las aceleraciones, que ha visto las cosas amontonarse contra una pared, etc., es el que sería el más joven. Es ésta la diferencia entre ellos en un sentido "absoluto", y es ciertamente correcta.[5]

Todo lo que se dice en esta "explicación" es erróneo.

Obsérvese, primero, que fuimos capaces de predecir el efecto sin calcular la aceleración de cosa alguna: únicamente calculamos la proporción entre las longitudes de las dos trayectorias. Las aceleraciones no juegan rol alguno en la explicación del resultado final. De hecho, resulta una cuestión sencilla modificar la situación para que B acelere de la misma manera que A o incluso más que A, pero aún así termina siendo más viejo que A. En la figura IV.3 hemos añadido aceleraciones exactamente iguales a la trayectoria de B. Ahora B enciende sus cohetes con exactamente la misma aceleración que A, pero el pequeño "empujón" en su línea mundial no hará que ésta sea significativamente más corta: él seguirá siendo el más viejo de los dos cuando ambos se reúnan. Si añadimos más empujones de ese tipo, podemos hacer que B acelere de forma arbitraria *más* que A, y que aún así siga siendo el más viejo.

En el espacio-tiempo de Minkowski, por lo menos uno de los gemelos deberá acelerar para que puedan volver a reunirse: como se dijo arriba, un par de líneas rectas en el espacio-tiempo de Minkowski sólo pueden encontrarse una vez. Esto es incidental en el efecto: en la relatividad general, gemelos que se encuentran ambos todo el tiempo en trayectorias *inerciales* pueden volver a encontrarse más de una vez, mostrando una diferencia en su envejecimiento cuando eso sucede. Tanto Rindler como Feynman señalan que en la relatividad la aceleración es objetiva, al igual que en el espacio y el tiempo absolutos newtonianos y en el espacio-tiempo galileano. Esto es verdad pero irrelevante; el asunto reside en cuál es la *longitud* de las líneas de mundo, no cuál es el grado de su *doblamiento* (o curvatura).

La confusión 3 liga la paradoja de los gemelos con la "dilatación relativista del tiempo", la cual se ilustra en la premisa de que "mientras más rápido se mueve el reloj, más lentamente mide los segundos". Si añadimos a esta premisa la proposición de que el gemelo A es el que "realmente está en movimiento", mientras que el gemelo B "en verdad está en reposo" (véase Rindler y Feynman, arriba), caeremos en un torbellino de malentendidos. El fenómeno de los gemelos se explica sin necesidad de atribuirle a nadie ningún tipo de "movimiento" o "velocidad" o "reposo"; es una simple cuestión de geometría del espacio-tiempo. El espacio-tiempo de Minkowski, al igual que el espacio-tiempo galileano (y a diferencia del espacio y el tiempo absolutos de Newton), no habla de una noción del "reposo" opuesta al "movimiento". La trayectoria de un cuerpo puede ser *recta* o *curva* y por lo tanto *inercial* o *acelerada*, pero no hay ningún hecho sobre "cuán rápido se mueve un reloj". Con mayor razón, no es posible explicar el fenómeno con relación a cuán rápido se mueve cada uno de los gemelos.

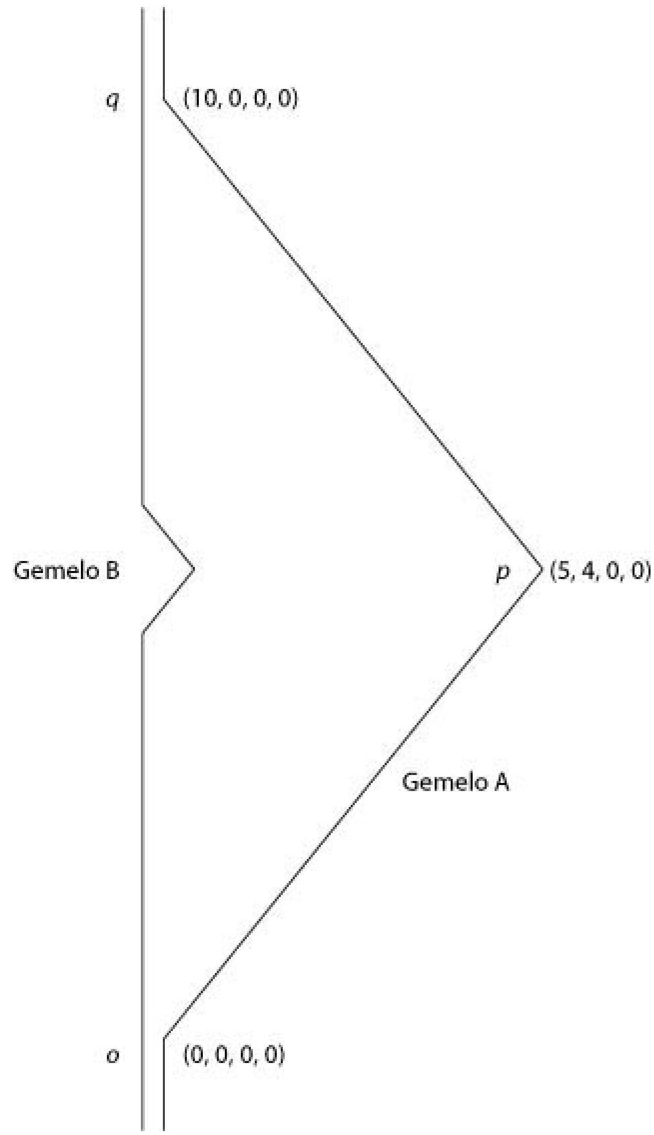

FIGURA IV.3

La ley de la luz, la ley relativista de la inercia y la hipótesis del reloj, todas relacionan el comportamiento de ciertos objetos perceptibles con el espacio-tiempo de Minkowski. Para completar una dinámica relativista, necesitaríamos una versión de la segunda ley de Newton que relacionara la fuerza impresa en un objeto con la curvatura de su línea de mundo. Esto requiere más matemáticas de las que nos podemos permitir aquí. Pero estos tres principios por sí mismos son suficientes para conectar la geometría del espacio-tiempo de Minkowski con los fenómenos observables.

LA REGLA DE MINKOWSKI, EL COMPÁS DE MINKOWSKI

La relatividad especial es, fundamentalmente, una hipótesis sobre la estructura del espacio-tiempo. Pero puesto que la geometría del espacio-tiempo no es de inmediato perceptible, la hipótesis sólo es relevante para las predicciones empíricas mediante las leyes que describen el comportamiento de las cosas perceptibles. De forma similar, la supuesta estructura euclidiana del espacio puede jugar un rol en la explicación de las propiedades de los diagramas visibles sólo si los instrumentos responsables del diagrama (la regla recta y el compás) reflejan, y por lo tanto revelan, la geometría subyacente del espacio mismo. La naturaleza del espacio-tiempo de Minkowski resulta para nosotros más intuitiva cuando se explicita esta analogía.

El rol de una regla en la geometría euclidiana consiste en indicar cuál es la estructura afín, es decir, escoger y marcar las líneas rectas en el espacio. La ley de la luz y la ley relativista de la inercia juegan exactamente el mismo rol en la relatividad especial con relación a una subclase de líneas rectas en el espacio-tiempo de Minkowski. En un vacío, protegido de las influencias externas, la trayectoria de un rayo de luz será una línea recta de tipo luz. De manera similar, si se le protege de las influencias externas, la trayectoria de un cuerpo masivo será una línea recta de tipo tiempo. De manera que nuestros principios físicos, enmarcados en los términos de la geometría minkowskiana, posibilitan el hecho de que parte de la estructura afín pueda hacerse visible en ciertas circunstancias.

No hemos propuesto postulados que nos puedan ayudar a identificar las líneas rectas de tipo espacio, líneas que se encuentren en el exterior de los conos de luz de los eventos que los componen. Estas líneas tienen un carácter físico diferente del que tienen las líneas de tipo luz y las líneas del tipo tiempo, puesto que (de acuerdo con el rol limitador del cono de luz) no pueden servir como trayectorias para los entes físicos. La diferencia esencial entre estos tipos de líneas no debe sorprender a nadie: también en el espacio-tiempo galileano, las líneas rectas que se encuentran dentro de una rebanada de simultaneidad específica son fundamentalmente diferentes de las trayectorias inerciales a través del tiempo. Pero resulta que no tenemos necesidad de añadir más premisas físicas que nos permitan identificar las líneas rectas de tipo espacio: el comportamiento de los rayos de luz, de los cuerpos masivos en movimiento inercial y de los relojes es suficiente para determinar toda la geometría.

El equivalente minkowskiano de un compás euclidiano sería un sistema físico que pudiera indicar todos los eventos que se encuentren en el mismo intervalo (a lo largo de una línea recta) desde algún evento central. Este requisito puede cumplirse si aceptamos la hipótesis del reloj. Supongamos que tenemos varios relojes de alarma cuya estructura es idéntica y cuyas alarmas han sido programadas para sonar un minuto después de que se pulse un botón. En un cierto momento, pulsamos el botón y los relojes son enviados en todas direcciones y con todos los impulsos posibles desde un evento central (*o* en la figura IV.4), permitiendo que todos continúen en su trayectoria inercialmente. De acuerdo con la hipótesis del reloj, todas las alarmas habrán de sonar cuando la longitud de sus trayectorias sea un minuto. ¿Qué tipo de forma asumirán los eventos de los sonidos de las alarmas?

Sea *o* el origen de algún sistema de coordenadas de Lorentz, elegido con el fin de que el intervalo, tal como se define en ese sistema, se corresponda con minutos. Entonces la ecuación para identificar un evento de sonido de alarma p es

$$I(p, o) = \sqrt{T(p)^2 - X(p)^2 - Y(p)^2 - Z(p)^2} = 1.$$

Es decir, p será un evento de sonido de alarma si se encuentra en el cono de luz futuro de o y si sus coordenadas (t, x, y, z) en este sistema de coordenadas satisfacen $t_p^2 - x_p^2 - y_p^2 - z_p^2 = 1$. Cuando dibujamos este lugar de los eventos en nuestro diagrama de espacio-tiempo euclidiano, obtenemos como resultado un *hiperboloide de revolución* (figura IV.4).

Una dificultad en la interpretación de la figura IV.4 consiste en que los puntos en el hiperboloide, los cuales se encuentran a una distancia ilimitada de o en el diagrama, están todos a *exactamente la misma "distancia" de* o *en la geometría intrínseca del espacio-tiempo de Minkowski*. O, más precisamente, las "longitudes" de todas las líneas rectas desde o hasta el hiperboloide son todas las mismas, aun cuando parecen tener diferentes longitudes en el diagrama. El hecho de que la trayectoria "vertical" punteada parezca más corta es únicamente una cuestión de convención: estamos obligados a dibujar de forma arbitraria la trayectoria de un reloj como si fuera vertical, aun cuando nada lo diferencie de cualquiera de los otros relojes. Si hubiéramos escogido de manera arbitraria que la trayectoria punteada que vemos en el lado derecho fuera vertical en el diagrama, la misma situación física se mostraría como en la figura IV.5.

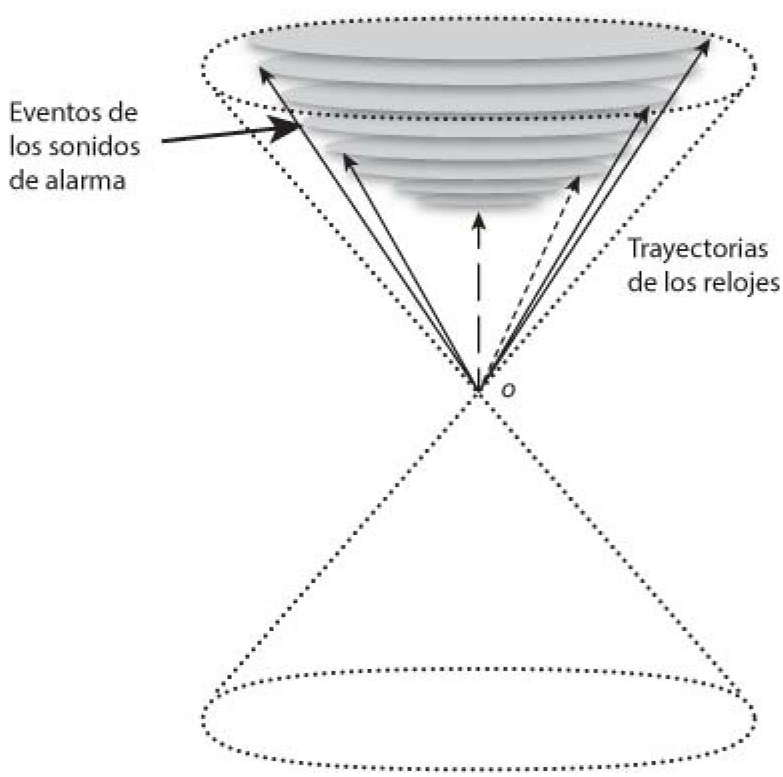

FIGURA IV.4

En la interpretación de las figuras IV.4 y IV.5 nos podría ayudar compararlas con las figuras III.3 y III.4. La "longitud temporal" de *cualquier* trayectoria que empieza en una rebanada de simultaneidad y termina en la siguiente es igual a cualquier otra, aun cuando pueda parecer que las líneas en el diagrama tienen longitudes bastante diferentes. Además, la relación que existe entre las figuras III.3 y III.4 es exactamente análoga a la relación que existe entre las figuras IV.4 y IV.5; en cada uno de los casos, ambos diagramas representan la misma situación física. Así que aunque en la figura IV.4 y en la figura IV.5 al parecer la alarma del reloj "vertical" suena primero y las demás se dilatan, no existe nada en la geometría del espacio-tiempo misma que justifique semejante suposición.

Una diferencia esencial entre la estructura temporal del espacio-tiempo galileano y el espacio-tiempo de Minkowski es que en el espacio-tiempo galileano el tiempo transcurrido a lo largo de una línea mundial es una función tan sólo de las rebanadas de simultaneidad en que se inicia y termina: son irrelevantes los detalles adicionales de la trayectoria. Y sin mencionar siquiera las rebanadas de simultaneidad, podemos decir que en el espacio-tiempo galileano cualquier par de relojes de precisión que partan juntos y terminen juntos habrán de mostrar el mismo tiempo transcurrido entre los dos eventos. Pero como se muestra en la paradoja de los gemelos, no es éste el caso en el espacio-tiempo de Minkowski.

Los cuerpos en movimiento inercial, los rayos de luz en un vacío y los relojes ideales, todos ellos sirven como instrumentos que hacen visibles ciertos aspectos de la geometría del espacio-tiempo de Minkowski. Utilizando precisamente estos tipos de objetos, podemos plantear y resolver problemas en la relatividad especial. Unos cuantos problemas de este tipo aparecen en el apéndice de este libro. Los problemas fueron diseñados para poner énfasis en la geometría fundamental del espacio-tiempo de Minkowski y es posible solucionarlos utilizando únicamente la ecuación IV.1. (En particular, es posible solucionar fácilmente estos problemas sin la ayuda de las transformaciones lorentzianas y por lo tanto tienen un carácter diferente de la mayoría de los problemas que aparecen en los manuales de física.) No hay mejor manera de llegar a apreciar el carácter del espacio-tiempo de Minkowski que la que se deriva de solucionar problemas de esta especie, y para hacerlo sólo es necesario tener conocimientos básicos de álgebra. Por lo tanto, al lector le podría ser provechoso hacer una pausa aquí para darles un vistazo a estos problemas.

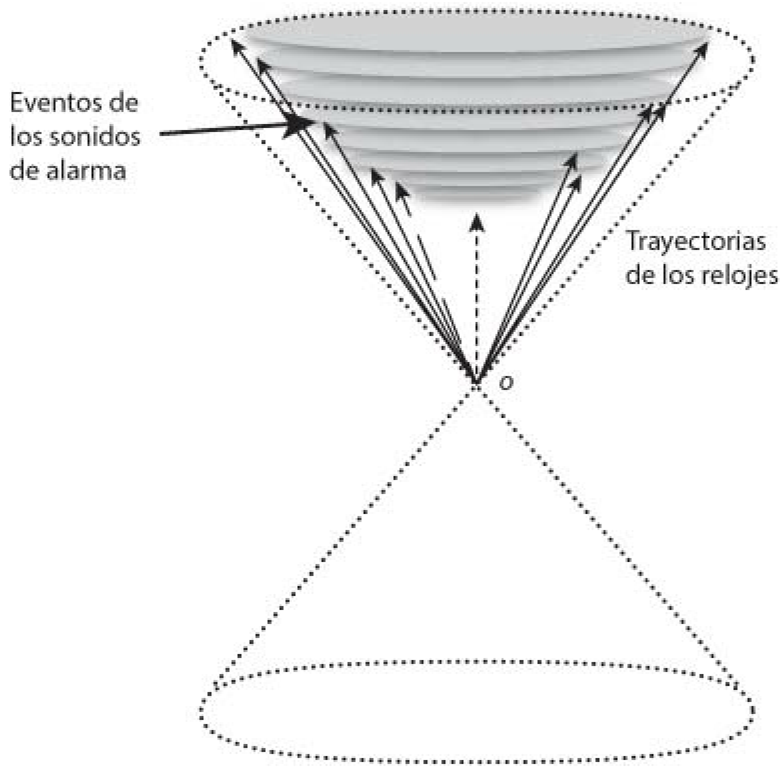

FIGURA IV.5

La utilización de los rayos de luz en un vacío y de los relojes en movimiento inercial nos permite reflexionar acerca de cómo construir sistemas de coordenadas en nuestro espacio-tiempo. Puesto que los relojes y los rayos de luz son visibles, estos sistemas de coordenadas no son meramente abstractos: tienen una significancia de tipo observacional.

CONSTRUCCIÓN DE LAS COORDENADAS DE LORENTZ

Hasta aquí la utilización que hemos hecho de las coordenadas de Lorentz ha sido enteramente abstracta: todo lo que hemos aseverado es que dada la geometría del espacio-tiempo de Minkowski, *existen* asignaciones de cuádruplos ordenados de números reales a eventos que tienen una relación especificada con la geometría intrínseca del espacio-tiempo. No hemos indicado cómo sería posible construir de manera práctica estos sistemas de coordenadas. De hecho, no sería posible describir una semejante realización práctica sin algunos principios que conecten la geometría del espacio-tiempo con cuerpos observables. Dadas la ley de la luz, la ley relativista de la inercia y la hipótesis del reloj, estamos ahora en una posición que nos permite indicar cómo instaurar un sistema de coordenadas visible.

Nuestra primera tarea consiste en asignar una coordenada t a cada uno de los eventos. Empezamos por suponer que tenemos una cantidad ilimitada de relojes ideales y que podemos saber si un reloj no está sujeto a cualquier tipo de fuerzas externas. Todos los relojes en cuestión tendrán, de acuerdo con la ley relativista de la inercia, trayectorias rectas de tipo tiempo. Elijamos *arbitrariamente* uno de estos relojes como el "reloj regidor" del sistema de coordenadas. Esta elección arbitraria habrá de determinar muchos de los aspectos sobresalientes de las coordenadas resultantes. El reloj regidor tendrá algún tipo de escala temporal arbitrariamente elegida (digamos, los segundos) y una "hora" cero arbitrariamente elegida. Dado esto, a todos los eventos en la línea de mundo del reloj regidor se les asignará una coordenada *t* por el reloj.

La proyección de la coordenada *t* a los eventos en el exterior de la trayectoria del reloj regidor requiere la utilización de otros instrumentos, incluyendo otros relojes. Pero debemos utilizar estos instrumentos con mucha cautela. En vista del tipo de comportamiento que aparece en la paradoja de los gemelos, no podemos simplemente colocar y sincronizar algunos relojes auxiliares junto al reloj regidor para luego lanzarlos a través del espacio: las trayectorias específicas que tomen desde la proximidad del reloj regidor hasta los sitios distantes tendrá una influencia en sus lecturas. De manera que es preferible suponer que existen relojes en todas las trayectorias rectas de tipo tiempo en el espaciotiempo de Minkowski para identificar una colección de relojes que estén en *co-movimiento*.

Intuitivamente, dos relojes están en co-movimiento si ambos se encuentran en trayectorias inerciales y no se aproximan el uno al otro ni se alejan el uno del otro. Es importante que las trayectorias sean inerciales: aunque los relojes que giren en torno a un centro común puedan mantener una separación constante, no es posible considerar que están en co-movimiento. No hay ninguna garantía, ya sea analítica o ya sea conceptual, de que puedan siquiera existir los relojes en co-movimiento: en algunos espacio-tiempos, ningún par de trayectorias inerciales puede mantener una separación constante. Pero las simetrías del espacio-tiempo de Minkowski permiten una fértil estructura de relojes en co-movimiento.

Un observador situado junto al reloj regidor es capaz de identificar un reloj inercial en co-movimiento mediante la utilización de una especie de radar. Es decir, el observador lanza rayos de luz desde el reloj regidor y luego observa cuánto tiempo les toma (de acuerdo con el reloj regidor) a los rayos de luz ir hasta un blanco, otro reloj donde se reflejan, y regresar. Resulta especialmente sencillo mostrar los rayos de luz en nuestro diagrama porque la trayectoria de un rayo de luz en un vacío siempre se representa por una línea recta con un ángulo de 45° en el diagrama. Si el reloj que sirve de blanco está en co-movimiento, el tiempo del viaje de ida y vuelta de la luz siempre será el mismo; si el reloj-objetivo se encuentra en movimiento relativo, entonces los tiempos del viaje de ida y vuelta habrán de variar (figura IV.6). Con el fin de que nuestros diagramas de espacio-tiempo sean más legibles, los hemos de restringir a sólo dos dimensiones del espacio-tiempo, pero la misma técnica funcionaría en cualquier dirección.

De esta forma, el observador junto al reloj regidor puede identificar los relojes en co-movimiento. El reloj-objetivo 1 está en co-movimiento con el reloj regidor, ya que cada viaje de ida y vuelta de la luz toma dos minutos. El reloj-objetivo 2, al contrario, no está en co-movimiento: el primer viaje de ida y vuelta toma menos de un minuto y los posteriores más de un minuto. En el espacio-tiempo de Minkowski, un grupo completo de relojes en co-movimiento llena todo el espacio-tiempo, seccionándolo como un conjunto de trayectorias rectas paralelas de tipo tiempo. Únicamente un reloj del grupo se encuentra presente en cada evento, y nuestro objetivo consiste en que la coordenada *t* de ese evento se pueda leer en el reloj apropiado. Pero antes de que podamos hacer esto, tenemos que calibrar y sincronizar los relojes en co-movimiento. La calibración, es decir, la elección de una unidad de medición, es fácil. Dejemos que el reloj regidor emita un pulso cada minuto, y luego que los relojes en comovimiento ajusten sus calibraciones de manera que una unidad de tiempo transcurra entre la llegada de las señales sucesivas. Ahora todos los relojes marchan "al unísono".

El último paso, la sincronización, es el que ha interesado más a los académicos. En el espacio-tiempo clásico, tal como en el espacio-tiempo galileano, es obvio en qué consiste la sincronización de los relojes: todos deben asignar el mismo tiempo a los eventos en la misma rebanada de simultaneidad. Puesto que en un espacio-tiempo clásico hay un hecho físico objetivo sobre cuáles eventos se dan al mismo tiempo (es decir, en el mismo momento del tiempo absoluto), tiene sentido exigir que los relojes se ajusten de manera que asignen la misma coordenada de tiempo a los eventos simultáneos. Pero en el espacio-tiempo de Minkowski no existe ese tipo de estructura objetiva de simultaneidad que los relojes puedan reflejar. No hay ningún hecho físico respecto a la cuestión de si dos eventos separados de tipo espacio (es decir, eventos fuera de los conos de luz de cada uno) "suceden en el mismo instante". En este sentido, la sincronización de los relojes en la relatividad requiere de una convención.

Dada la geometría del espacio-tiempo de Minkowski, algunos métodos convencionales son más sencillos y más naturales que otros. En la figura IV.6 los observadores son capaces de determinar mediante el método del radar que el reloj-objetivo 1 está en co-movimiento con el reloj regidor, mientras que el reloj-objetivo 2 no lo está. Pueden determinar además que el tiempo del viaje de ida y vuelta de la luz entre ambos relojes es siempre de dos minutos. Es muy natural que el reloj-objetivo 1 se ajuste de modo que cuando un rayo de luz le llegue del reloj regidor marque un minuto más de lo que el reloj regidor marcaba cuando se emitió el rayo de luz. Si se adopta este método convencional, el reloj-objetivo 1 marcará 12:01 al recibir el primer rayo de luz, 12:03 al recibir el segundo, etc. Tras ajustar cada reloj del grupo mediante este método convencional damos término a nuestra tarea: ahora a cada evento en la historia del universo se le ha asignado una coordenada *t*. En la figura IV.6, el conjunto de eventos al que se asigna la misma coordenada *t* en este sistema de coordenadas habrá de formar una línea horizontal o un plano horizontal que se ve exactamente como las rebanadas de simultaneidad en la figura III.3. Pero puesto que no hay una relación física de simultaneidad entre los eventos, y puesto que esta forma en particular de asignar coordenadas *t* es arbitraria en muchos sentidos (incluyendo la elección de un reloj regidor y la elección de un método convencional para ajustar los relojes), las llamaremos *rebanadas de* t *equivalente*, en vez de rebanadas de simultaneidad.

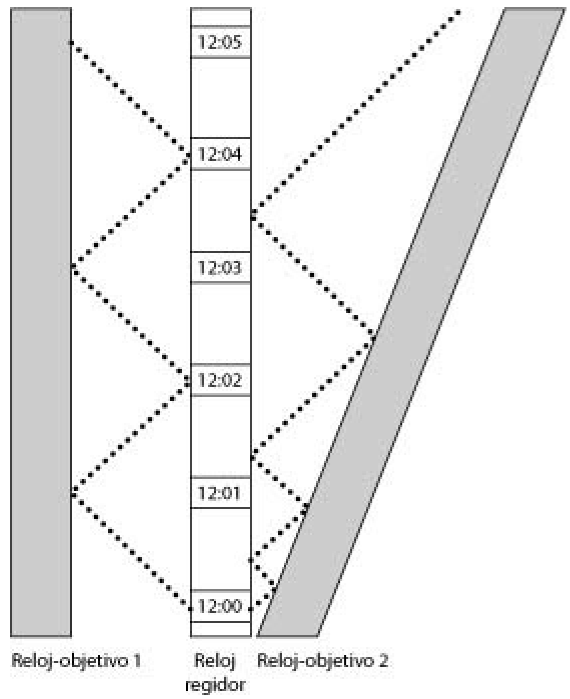

FIGURA IV.6

¿Qué habría pasado si hubiéramos elegido un reloj regidor diferente? En específico, ¿qué si hubiéramos escogido el reloj-objetivo 2 de la figura IV.6 con el fin de que fungiera como un reloj regidor, identificado los relojes en co-movimiento en relación con él, y luego ajustado la calibración de acuerdo con el mismo método? Se muestra el resultado de esta tarea en la figura IV.7, donde el reloj-objetivo 1 se sincronizó con el reloj regidor cuando estuvieron adyacentes a las 12:00. Es notorio en particular que las rebanadas de *t* equivalente en estas nuevas coordenadas se encuentran inclinadas con respecto a las rebanadas de *t* equivalente en el sistema original. En la figura IV.7, la línea horizontal indica todos los eventos a los cuales se les asignó la hora 12:02 en el sistema del reloj regidor, y la línea inclinada, todos los eventos a los cuales se les asignó la hora 12:02 en el sistema del reloj-objetivo 2. Si la "simultaneidad de las coordenadas" —es decir, la asignación de la misma coordenada *t* en algún sistema de coordenadas— supuestamente se correspondiera con cierta relación física real entre los eventos, a lo sumo tan sólo uno de estos sistemas de coordenadas sería correcto. Pero la inexistencia de cualquier tipo de relación de simultaneidad objetiva significa que las coordenadas *t'* no están ni mejor ni peor adaptadas a la geometría objetiva que las coordenadas *t*.

En nuestros diagramas euclidianos del espacio-tiempo, las trayectorias de los relojes en co-movimiento se relacionan con sus rebanadas de *t* equivalente de la siguiente manera: al inclinarse las trayectorias de los relojes en el diagrama, las rebanadas de *t* equivalente asociadas también se inclinan para que los rayos de luz siempre dividan la diferencia entre las trayectorias de los relojes y las rebanadas de *t* equivalente. En otras palabras, si la pendiente del reloj regidor en el diagrama es *s*, la pendiente de superficie de *t* equivalente de la coordenada del tiempo de ese reloj será 1/*s*. Como consecuencia, en el diagrama el ángulo entre un rayo de luz y la trayectoria de un reloj siempre es igual al ángulo entre el rayo de luz y la rebanada de *t* equivalente del reloj. Esto es meramente una consecuencia del método convencional que se utilizó para dibujar los diagramas.

Al hecho de que las rebanadas de *t* equivalente sean diferentes de las rebanadas de *t'* equivalente se le denomina *relatividad de la simultaneidad*. Se dice comúnmente que en la relatividad la noción de la simultaneidad es relativa, puesto que depende del observador o del estado de movimiento. Hay una pizca de verdad en esta caracterización, pero es posible que sea tan dañina como benéfica. La premisa clave de la relatividad es la *inexistencia* de la simultaneidad en cuanto relación física real entre los eventos. Además, la "relatividad de la simultaneidad" es una circunstancia muy derivativa, muy basada en las coordenadas. Los diferentes sistemas de coordenadas emplean diferentes coordenadas *t*, y sus superficies de valor *t* constante son diferentes. Las coordenadas de Lorentz constituyen tan sólo un subconjunto específico de todos los sistemas de coordenadas posibles, el cual se distingue por una cierta sencillez de definición en el espacio-tiempo de Minkowski. Pero en la relatividad general, como veremos, las coordenadas de Lorentz por lo general no existen y los procedimientos prácticos arriba descritos no logran definir ningún sistema de coordenadas. (Ya se mencionó que en la relatividad general los relojes en co-movimiento inercial generalmente no existen, y por lo tanto nuestro método falla en el primer paso.) El fenómeno de los gemelos, al contrario, es un efecto físico evidente que es posible describir sin la utilización de ningún tipo de sistema de coordenadas y que sigue existiendo en la relatividad general.

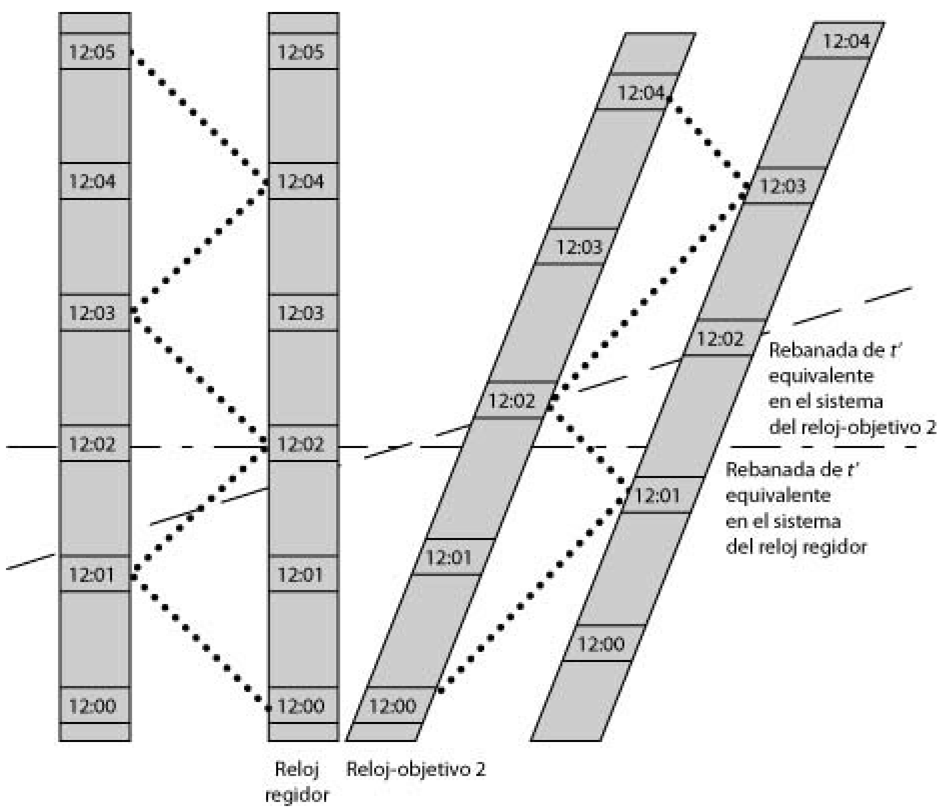

FIGURA IV.7

Nuestras dos coordenadas *t* también son una demostración de la *dilatación temporal basada en las coordenadas*. La dilatación temporal relativista a veces se describe diciendo que los "relojes en movimiento son más lentos", pero ya sabemos que esta frase no tiene sentido: en la relatividad, no existe ninguna diferencia entre los relojes en movimiento y los relojes que no están en movimiento, y todos los relojes ideales de idéntica estructura, por definición, miden los segundos exactamente al mismo tiempo en proporción al intervalo a lo largo de sus líneas de mundo. Pero si examinamos la coordenada *t* que resulta de utilizar el reloj regidor en nuestro procedimiento en comparación con la coordinada t′ que resulta de utilizar en su lugar el reloj-objetivo 2, podemos ver un fenómeno interesante. El reloj regidor y el reloj-objetivo 2 se encuentran en sincronía cuando se encuentran adyacentes a las 12:00. Pero el reloj-objetivo 2 indica 12:02 en un punto que en el sistema del reloj regidor está *arriba* de la superficie de 12:02. En este sentido, *con respecto a las coordenadas* t *del reloj regidor*, "es más lento" el reloj-objetivo 2. Sin embargo, la situación es perfectamente simétrica: el reloj regidor indica 12:02 arriba de la superficie de 12:02 de las coordenadas del reloj-objetivo 2, por lo que, *en relación con las coordenadas del reloj-objetivo 2*, el reloj regidor "es más lento". Claro, no existe ningún hecho sobre cuál de los relojes es "realmente" más lento: es ésta una situación que depende totalmente de las coordenadas. Sin embargo, sí existe un hecho sobre cuál de los gemelos en la paradoja de los gemelos termina siendo más joven, y esta situación no se relaciona en absoluto con la existencia de coordenadas. Poner en el mismo saco el fenómeno perfectamente objetivo de los gemelos y la "dilatación temporal" basada en coordenadas es una buena manera de sembrar la confusión.

Hasta aquí hemos descrito una técnica física para la asignación de coordenadas *t* a todos los eventos en el espacio-tiempo de Minkowski relacionados con un reloj inercial arbitrariamente elegido. Para completar nuestro sistema de coordenadas, también tenemos que asignar las coordenadas *x*, las coordenadas *y* y las coordenadas *z*. Tal cosa obviamente exige una elección arbitraria de direcciones *x, y* y *z* ortogonales: podemos imaginar que en nuestro reloj regidor se conectan de algún modo tres flechas ortogonales. El reloj regidor no debe estar girando, pero, como ya Newton lo había señalado, esto puede determinarse empíricamente. El reloj regidor sirve como el origen espacial de nuestras coordenadas, de modo que sus coordenadas *x, y* y *z* estarán todas en 0. A todos los demás relojes en co-movimiento también se les deben asignar ciertas coordenadas "espaciales". Veamos cómo hacer esto a lo largo del eje *x* de nuestro sistema de coordenadas: a los demás ejes se les tratará de la misma manera.

Puesto que el reloj regidor tiene una flecha conectada que apunta en la dirección *x*, puede emitir señales luminosas en esa dirección, las cuales serán captadas por sólo algunos de los relojes en co-movimiento. Esos relojes ocuparán el eje *x* positivo de nuestro sistema de coordenadas: su coordenada *y* y su coordenada *z* serán 0. La emisión de luz en la dirección opuesta señala el eje *x* negativo. La única cuestión que queda pendiente es cuál valor de coordenada x se le debe asignar a cada uno de estos relojes.

El único hecho de que dispone el reloj regidor es el tiempo que le tomará a una señal de luz llegar hasta un reloj en co-movimiento y regresar. En la figura IV.7, el tiempo del viaje de ida y vuelta hasta el reloj en co-movimiento indicado es de dos minutos invariablemente. Así, resulta natural hacer que la coordenada x de ese reloj, que se corresponde con su "distancia espacial" del reloj regidor, sea proporcional al tiempo del viaje de ida y vuelta. Aún más sencillamente, podemos decir que puesto que el viaje de ida y vuelta toma dos minutos, el reloj en co-movimiento está a un minuto-luz de distancia. Por lo tanto, al reloj en co-movimiento cuyo viaje redondo toma dos minutos y que se encuentra en la dirección x positiva en relación con el reloj regidor se le asigna el valor de +1 para la coordenada x, y el reloj en comovimiento cuyo viaje redondo toma dos minutos y que se encuentra en la dirección x negativa se le asigna el valor de −1 para la coordenada x. Siguiendo este esquema, a cada uno de los relojes en co-movimiento que se encuentre en el eje x se le asigna una coordenada x, y lo mismo se hace con los otros ejes. La repetición del procedimiento respecto a todos los relojes en los ejes finalmente resulta en que a todos y cada uno de los relojes en co-movimiento se les asigne la coordenada x, la coordenada y y la coordenada z, completando así nuestro sistema de coordenadas.

He aquí que mediante este método hemos podido construir un sistema de coordenadas de Lorentz en el espacio-tiempo de Minkowski. Las diferentes opciones de relojes regidores, unidades de tiempo, origen del tiempo y orientación de los ejes espaciales dan como resultado coordenadas de Lorentz diferentes, y cada conjunto de coordenadas de Lorentz se deriva de una de estas opciones.

Aquí debemos aclarar la situación lógica. Cuando introdujimos la noción de un sistema de coordenadas de Lorentz, no lo asociamos en absoluto con cualquier tipo de procedimiento físico: las coordenadas se utilizaron sólo como una forma abstracta de especificar la geometría intrínseca del espacio-tiempo de Minkowski. En seguida ligamos la geometría con el comportamiento de la materia mediante un conjunto de principios físicos: la ley de la luz, la ley relativista de la inercia y la hipótesis del reloj. Por último, demostramos que si estos principios se aceptaran, entonces un cierto procedimiento físico que utiliza relojes ideales en movimiento inercial y rayos de luz en un vacío habría de resultar en la asignación de coordenadas de Lorentz al espacio-tiempo de Minkowski. En ningún punto de este procedimiento hemos siquiera mencionado la "velocidad de la luz" o postulado que la "velocidad de la luz es constante": en el espacio-tiempo de Minkowski no es posible realizar ninguna medida objetiva respecto a la velocidad de cualquier cosa. Y tampoco hemos invocado en ningún momento la noción de un "sistema de coordenadas inercial", o postulado que "todos los sistemas inerciales son equivalentes" o que "las leyes de la física adoptan la misma forma en todos los sistemas inerciales". Más bien, hemos propuesto una cierta estructura geométrica del espacio-tiempo, proporcionándole a esta estructura mediante unos postulados físicos una significancia física ligada al comportamiento de la materia visible y, finalmente, hemos descrito cómo emplear la materia para construir sistemas de coordenadas.

Habiendo hecho todo esto, podemos ahora comprender qué se da a entender por “la constancia de la velocidad de la luz”, por un “sistema de coordenadas inercial” y por la “equivalencia de todos los sistemas inerciales”.

La luz, en sí misma, carece de velocidad, ya que no existe ni el tiempo absoluto ni el espacio absoluto en la relatividad. Pero *en relación con un sistema de coordenadas*, podemos asignarle a un rayo de luz una *velocidad coordenada*. Por ejemplo, en la figura IV.7, ¿a qué “velocidad” viaja el rayo de luz desde el reloj regidor al reloj en co-movimiento? Pues bien, *en términos del sistema de coordenadas sin prima*, al emitirse desde el reloj regidor el valor t es 12:00 y al llegar al reloj en co-movimiento es 12:01, de manera que la diferencia en valores t (el “tiempo transcurrido” en este sistema de coordenadas) es de un minuto. Es de notar que el “tiempo transcurrido” no se mide por ningún reloj individual: depende de nuestros procedimientos para ajustar los relojes en co-movimiento. De modo que en un sentido dependiente de las coordenadas, al rayo de luz le tomó un minuto para llegar desde el reloj regidor hasta el reloj en co-movimiento. ¿Qué distancia cubrió el rayo de luz al realizar ese viaje? Nuevamente, no hay una respuesta *objetiva* para esta pregunta: no tenemos nada parecido al espacio absoluto persistente de Newton. Pero la coordenada x del reloj en co-movimiento en el sistema de coordenadas del reloj regidor es -1, y tanto su coordenada y como su coordenada z son 0. Así (utilizando la ecuación pitagórica), podemos decir que el reloj en co-movimiento se encuentra a un minuto-luz de distancia. El rayo de luz viaja un minuto-luz (en estas coordenadas) en un minuto, de modo que su *velocidad de coordenada* en este sistema de referencia es de un minuto-luz por minuto.

Ahora bien, todos estos cálculos le podrían parecer al lector una especie de broma. Dada la forma en que asignamos las coordenadas, es *obvio* que la “velocidad de coordenada” de la luz es un minuto-luz por minuto: si se mide el tiempo en minutos y se asignan las distancias en minutos-luz, entonces *por definición* un rayo de luz cubrirá un minutoluz por minuto. Y en este sentido estos cálculos son una broma: el resultado ya estaba implantado en el procedimiento de la asignación de las coordenadas. Pero no es un mal resultado. Puesto que hemos estado intentando deshacernos del espacio y el tiempo absolutos newtonianos, también hay que desechar las velocidades absolutas newtonianas. Esto es válido para la luz tanto como para cualquier otra cosa. En este sentido, “la constancia de la velocidad de la luz” no puede ser un principio físico fundamental.

Sin embargo, no debemos ir demasiado lejos en la otra dirección. La velocidad de la luz coordenada es constante en todos los sistemas de coordenadas lorentzianos, y la velocidad de la luz coordenada *no* es constante en otros sistemas de coordenadas que sea posible definir en el espacio-tiempo minkowskiano. (Por ejemplo, simplemente tratemos de combinar la coordenada t de un sistema lorentziano con las coordenadas z, las coordenadas x y las coordenadas y de otro sistema.) Pero el hecho de que las coordenadas de Lorentz, con sus relaciones con el comportamiento de la luz y de los relojes, sean *siquiera* posibles no es una cuestión de convención. El hecho de que toda la luz que se emite desde un evento (en un vacío) se propague a lo largo de un cono de luz no es una cuestión convencional. La existencia de relojes en co-movimiento, tal como nosotros los hemos definido, no es una cuestión convencional. La propuesta de un espacio-tiempo minkowskiano es una tesis física, no una convención.

Como analogía “espacial” con la dilatación temporal basada en coordenadas, las coordenadas de Lorentz también muestran una *contracción Lorentz-FitzGerald basada en coordenadas*. Refiriéndonos una vez más a la figura IV.7, sabemos que en las coordenadas del reloj regidor, el reloj en co-movimiento de la izquierda tiene coordenadas constantes *x, y* y *z* iguales a −1, 0 y 0, respectivamente. Así, en estas coordenadas el reloj del lado izquierdo mantiene una “separación espacial” constante de un minuto-luz respecto al reloj regidor. Con base en la paridad de razonamiento, el reloj del lado derecho mantiene una separación constante de un minuto-luz respecto al reloj-objetivo 2 en el marco lorentziano asociado con esos relojes. Pero esto deja totalmente abierta la cuestión de cuál es la separación “espacial” entre el reloj-objetivo 2 y el correspondiente reloj en comovimiento *en las coordenadas del reloj regidor*. Obviamente, estos dos relojes mantienen una separación constante entre sí: sus líneas de mundo son paralelas. Pero no es evidente cuál es exactamente esa separación en el marco de referencia del reloj regidor.

Disponemos de la suficiente información como para solucionar este problema una vez que le atribuyamos una trayectoria definitiva al reloj-objetivo 2. Supongamos que al igual que el gemelo A en la primera parte de su viaje, el reloj-objetivo 2 viaja inercialmente desde el origen a través del evento con las coordenadas (5, 4, 0, 0) en las coordenadas del reloj regidor. (El reloj marcará 12:05 en vez de cinco porque lo echamos a andar a las 12:00, pero utilizaremos coordenadas más sencillas.) De modo que en las coordenadas del reloj regidor, el reloj-objetivo 2 cubre cuatro minutos-luz en cinco minutos: se mueve a 80% de la velocidad de la luz. Utilizando estas coordenadas de Lorentz como coordenadas cartesianas en el diagrama de espacio-tiempo, podemos convertir nuestro problema en uno de la geometría euclidiana. Puesto que las coordenadas *y* y las coordenadas *z* de todos los relojes siempre son 0, las dejaremos fuera de la explicación. La trayectoria del reloj-objetivo 2 se corresponde con la ecuación $x = \frac{4}{5}t$. Es decir, los eventos en la trayectoria del reloj-objetivo 2 todos tienen coordenadas *x* y coordenadas *t* que solucionan esta ecuación. La trayectoria del reloj del lado derecho se describe similarmente con la ecuación $x = \frac{4}{5}t + x_0$, donde x^0 es el valor *x* del punto donde su trayectoria se encuentra con el eje $t = 0$. Obsérvese que estas dos ecuaciones describen líneas con la misma pendiente y por lo tanto líneas paralelas. El rayo de luz que surge del origen y se mueve hacia la derecha se describe por la ecuación $x = t$. Puesto que los rayos de luz siempre se representan por líneas de 45°, sus pendientes son siempre + 1 o −1. Puesto que tenemos una ecuación para el reloj de la derecha y una ecuación para el rayo de luz que se emite desde el origen, podemos solucionar para las coordenadas del evento en que este rayo de luz se encuentra con el reloj que está más a la derecha. Utilizando $x = t$ eliminamos *x* de la otra ecuación y obtenemos:

$$x = \tfrac{4}{5}t + x_0$$
$$\tfrac{1}{5}t = x_0$$
$$t = 5x_0.$$

Puesto que $x = t$, el valor t de este evento es también su valor x. Así que las coordenadas del evento donde el rayo de luz impacta el reloj son $(5x_0, 5x_0)$.

Ahora que tenemos una ecuación para el punto donde el rayo de luz se refleja, podemos encontrar una ecuación para el rayo reflejado. La trayectoria del rayo de luz reflejado de vuelta desde el reloj más a la derecha hasta el reloj-objetivo 2 tiene una pendiente de −1 y se origina en $(5x_0, 5x_0)$, de manera que su ecuación es $x = -t + 10x_0$. (Esta es la única línea con una pendiente de −1 que pasa a través de $(5x_0, 5x_0)$.) El evento en que este rayo de luz se encuentra con el reloj-objetivo 2 debe satisfacer tanto esta ecuación como la ecuación para el reloj-objetivo 2. Debemos solucionar simultáneamente la ecuación $x = -t + 10x_0$ y la ecuación $x = \tfrac{4}{5}t$. Utilizando la segunda ecuación para eliminar x de la primera, derivamos

$$\tfrac{4}{5}t = -t + 10x_0$$
$$\tfrac{9}{5}t = 10x_0$$
$$t = \tfrac{50}{9}x_0.$$

La coordenada t del evento donde el rayo de luz que regresa se encuentra con el reloj-objetivo 2 es $\tfrac{50}{9}x_0$. Puesto que sabemos que $x\ \tfrac{4}{5}t$, las coordenadas son $(\tfrac{50}{9}x_0, \tfrac{40}{9}x_0)$. Pero tenemos un hecho adicional. En el momento en que el rayo de luz que retorna impacta el reloj-objetivo 2, el reloj muestra dos minutos de tiempo transcurrido desde el origen. Ya que, según la hipótesis del reloj, ésta es una medida del intervalo a lo largo de esa trayectoria, tenemos $\sqrt{(\tfrac{50}{9}x_0)^2 - (\tfrac{40}{9}x_0)^2} = 2.$ Cuadrando ambos lados y solucionando para x_0:

$$4 = \tfrac{2500}{81}x_0^{\,2} - \tfrac{1600}{81}x_0^{\,2} = \tfrac{900}{81}x_0^{\,2}$$
$$2 = \tfrac{30}{9}x_0$$
$$x_0 = \tfrac{18}{30} = \tfrac{3}{5}.$$

El punto donde el reloj más a la derecha se encuentra con el eje $t = 0$ tiene las coordenadas $(0, \tfrac{3}{5}, 0, 0)$ en el marco del reloj regidor.

Esto significa que *con relación a las coordenadas del reloj regidor*, los dos relojes en co-movimiento se encuentran a una distancia constante entre sí $\tfrac{3}{5}$ de minutos-luz, mientras que las coordenadas del reloj-objetivo 2 siempre se encuentran a una distancia de un minuto-luz entre sí. Y esta circunstancia es simétrica: en las coordenadas del reloj-objetivo 2, el reloj regidor y su acompañante en co-movimiento sólo se encuentran a $\tfrac{3}{5}$ minutos-luz entre sí. Esta relación de los diferentes sistemas de coordenadas de Lorentz es tan sólo una manera de entender la llamada contracción Lorentz-FitzGerald.

La contracción Lorentz-FitzGerald basada en coordenadas no es, en sentido cabal, la contracción física de cosa alguna. En nuestro ejemplo, todos los relojes siempre se muéven inercialmente: nada está sujeto a ningún tipo de fuerzas y nada se "contrae". Lo único que hemos señalado es un hecho con relación a las coordenadas asignadas a diversos eventos en los sistemas de coordenadas que han sido construidos mediante ciertas reglas. Esta manera de entender el "efecto" la recomienda, por ejemplo, Rindler:

> Según Lorentz, el mecanismo responsable de la contracción era un cierto incremento en las fuerzas eléctricas cohesivas que fortalecían la estructura atómica. La relatividad, por otra parte, pasa por alto todo tipo de explicación en función de las fuerzas o algo semejante, y sin embargo predice la inevitabilidad de los fenómenos [...] En la relatividad el efecto es esencialmente un efecto basado en la "proyección" geométrica, muy análoga a mirar una barra *estacionaria* que no se encuentra paralela al plano de la retina. El hecho de impartir una velocidad uniforme en la relatividad equivale a realizar una seudo-rrotación en el "espacio-tiempo".[6]

Pero de la misma manera que existe una dilatación temporal basada en coordenadas, la cual no es un efecto físico, y que existe el fenómeno de los gemelos, que en nada se relaciona con las coordenadas, también existe la contracción Lorentz-FitzGerald basada en coordenadas que Rindler describe y además una contracción *física* Lorentz-FitzGerald que de hecho se puede explicar "en función de las fuerzas o algo semejante". La incomprensión de estos dos "efectos" sólo puede crear confusiones respecto a la relatividad y a las determinaciones experimentales de la "velocidad de la luz". Dedicaremos el próximo capítulo a la aclaración de este asunto.

Antes de examinar ese difícil tema, desarrollemos un sencillo problema más que ilustra la relación entre los sistemas de coordenadas, así como el rol que juegan nuestros principios físicos. Supongamos que configuramos un par de marcadores en movimiento inercial que se encuentren, en cualquier sistema de coordenadas lorentzianas en que estén en reposo, a una distancia de un minuto-luz entre sí. Un reloj en movimiento inercial rebasa primero uno de los marcadores y luego el otro. *Según el reloj*, transcurre exactamente un minuto entre el evento en que se encuentra con el primer marcador y el evento en que se encuentra con el segundo marcador. Preguntas: *En el marco de referencia de los marcadores*, ¿cuánto tiempo transcurre entre estos dos eventos? En este marco de referencia, ¿a qué velocidad se mueve el reloj? *En el marco de referencia del reloj*, ¿a qué distancia se encuentran los marcadores entre sí?

La manera más fácil de resolver estos problemas es mediante un conveniente diagrama de espacio-tiempo. Cuando se intenta determinar la forma en que se describe una situación en un sistema de coordenadas de Lorentz, el diagrama de espacio-tiempo más conveniente es uno en que las rebanadas t constantes son horizontales y las rebanadas x constantes son verticales: estas coordenadas de Lorentz en el espacio-tiempo se corresponden con las coordenadas cartesianas en el diagrama. Si hacemos del evento donde el reloj se encuentra con el primer marcador el origen de las coordenadas, obtendremos el diagrama para el marco de referencia de los marcadores que se muestra en la figura IV.8. Queremos calcular t_m, la coordenada t que las coordenadas lorentzianas de los marcadores asignan al evento donde el reloj se encuentra con el marcador 2.

No debemos llegar apresuradamente a la tentadora conclusión de que si el reloj viaja un minuto-luz en un minuto (como lo señala el reloj mismo), entonces debe estar viajando a la velocidad de la luz. La razón está en que lo que marca el reloj no es lo mismo que la coordenada *t* en el marco de referencia del marcador. Sin embargo, el problema se volverá trivial una vez que recordemos que un reloj ideal —cualquier tipo de reloj ideal en cualquier tipo de trayectoria— mide el intervalo a lo largo de su línea mundial, y que recordemos también en qué forma se expresa el intervalo en términos de las coordenadas de Lorentz. El intervalo de (0, 0) hasta (t_m, 1) a lo largo de una trayectoria recta es

$$\sqrt{(t_\mathrm{m} - 0)^2 - (1 - 0)^2} = \sqrt{(t_\mathrm{m}^2 - 1}.$$

De acuerdo con la medición del reloj, sabemos que este mismo intervalo es 1. De manera que nuestra ecuación completa es

$$\begin{aligned} &\sqrt{t_\mathrm{m}^2 - 1 = 1} \\ &t_\mathrm{m}^2 - 1 = 1 \\ &t_\mathrm{m}^2 = 2 \\ &t_\mathrm{m}^2 = \sqrt{2}. \end{aligned}$$

En el marco de referencia del marcador, el reloj toma cerca de 1.414 minutos para viajar entre los marcadores, y por lo tanto su velocidad de coordenada en ese marco de referencia es casi 70.7% de la velocidad de la luz. A juzgar por ese marco, el reloj en “movimiento” “avanza lentamente”: sólo registra que ha transcurrido un minuto. Este registro constituye la dilatación temporal que depende de las coordenadas.

Si queremos hacer lo mismo desde el punto de vista del reloj (es decir, las coordenadas de Lorentz en que el reloj se encuentra en reposo), entonces debemos reconfigurar el mismo diagrama de espacio-tiempo haciendo que la trayectoria del reloj sea vertical (figura IV.9). Las coordenadas en la figura IV.9 ahora se encuentran en el marco lorentziano del reloj (es decir, el marco lorentziano donde el reloj se encuentra en reposo en el origen espacial), en vez de en el marco de referencia de los marcadores. Todos los eventos son exactamente los mismos, pero se les asignan valores de coordenadas diferentes. Quisiéramos determinar x_n, la cual es la distancia coordenada constante entre los marcadores en este marco.

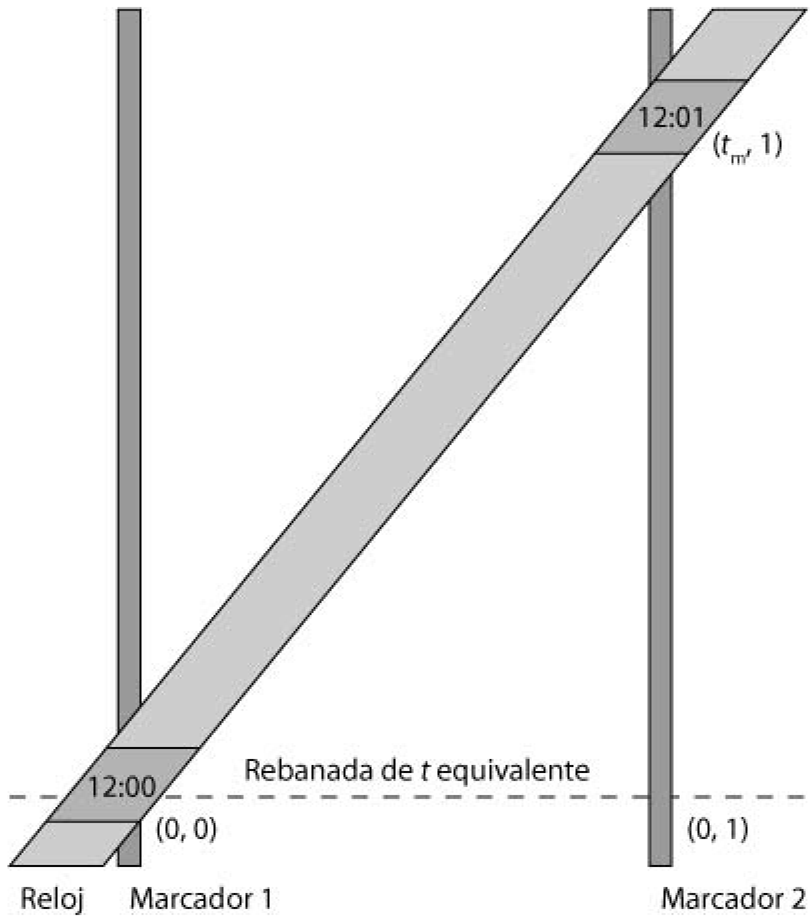

FIGURA IV.8

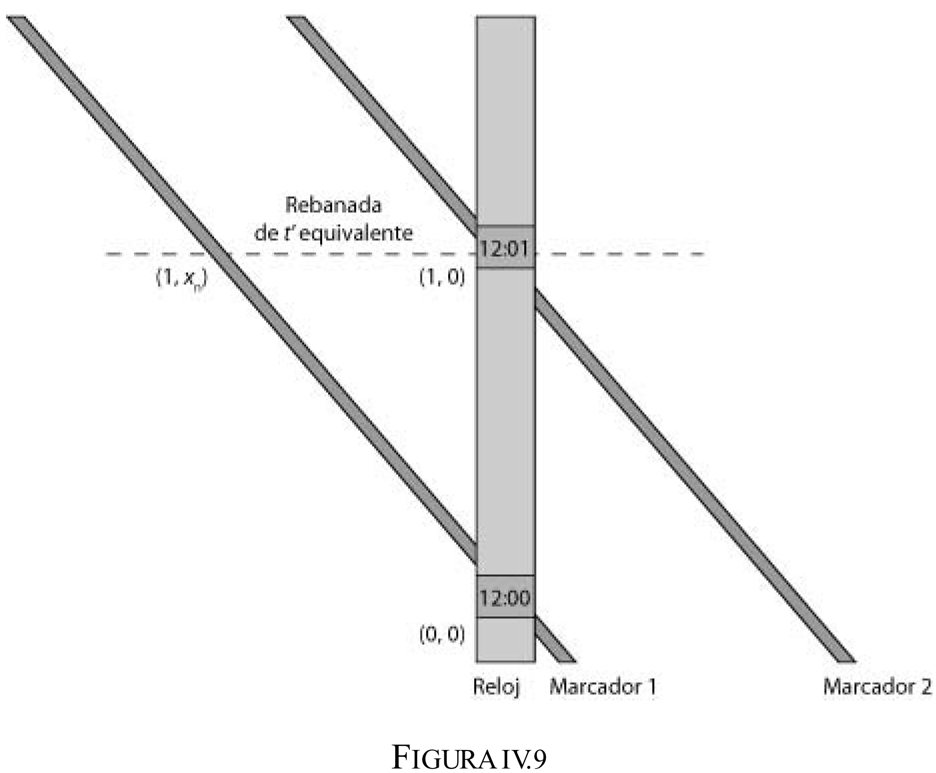

FIGURA IV.9

La derivación más sencilla utiliza los resultados de nuestra última operación, junto con el hecho de que si la velocidad de A en el marco del reposo de B es **v**, entonces la velocidad de B en el marco de reposo de A es − **v**. De manera que si el reloj está viajando a 70.7% de la velocidad de la luz en el marco de reposo de los marcadores, entonces los marcadores están viajando a 70.7% de la velocidad de la luz (en la dirección opuesta) en el marco de reposo del reloj. Pero entonces el marcador 1, un minuto después de que haya pasado al reloj, se encontrará en el evento con las coordenadas (1, −.707). Es decir, en el sistema de coordenadas del reloj los marcadores siempre estarán separados por una distancia de .707 minuto-luz. En esta forma hemos derivado la contracción Lorentz-FitzGerald basada en coordenadas.

Los estudiantes de física pudieron haberse dado cuenta de que hemos derivado todos estos resultados sin tener la necesidad de utilizar jamás la transformación de Lorentz, es decir, la transformación coordenada general entre las coordenadas de Lorentz que tienen el mismo origen y la misma calibración. El lector puede encontrar más ejemplos de este tipo de problema en el apéndice.

Antes de seguir con la contracción física Lorentz Fitz-Gerald, debemos examinar un conjunto final de confusiones potenciales. Nos hemos preocupado por utilizar con cuidado las coordenadas en nuestras descripciones: cuando mencionamos el “marco de referencia del reloj”, queremos dar a entender un conjunto de coordenadas de Lorentz en las cuales la velocidad coordenada del reloj es igual a cero. Si hablamos coloquialmente de “cómo parecen ser las cosas” en semejante marco de referencia, queremos dar a entender solamente cómo se les asignan coordenadas a los eventos en ese sistema. Tal cosa no debe confundirse con el sentido muy diferente de “cómo las cosas le parecen a un observador”, es decir, literalmente lo que el observador *vería* si abriera los ojos. Esto se determina por la luz que atraviesa sus pupilas y nada tiene que ver con las coordenadas o los sistemas de referencia.

Por ejemplo, la dilatación temporal basada en coordenadas describe cómo las coordenadas de Lorentz asociadas con un observador en movimiento inercial se conectan con un reloj que se mueve en relación con ese observador. En este sentido, un reloj “en movimiento” siempre le parecerá “lento” al observador. Pero esto no describe en absoluto qué es lo que el observador literalmente *vería* si utilizara un telescopio para observar el reloj. La observación literal en este sentido (en vez de la adscripción de coordenadas) algunas veces encuentra que el reloj se atrasa y que algunas veces se adelanta. Si regresamos al escenario de los gemelos, resulta fácil añadir rayos de luz al diagrama y determinar, sin el uso de *cualquier tipo* de coordenadas, qué es lo que literalmente vería cada uno de los gemelos. Supongamos, para que las cosas sean sencillas, que cada gemelo emitiera un pulso luminoso cuando juzgara, gracias a su reloj, que han transcurrido 10 días. Estos pulsos de luz se muestran en la figura IV.10.

Como se indica en el diagrama, el primer *flash* que emitió el gemelo B llega hasta el gemelo A exactamente en el evento donde da la vuelta para retornar. (El flash se emite desde el evento con las coordenadas (1, 0, 0, 0) y la pendiente de la trayectoria es +1, de manera que el *flash* llega hasta el punto (5, 4, 0, 0).) Puesto que el gemelo A ha envejecido 30 días en el ínterin, si ha estado observando continuamente a su hermano a lo largo del viaje, lo ha visto a él y a sus relojes moverse muy lentamente, a ⅓ de su velocidad normal. Es decir, a lo largo de 30 días (de acuerdo con sus propios relojes) el gemelo A observa que su hermano pasa por 10 días de actividad (de acuerdo con los relojes de su hermano). En el viaje de regreso la situación se revierte. El gemelo A recibe nueve *flashes*, aunque sólo experimenta 30 días de tiempo: su hermano, visto a través del telescopio, parece actuar cómicamente, de manera acelerada. Así pues, en el viaje de vuelta de 30 días, el gemelo A ve al gemelo B envejecer 90 días. Al final, percibirá que su hermano ha envejecido 100 días a lo largo del viaje, inevitablemente.

La experiencia del gemelo B es a la vez similar y diferente. Cuando ve a su hermano alejarse de la Tierra, éste también parece, desde el punto de vista de la inspección visual, haberse vuelto más lento por un factor de tres. El pulso luminoso de 30 días que el gemelo A envía desde el punto en que se da la vuelta para regresar, le llega al gemelo B tras 90 días, según sus cálculos. Así, durante 90 días el gemelo B ve al gemelo A envejecer lentamente y le parece que sólo ha envejecido 30 días cuando inicia el viaje de regreso. En el viaje de vuelta, al gemelo B le parece que el gemelo A envejece tres veces más rápido. Pero el gemelo B sólo experimenta el viaje de vuelta durante 10 de sus días. En ese periodo, el gemelo A parece envejecer 30 días, de manera que habrá envejecido 60 días cuando se vuelven a reunir, inevitablemente.

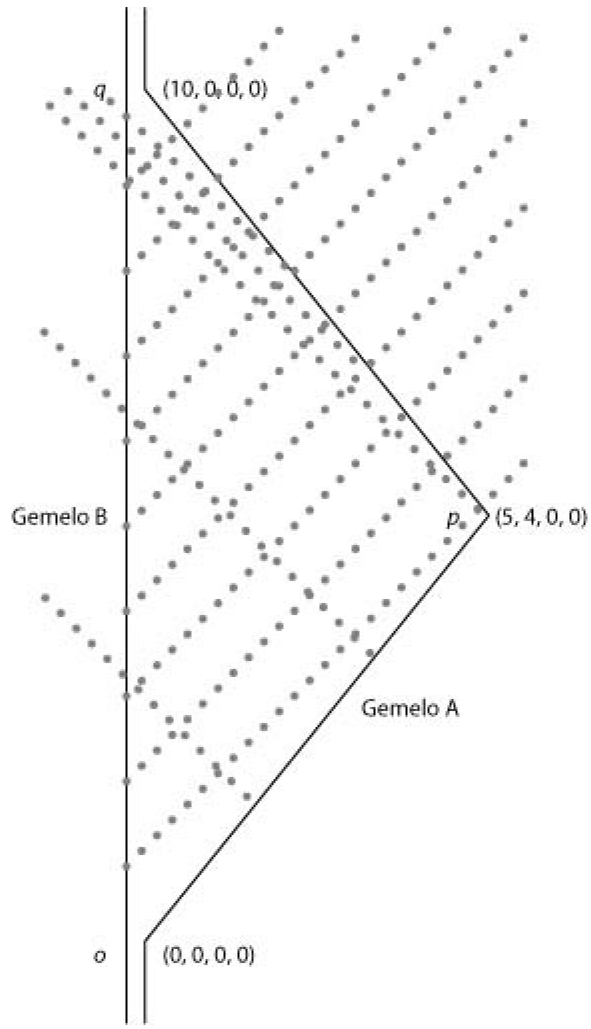

FIGURA IV.10

No hay reglas generales con relación a la aceleración *aparente* de un reloj en movimiento, si por "aparente" entendemos la apariencia visual que literalmente tiene para un observador. Las apariencias se pueden determinar mediante el examen del diagrama de espacio-tiempo, y no es necesario en absoluto utilizar ningún sistema de coordenadas. Es ésta una cuestión física, y la integración de las coordenadas se realiza simplemente porque resulta conveniente en los cálculos.

[1] Aunque éste sea el punto de vista mayoritario respecto a la cuestión, yo mismo lo cuestiono. Véase mi comentario en el capítulo I, nota 1.

[2] Es aquí que nuestra definición del intervalo mediante la raíz cuadrada se justifica a sí misma: utilizando el método convencional más usual, los relojes miden la raíz cuadrada del intervalo.

[3] Por esta razón también la "métrica" de Minkowski no es una métrica.

[4] W. Rindler, *Essential Relativity*, Springer-Verlag, Nueva York, 1977, p. 46.

[5] R. Feynman, R. Leighton y M. Sands, "The Twin Paradox", *The Feynman Lectures on Physics*, vol. 1, Addison-Wesley, Reading (Mass.), 1975, p. 16-3.

[6] W. Rindler, *op. cit.*, p. 41.

V. FÍSICA DE LA MEDICIÓN

LA HIPÓTESIS DEL RELOJ

Para ligar la geometría del espacio-tiempo al comportamiento observable de las cosas perceptibles, necesitamos un principio físico de algún tipo. En el capítulo precedente investigamos algunas de las consecuencias observables de la premisa de que el espacio-tiempo posee una geometría minkowskiana, junto con la ley de la luz, la ley relativista de la inercia y la hipótesis del reloj. También supusimos que podíamos determinar el hecho de que un cuerpo no se encuentre sujeto a fuerzas externas y de que una región en particular sea un vacío. Dando por sentado todo esto, expusimos el fenómeno de los gemelos y mostramos cómo las coordenadas de Lorentz podían construirse física y visiblemente en el espacio-tiempo.

Entre los tres principios físicos que hemos mencionado, la hipótesis del reloj llama la atención especialmente. La ley de la luz y la ley relativista de la inercia describen, utilizando el vocabulario que puede esperarse de una ley de la física, cómo los rayos de luz y los cuerpos masivos libres de fuerzas funcionan en el espacio-tiempo de Minkowski. La hipótesis del reloj es una cuestión totalmente diferente, ya que especifica cómo funcionan los relojes (con relación a la geometría de Minkowski); pero es evidente que la palabra “reloj” no es el tipo de término que debería aparecer en una ley de la física fundamental. Es posible que la naturaleza reconozca la diferencia entre la luz y las partículas masivas, pero la naturaleza no tiene que determinar si a un mecanismo dado se le puede identificar con un “reloj” para establecer cómo debería comportarse. Un término de la especie de “reloj”, a diferencia de “rayo de luz” o “partícula masiva”, no debe aparecer en ningún planteamiento de ninguna ley de la física fundamental.

Esto lo sabía bien Einstein, reconociendo que cualquier discusión física que utilizara los términos de “relojes” o de “barras de medición” sólo se podía aceptar como algo que fuera temporalmente conveniente:

> En primer lugar, una observación crítica a la teoría [es decir, la relatividad especial], tal y como queda caracterizada anteriormente. Llama la atención el que la teoría (fuera del espacio cuadridimensional) introduzca dos tipos de cosas físicas, a saber: *1*) reglas de medir y relojes, *2*) todas las demás cosas, por ejemplo el campo electromagnético, el punto material, etc. Esto es, en cierto sentido, inconsecuente; las reglas de medir y los relojes deberían representarse en realidad como soluciones de las ecuaciones fundamentales (objetos consistentes en configuraciones atómicas móviles), no como entidades en cierta medida autónomas desde el punto de vista teórico. Semejante proceder se justifica, sin embargo, porque desde un principio se vio claro que los postulados de la teoría no son lo bastante fuertes como para que las ecuaciones de los fenómenos físicos deducidas de ellos sean tan completas y libres de arbitrariedad que permitan fundar sobre esa base una teoría de las reglas de medir y de los relojes. De no querer renunciar por entero a una interpretación física de las coordenadas (lo que, de suyo, sería posible), era mejor permitir semejante inconsecuencia —aunque con la obligación de eliminarla en un estudio posterior de la teoría—. Ahora bien, no cabe legitimar el susodicho pecado hasta el punto de imaginar, pongamos por caso, que las distancias sean entes físicos de naturaleza especial, esencialmente distintos de las demás magnitudes físicas.[1]

Nosotros hemos evitado la mención de las “barras de medición” en nuestro texto, refiriéndonos sólo a los relojes y a los rayos de luz, pero sigue habiendo un problema en relación con los relojes, y, como veremos, el asunto de las “barras rígidas” pronto habrá de aparecer. ¿Cuál es, pues, la naturaleza exacta de la hipótesis del reloj y hasta dónde podemos llevar a la práctica la exigencia de Einstein en el sentido de que los relojes no sean más que configuraciones atómicas en movimiento que obedecen las leyes de los campos electromagnéticos, las partículas materiales, etcétera?

Primero hay que darse cuenta de que el término mismo de “hipótesis del reloj” es poco afortunado. Es una frase que se utiliza en la literatura de la física, pero no hay forma alguna de expresar el contenido de semejante hipótesis de manera general. Más bien, con lo que tenemos que empezar no es con una hipótesis, sino con una *definición:*

> Definición de reloj: Un reloj ideal es un cierto tipo de instrumento observable mediante el cual es posible asignar ciertas cifras a los eventos en la línea de mundo del instrumento, en forma tal que las proporciones de las diferencias entre las cifras correspondan a las proporciones entre las longitudes de los intervalos de los segmentos de la línea de mundo cuyos extremos sean esos eventos.

Así, por ejemplo, si un reloj ideal de alguna manera le asigna las cifras 4, 6 y 10 a los eventos *p, q* y *r* en su línea de mundo, entonces la proporción entre las longitudes del segmento $\overline{pq}$ y del segmento $\overline{qr}$ es 1:2, etc. Éste es exactamente el sentido que para nosotros tiene “un reloj ideal” en el contexto de la relatividad.[2]

Dada esta definición de un reloj ideal, podemos entonces proponer la *hipótesis* de que un cierto sistema físico explícito es un reloj ideal, o que se aproxima a un reloj ideal en cierto nivel de precisión en ciertas circunstancias específicas. Semejante hipótesis se justificaría de diferentes formas. Por ejemplo, si proponemos que dos sistemas distintos constituyen relojes ideales, podríamos ponerlos lado a lado con el fin de ver si las “diferencias temporales” que registran son proporcionales entre sí. Seríamos capaces de argüir que la mejor explicación para la existencia de los diferentes tipos de instrumentos que consideramos como relojes, y que concuerdan entre sí de este modo, consiste en que todos (o casi todos) son relojes ideales: su conformidad se podría explicar en el sentido de que todos miden con precisión la misma cantidad física observable. Este tipo de razonamiento justificaría el hecho de que consideráramos los relojes finos suizos como relojes (aproximadamente) ideales, aun cuando no pudiéramos explicar con claridad y con palabras fundamentales cómo funcionan. Pero la justificación de la idea de que un instrumento en particular se aproxime a un reloj ideal, tal como se ha definido, debe basarse necesariamente en demostrar que satisface la definición. Según Einstein, esto es lo que una teoría de la física completa y coherente debe ser capaz de proporcionar en última instancia.

Cumplir con esta obligación teórica en el caso de un reloj suizo mecánico, o de un reloj de cuarzo, o de un reloj atómico, requiere una cantidad enorme de análisis detallado: tenemos que ser capaces de describir totalmente el instrumento con el vocabulario de la física fundamental y de especificar las leyes que gobiernan todas las partes. No estamos en una posición de llevar a cabo semejante análisis aquí. Pero podemos desarrollar el análisis de un tipo muy sencillo de reloj idealizado en función de los principios de la física que tenemos al alcance, y mucho aprenderemos en el proceso. El reloj que tengo en mente es un *reloj de luz*.

Un reloj de luz consiste en un rayo de luz que se refleja entre dos espejos. Supongamos que haya un mecanismo (y también en este caso no nos obligamos a proporcionar un análisis físico completo) que sea capaz de registrar el momento en que el rayo de luz llega hasta uno de los espejos y de hacer "tic-tac" cuando eso sucede. También supongamos que algún tipo de mecanismo cuenta los tic-tacs. En esta forma, el instrumento asignaría números enteros a eventos específicos que se encuentran a lo largo de su línea de mundo.

Einstein menciona este tipo de reloj en su exposición:

> La presuposición de la existencia (en principio) de reglas de medida (ideales o perfectas) no es independiente de la existencia de relojes (también ideales), porque una señal luminosa que es reflejada una y otra vez entre los extremos de una regla rígida representa un reloj ideal, siempre y cuando el postulado de la constancia de la velocidad de la luz en el vacío no conduzca a contradicciones.[3]

Einstein especifica que los espejos deben fijarse a los extremos de una "barra rígida", pero puesto que éste es un concepto que todavía no hemos introducido, veremos hasta dónde podemos llegar sin utilizarlo.

Pensemos en un sistema que se constituye por dos espejos en movimiento inercial que tienen sendas trayectorias paralelas en el espacio-tiempo de Minkowski, por un rayo de luz que se refleja sucesivamente entre ambos espejos y por un contador que registra el número de veces que el rayo de luz llega hasta uno de los espejos. *En la medida en que los espejos estén libres de fuerzas impresas*, este sistema se comportará como un reloj ideal: los "tic-tacs" sucesivos del rayo de luz que llega al espejo delimitan segmentos congruentes en la línea de mundo del espejo. En esencia, podemos sustituir con espejos el reloj regidor y el reloj-objetivo 1 que aparecen en la figura IV.6, y entonces considerar que la llegada del rayo de luz a uno de los espejos constituye un reloj. Aquí no nos interesa explicar *cómo podríamos saber* que las trayectorias de los relojes son paralelas entre sí: sencillamente consideramos que tal cosa es una parte de la especificación física del sistema en cuestión.

Tal como se ha definido, este instrumento está sujeto a una restricción severa: únicamente podemos probar que funciona como un reloj ideal si ambos espejos se encuentran en movimiento inercial. Para que sea un reloj ideal, un sistema tiene que poder medir continuamente el intervalo, sin que importe la trayectoria que siga. Pero empecemos con una pregunta más sencilla: ¿cómo funcionaría este sistema si quisiéramos que pasara de un movimiento inercial a otro movimiento inercial acelerándolo repentinamente en cierta dirección? En suma, ¿cómo funcionaría nuestro sistema si lo *impulsáramos*, haciendo que pasara de su movimiento inercial actual a otro movimiento inercial?

Supongamos que decidimos realizar esta tarea dándole a cada reloj un impulso idéntico. Hasta ahora, nuestra descripción de semejante intervención no ha sido suficientemente precisa: también tenemos que especificar los eventos exactos en que estos impulsos se llevan a cabo. En la figura V.1 mostramos un par de impulsos que se aplican *a los eventos con el mismo valor* t *en un marco de Lorentz donde los espejos se encuentran inicialmente en reposo*. El contador está conectado al espejo del lado derecho y los números producidos por el dispositivo se indican en el diagrama.

Tanto antes como después del impulso, el sistema que funciona como un reloj ideal debe hacerlo: la trayectoria entre los eventos 1 y 2 es congruente con la trayectoria entre los eventos 2 y 3, y la trayectoria entre los eventos 3 y 4 es congruente con la trayectoria entre los eventos 4 y 5. ¿Pero cuál es el caso respecto a la relación entre los segmentos antes y después del impulso? ¿Es congruente el segmento entre 2 y 3 con el segmento entre 3 y 4, por ejemplo?

Si especificamos el impulso con mayor exactitud, podemos calcular la respuesta. Adoptemos un sistema de coordenadas de Lorentz en que el evento 3 sea el origen y las coordenadas del evento 2 sean (−1, 0). En estas coordenadas, los intervalos entre 1 y 2 y entre 2 y 3 tienen ambos la medida 1. Supongamos que luego del impulso ambos espejos se mueven a $\frac{4}{5}$ de la velocidad de la luz en estas coordenadas. Entonces la ecuación para el espejo del lado derecho tras el impulso que se le da en estas coordenadas es $x = \frac{4}{5}t$ y la ecuación para el espejo del lado izquierdo es $x = \frac{4}{5}t - \frac{1}{2}$. (Los espejos deben estar separados por ½ de una "unidad", puesto que a la luz le toma una "unidad" realizar el viaje redondo entre ambos.) La ecuación para el rayo de luz emitido en el evento 3 es $x = -t$, de manera que las coordenadas del evento *p* son $(\frac{5}{18}, \frac{-5}{18})$. (Solucione $-t = \frac{4}{5}t - \frac{1}{2}$ para encontrar este hecho.) La ecuación para el rayo de luz desde *p* es $x = t - \frac{5}{9}$, de manera que las coordenadas del evento 4 fácilmente se calculan como $(\frac{25}{9}, \frac{20}{9})$. El intervalo entre el evento 3 y el evento 4 es

$$\sqrt{(\tfrac{25}{9})^2 - (\tfrac{20}{9})^2} = \tfrac{5}{3}.$$

En resumen, nuestro reloj de luz "hace tic-tac más lentamente" luego del impulso de antes: los tic-tacs sucesivos posteriores se corresponden con un intervalo mayor que los tic-tacs sucesivos anteriores. Nuestro sistema no es un reloj ideal.

Con el fin de continuar midiendo el intervalo con la misma rapidez, los espejos deberán de algún modo *aproximarse* más después del impulso de lo que lo hacen en la figura V.1. Más exactamente, en el marco original lorentziano deberán mostrar una separación de sólo las $\frac{3}{5}$ partes de la que mostraban originalmente. Así que además de la fuerza que coloca a los espejos en sus nuevas trayectorias inerciales, tiene que haber algún tipo de mecanismo mediante el cual, a juzgarse desde este marco de Lorentz, los espejos se acerquen. Semejante mecanismo habría de producir una contracción *física* Lorentz-FitzGerald. No resulta una coincidencia que el grado de esta contracción física sea exactamente el grado de la contracción Lorentz-FitzGerald basada en coordenadas. Sin embargo, todavía no tenemos una explicación del *mecanismo físico mediante el cual se produce ese tipo de contracción*.

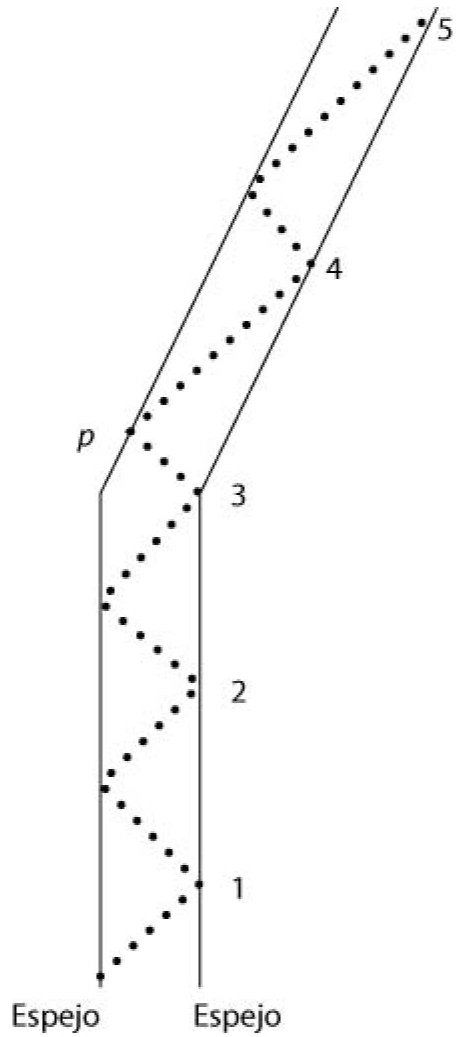

FIGURA V.1

Es de notar que las interacciones físicas que seguidamente examinaremos sólo se puedan describir de manera adecuada como una *contracción* de la distancia entre los espejos desde el punto de vista de algunos marcos lorentzianos, particularmente desde el punto de vista del marco inicial de reposo de los espejos. En otros marcos lorentzianos, el mismo cambio se describe como una *expansión* de la distancia entre los espejos. Sin embargo, tiene que haber una explicación física concreta de cómo se comportan los espejos: aquí no sólo analizamos los sistemas de coordenadas, sino que aplicamos la física con el propósito de explicar el fenómeno.

Vimos arriba que Rindler caracteriza la contracción Lorentz-FitzGerald meramente como un asunto de proyección geométrica más que como una consecuencia de las fuerzas físicas. Este punto de vista es ampliamente aceptado por los físicos, como fue demostrado por John Stewart Bell. Éste, que trabajaba en el acelerador de partículas del CERN [Consejo Europeo para la Investigación Nuclear], en una ocasión describió el resultado de un experimento informal que él había realizado, presentándoles a sus colegas una situación muy parecida a la de nuestros dos espejos: dos cohetes de idéntica construcción al comienzo en movimiento inercial y mutuamente en reposo. Se envía una señal a los motores de reacción con el fin de que se enciendan “en el mismo momento”, a juzgar por su marco de reposo lorentziano. Puesto que su construcción es idéntica, las trayectorias de los cohetes pueden representarse en un diagrama de espacio-tiempo con figuras paralelas congruentes, de manera que, *a juzgar por el marco de Lorentz inicial*, los cohetes mantendrán una separación constante entre sí, precisamente de la misma manera que nuestros espejos lo hacen en la figura V.1. Bell luego añadió un pequeño detalle: entre los cohetes se ha tendido un hilo tenso. Y Bell les hizo una pregunta a sus colegas, algunos de los cuales se encontraban entre los más respetados físicos teóricos y experimentales del mundo, una simple pregunta de orden físico: ¿se romperá el hilo?

Bell dice;

> Este viejo problema en una ocasión se discutió en el comedor del CERN. Un distinguido físico experimental rechazaba la posibilidad de que el hilo se rompiera y pensaba que mi afirmación de que efectivamente lo haría mostraba mi incomprensión de la relatividad especial. Decidimos apelar al arbitraje de la División Teórica del CERN e hicimos una pequeña encuesta (no muy sistemática) al respecto. ¡El consenso unánime fue que el hilo *no* se rompería![4]

Pero en efecto el hilo se rompe, como se muestra en la figura V.2.

Los físicos del CERN probablemente veían el problema desde la misma perspectiva que Rindler: si la contracción Lorentz-FitzGerald fuera simplemente una cuestión de *observar los mismos eventos desde un ángulo diferente*, o simplemente una cuestión de *describir los mismos eventos en un sistema de coordenadas diferente*, entonces sin lugar a dudas no podría ser la causa del rompimiento del hilo. Si en realidad no tuviera nada que ver con las fuerzas eléctricas o interatómicas, entonces no podría tener cualquier tipo de efecto físico observable. Y de hecho, eso que hemos llamado la *contracción Lorentz-FitzGerald basada en coordenadas* es sencillamente una forma de observar las relaciones entre los diferentes sistemas de coordenadas de Lorentz. Sin embargo, como insiste Bell correctamente, también hay una contracción Lorentz-FitzGerald *física* que *efectivamente* depende de las fuerzas interatómicas y que puede tener efectos físicos. Tenemos que entender este efecto físico con el fin de ver cómo podríamos arreglárnoslas para construir un reloj ideal.

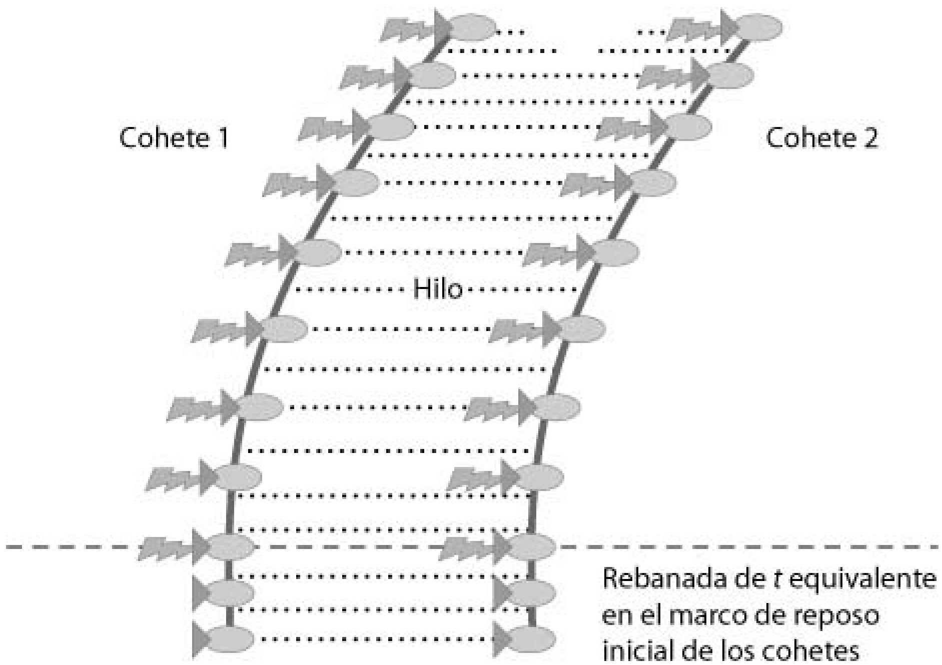

FIGURA V.2

EMPUJONES ABSTRACTOS Y EMPUJONES FÍSICOS

La contracción Lorentz-FitzGerald basada en coordenadas puede derivarse, como nosotros lo hemos hecho, sin discutir jamás el caso de un objeto sujeto al impulso de una fuerza y que por eso se acelera. Incluso cuando hubimos de construir físicamente las coordenadas de Lorentz, jamás tuvimos que lidiar con un cuerpo en aceleración: todos los relojes siempre tuvieron trayectorias inerciales. De manera que la contracción basada en coordenadas, por sí misma, no puede tener implicaciones sobre las consecuencias físicas de acelerar un sistema, tal como las consecuencias de acelerar los espejos en la figura V.1 o los cohetes en la figura V.2.

En la literatura de la física, a veces se denomina *empujón de Lorentz* al simple hecho de transformar las coordenadas de un marco lorentziano en un segundo marco en movimiento relativo respecto al otro. Es lícito preguntarnos en qué forma se construyen las coordenadas del mismo grupo de eventos en ambos sistemas. Obviamente, semejante "empujón" constituye un cambio abstracto descriptivo, no un cambio físico. A veces se le denomina *empujón pasivo*. Por otra parte, podemos quedarnos con un solo sistema de coordenadas de Lorentz (o incluso no utilizar en absoluto un sistema de coordenadas) y preguntarnos qué le sucedería a un sistema físico si se le sujetara a una fuerza y por lo tanto se le hiciera acelerar. La aceleración podría ser constante, como en el caso de los cohetes de Bell, o podría ser el resultado de un impulso relativamente fuerte, como en el caso de nuestros espejos. En el caso de los espejos, al mismo sistema lo llevamos desde una trayectoria inercial, antes del impulso, hasta otra trayectoria inercial, después del impulso. A este tipo de cambio *físico* de un sistema se le puede llamar un *empujón activo* o un *empujón físico*.

Recordemos el barco de Galileo en el capítulo III. Galileo sometió el barco a un conjunto de experimentos mientras estaba inercialmente "en reposo" en el puerto. Luego describió cómo al *mismo barco exactamente, con el mismo equipo*, se le dejaba "avanzar a cualquier velocidad que os agrade, siempre que el movimiento fuera uniforme y no fluctuara de un lado a otro". Es decir, una vez que hubiera alcanzado velocidad, el barco estaría de nuevo en movimiento inercial. Así, el barco de Galileo se corresponde con nuestros espejos en la figura V.1: se encontraba en una trayectoria inercial antes y después de una aceleración limitada. Y la afirmación de la física a la que se le ha dado el nombre de "relatividad galileana" dice que *los resultados observables de los experimentos serán los mismos antes y después de la aceleración*. Galileo no se pronuncia sobre lo que sucedería *durante* la aceleración: al avanzar el barco gracias al impulso del viento, las mariposas efectivamente se congregarían hacia la popa del barco, el agua no caería directamente, etc. De manera que la propuesta de Galileo consiste en que un sistema en movimiento inercial se comportará de la misma manera después y antes de un empujón físico.

Sin embargo, nuestro intento de construir un reloj de luz parece refutar lo anterior: el reloj en la figura V.1 hace tic-tac a un ritmo diferente (en función de la medición del intervalo) antes y después de que los impulsos se impriman. Entonces, ¿refuta la relatividad especial a la relatividad galileana, o refuta la equivalencia *física de todos los marcos inerciales?* ¡Estas consecuencias serían extremadamente extrañas, puesto que la relatividad galileana con frecuencia se presenta como un postulado fundamental de la relatividad especial!

Si regresamos a la figura V.1, debería ser obvio que el ritmo al que el reloj hace tic-tac después de que se le hayan aplicado los impulsos depende crucialmente de *dónde y cuándo* se administraron los impulsos. Por ejemplo, si el impulso en el lado derecho se hubiera aplicado un poco después, entonces los espejos hubieran terminado más cerca el uno del otro y el reloj hubiera hecho tic-tac a ritmo más rápido (en relación con el intervalo). Nosotros decidimos aplicar los impulsos con el mismo valor t en el marco de reposo inicial de los espejos, pero ésa fue una decisión libremente elegida. Acaso "debimos" aplicarlos en un momento distinto.

En el caso de un reloj físico real esta cuestión jamás se plantea, puesto que un reloj físico real no puede contener dos partes completamente inconexas. De hecho, una manera sencilla de transformar nuestro par de espejos en un reloj ideal adecuado es, como dice Einstein, *conectándolos ambos a una barra rígida*. Una vez que hayamos hecho esto, no tenemos que preocuparnos de cuándo aplicaremos el par de impulsos: simplemente podemos aplicarlos en *uno* de los espejos, dejando que la barra se encargue de acarrear el otro espejo. Durante y poco después del impulso, cuando las partes del sistema se estén acelerando, su funcionamiento a veces se desestabilizará. Pero a la larga, después de que dejemos de impulsarlo, el sistema volverá a equilibrarse en una nueva trayectoria inercial. Y si hacemos esto con nuestros espejos de la figura V.1, veremos que una vez que todo vuelva a estabilizarse, *los espejos se encontrarán más cerca de lo que se muestra en la figura V.1 y exactamente a la distancia adecuada para que el reloj de luz haga tic-tac al mismo ritmo antes y después del impulso*. La conexión de los relojes con una barra de acero produce una mayor aproximación a un reloj ideal.

Si aceptamos este resultado por el momento, podemos ver que la presencia de la barra tiene un efecto físico, real, en las trayectorias de los espejos: la barra lleva un poco hacia atrás el espejo del lado derecho y un poco hacia adelante el espejo del lado izquierdo (tal como se describe en el marco de reposo inicial). La barra genera estas fuerzas en los espejos mediante una especie de contracción física real, la cual es causada a su vez por las fuerzas interatómicas que forman la estructura rígida de la barra. En su texto "How to Teach Special Relativity" [Cómo enseñar la relatividad especial],[5] en el cual aparece la anécdota que citamos anteriormente, Bell produce un análisis físico detallado de esta contracción en el caso de un sistema estructurado por fuerzas electromagnéticas y que se ha acelerado muy suavemente. (Tiene que ser muy suave esta aceleración, pues de lo contrario el sistema sencillamente se quebrará en vez de contraerse.) El hilo en el ejemplo de Bell se rompe porque aunque estas fuerzas interatómicas tratan de contraerlo los cohetes impiden esta contracción, acrecentando la tensión hasta que ésta supera la resistencia a la tensión del hilo. Si reemplazáramos el hilo que menciona Bell con un cable fuerte, la tensión del cable causaría que los cohetes se aproximaran entre sí, produciendo la contracción física.

Pero si Bell se ve obligado a realizar un análisis detallado de las fuerzas electromagnéticas en el átomo para obtener este resultado, ¿cómo podemos estar seguros de que ciertas otras fuerzas en otros sistemas habrían de producir el mismo aparentemente fortuito resultado, reposicionando los espejos con gran exactitud con el fin de que se mantuviera el funcionamiento preciso del reloj?

La clave de un análisis general yace en la noción de un cuerpo *rígido*. Einstein habla de "barras de medición", y la característica esencial de una barra de medición es que tiene que ser, de alguna manera, rígida. Si intentáramos aplastarla, una barra rígida opondría resistencia a la compresión, y si intentáramos estirarla, también se resistiría. Es decir, un cuerpo rígido posee un *estado de equilibrio* que tiende a mantener ante las fuerzas externas (suficientemente pequeñas), y siempre retorna a ese estado de equilibrio cuando las fuerzas externas se retiran. Nuestro sencillo sistema de dos espejos no constituye un sistema rígido sin la barra conectora porque los espejos no tienden a resistir las fuerzas externas y a mantener un estado de equilibrio fijo.

La comprensión física total de un estado de equilibrio requeriría una explicación total de la estructura interna del sistema rígido, tanto de su composición como de las fuerzas entre sus partes. Pero incluso sin tener una explicación detallada de este tipo, podemos hacer algunas aseveraciones respecto a los cuerpos rígidos en cualquier teoría compatible con la relatividad especial. El requisito fundamental de una teoría relativista consiste en que las leyes físicas puedan especificarse utilizando únicamente la geometría del espacio-tiempo relativista. En el caso de la relatividad especial, esto significa específicamente el espacio-tiempo de Minkowski. La simetría del espacio-tiempo minkowskiano es lo que nos permite demostrar nuestro resultado general.

Supongamos que existe un sistema con un estado de equilibrio que tienda a mantenerse estable cuando se encuentre libre de fuerzas externas (y por lo tanto en movimiento inercial). La especificación de este estado típicamente requeriría la descripción precisa de la estructura interna y de las fuerzas del sistema. Llamemos a este estado de equilibrio S_{EQ}. En un sistema de coordenadas de Lorentz dado, tal como el marco de reposo del sistema, S_{EQ} tendrá una descripción específica que depende de las coordenadas: por ejemplo, la posición relativa de las partículas en el sistema puede darse en función de sus coordenadas en este marco lorentziano. Puesto que S_{EQ} es un estado de equilibrio, el sistema tenderá naturalmente a retornar a S_{EQ} si se da el caso de que esté en un estado cercano a S_{EQ} y se encuentre libre de fuerzas externas. Si el sistema se somete a una pequeña deformación que lo separe ligeramente del S_{EQ}, habrá de retornar espontáneamente a S_{EQ} como consecuencia de sus fuerzas estructurales intrínsecas.

Ahora consideremos el caso de un *diferente* estado físico, S', que se relacione con S_{EQ} de la siguiente manera: S tiene la misma descripción basada en coordenadas con relación a un sistema de coordenadas de Lorentz *diferente*, que S_{EQ} en relación con su marco de reposo. (Lorentz denominaba *estados correspondientes* a los pares de este tipo.) Por lo tanto, para *cualquier* tipo de leyes de fuerza relativistas, S' también sería un estado de equilibrio, y un sistema cerca del estado S' y libre de fuerzas externas tendería a entrar en el estado S'. Es que la geometría minkowskiana asume exactamente la misma forma descrita en cualquiera de los sistemas de coordenadas lorentzianas (por causa de la simetría del espacio-tiempo minkowskiano), y las leyes de la física asumen exactamente la misma forma basada en coordenadas cuando se formulan en un lenguaje basado en coordenadas en cualquier sistema de coordenadas de Lorentz (porque las leyes sólo pueden referirse a la geometría de Minkowski, y ésta tiene la misma descripción basada en coordenadas). Así, el comportamiento de S' descrito en términos de las nuevas coordenadas de Lorentz sería idéntico al comportamiento de S_{EQ} descrito en términos de las anteriores coordenadas. La tendencia que mostraría el sistema original para asumir el estado S_{EQ} si se le dejara solo implicaría la tendencia del nuevo sistema para asumir el estado S'.

Si S' se relaciona con las nuevas coordenadas lorentzianas de la misma forma exactamente en que S_{EQ} se relaciona con las coordenadas lorentzianas precedentes, es fácil entender cómo S' y S_{EQ} se relacionan entre sí, incluso si se encuentran en estados muy complicados. La figura V.3 muestra la situación de modo esquemático: en el diagrama únicamente necesitamos "angular" S_{EQ} de manera adecuada para obtener S'.

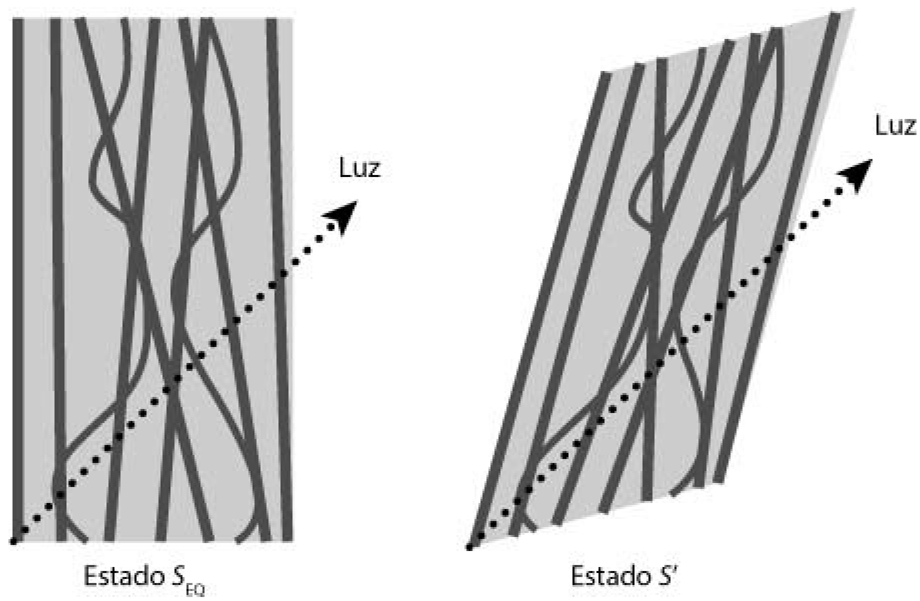

FIGURA V.3

A pesar de las apariencias, S_{EQ} y S' son geométricamente congruentes en el espacio-tiempo de Minkowski. Genéricamente, son el mismo estado físico. Por lo tanto, si al comienzo el sistema tiene la disposición de regresar a S_{EQ} tras el empujón físico apropiado, también se dispondrá a regresar a S'. Una vez más, la transición de S_{EQ} a S' debe ser lo suficientemente suave: una barra de hierro que se someta a presión o a estiramiento tiende a retornar a su longitud inicial a menos que se la presione o estire con demasiada fuerza, deformándola.

El acercamiento de Bell al empujón físico, el cual involucra la aplicación de la dinámica adecuada para el sistema acelerado, es mucho más detallada que nuestra presentación genérica. Si fuera posible desarrollar el programa de Bell, podríamos calcular exactamente cómo se comporta el sistema en todo momento y con todo detalle. Por ejemplo, si empujáramos un lado de nuestro reloj de luz rígido, éste se comprimiría momentáneamente cuando la onda de fuerza lo atravesara. Durante la aceleración, no se comportaría exactamente de la misma manera en que debiera hacerlo un reloj ideal, y las ecuaciones precisas de Bell podrían determinar el grado de esas divergencias respecto a la idealidad. Nuestra exposición genérica no puede explayarse en torno a esta cuestión: sólo señala que una vez que dejáramos de acelerar el reloj, permitiéndole encontrar su nuevo estado de equilibrio, nuevamente se convertiría en un reloj ideal que marchara al ritmo *original* en relación con el intervalo. Además, ni siquiera necesitaríamos conocer los detalles de las fuerzas que estructuran el reloj, sólo que las leyes dinámicas podrían especificarse en términos de la geometría de Minkowski. Si estas condiciones se mantuvieran y nosotros aceleráramos nuestro reloj rígido con suficiente suavidad, al final del experimento se habría contraído a la manera de la contracción Lorentz-FitzGerald hasta el grado correcto que le permitiera retornar al estado ideal.[6]

Se dijo arriba que el comportamiento físico que se describe como la *contracción* del hilo en un marco, en otro marco se describe como la *expansión* del hilo. Es una consecuencia de la simetría de la contracción de Lorentz-Fitz Gerald. Consideremos, por ejemplo, un marco de Lorentz para la figura V.2. Al encenderse los cohetes, el hilo se vuelve *más lento* en vez de *más rápido*. En este marco, sería de esperar que la estructura interatómica del hilo hiciera que éste se *expandiera* en vez de *contraerse*. Entonces, ¿por qué se rompe el hilo?

En la figura V.2, el eje *t* de este nuevo marco de referencia será paralelo a la trayectoria final de los cohetes. Es decir, el nuevo eje *t* habrá girado en dirección de las manecillas del reloj con relación al anterior. Y como ya hemos visto, cuando el eje *t* gira en dirección de las manecillas del reloj, el eje *x* correspondiente gira en dirección contraria a las manecillas del reloj, de manera que los rayos de luz siguen dividiendo la diferencia entre los ejes. Ahora la clave consiste en darse cuenta de que en este nuevo marco de referencia los cohetes no empiezan a acelerar "en el mismo momento" (es decir, en el mismo valor *t*). En este nuevo marco, el cohete del lado derecho empieza a acelerar hacia la derecha *antes* de que lo haga el cohete del lado izquierdo. No debe sorprender a nadie que el hilo se rompa: al irse separando en su movimiento los cohetes (en este marco), el hilo tendría que estirarse para soportar la distancia creciente entre las naves. El hilo no es capaz de estirarse tanto a pesar de que (en este marco) el efecto de las fuerzas interatómicas causa que el hilo se relaje, más que tensarse. En otros marcos, el cohete de la izquierda empieza a acelerar antes que el cohete de la derecha, de manera que se juzga que las naves se aproximan entre sí, pero el estrechamiento de los lazos interatómicos es tan grande que el hilo no puede abarcar incluso esta distancia reducida, y por lo tanto se rompe.

La contradicción aparente entre estas tres explicaciones de la razón por la que el hilo se rompe es una ilustración de que pueden ser engañosas las disquisiciones de los eventos en la relatividad que dependen de los marcos. Tan sólo existe un conjunto de eventos que se gobierna por leyes para las cuales son indiferentes los tipos de sistema de coordenadas que puedan utilizarse para describir una situación. En cada una de las explicaciones que dependen de marcos, las fuerzas interatómicas en el hilo juegan un rol en la exacta determinación del momento en que se romperá el hilo. Pero la forma en que se describa ese rol en un marco de referencia específico depende crucialmente de qué marco se elija.

LA "CONSTANCIA DE LA VELOCIDAD DE LA LUZ"

Como señalamos al comienzo de nuestra discusión de la relatividad, para que cualquier ente pueda tener una velocidad objetiva, independiente de los marcos de referencia, tiene que haber un hecho objetivo respecto a cuánto tiempo transcurre entre dos eventos y respecto a la distancia en el espacio entre ambos eventos. Una vez que abandonamos el tiempo absoluto newtoniano y la persistencia de los puntos en el espacio absoluto newtoniano, no existen las velocidades absolutas, ya sea de la luz o de cualquier otra cosa.

Dado un sistema de coordenadas, podemos definir la velocidad coordenada de cualquier objeto mediante el método de considerar que una coordenada juega el rol del tiempo absoluto y las demás coordenadas el rol del espacio absoluto. Por ejemplo, dadas las coordenadas lorentzianas (t, x, y, z), podemos definir el tiempo coordenado que transcurre entre los eventos p y q como $t_p - t_q$, y la distancia espacial coordenada entre los eventos como $\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}$. Dadas estas definiciones, la velocidad coordenada de cualquier rayo de luz en un vacío es 1, puesto que el intervalo para este tipo de rayo es siempre 0. Pero éste es un comentario sobre la definición de las coordenadas de Lorentz más que sobre el comportamiento de la luz.

Al explicar cómo se construyen los marcos lorentzianos mediante procedimientos experimentales, la "constancia de la velocidad de la luz" se vuelve empíricamente importante. Pero ya que la constancia de la velocidad coordenada se integra analíticamente en la definición de un marco lorentziano, la verdadera cuestión empírica consiste en comprobar si el procedimiento constructivo funcionará o no. Además, *nadie le asigna realmente coordenadas a los eventos en un laboratorio mediante el procedimiento arriba expuesto*. El problema práctico es obvio: el tiempo necesario para un viaje de ida y vuelta de la luz dentro de un laboratorio es demasiado breve como para que se pueda registrar con la debida precisión por un reloj común y corriente. Por lo tanto, la explicación de la importancia *empírica* de la aseveración de que la velocidad de la luz es constante mediante la referencia a la construcción de las coordenadas de Lorentz, poco tiene que ver con cualquier tipo de operación real en un laboratorio que jamás se haya realizado.

No obstante, sí existen resultados de operaciones reales en laboratorios reales que se podrían describir como indicaciones de la constancia de la velocidad de la luz. En estas operaciones *no se utilizan ni relojes, ni barras de medición, ni coordenadas*. Ahora nos encontramos en una posición para entender por qué este tipo de experimentos debe tener un cierto resultado si la relatividad especial es correcta. La contracción física Lorentz-FitzGerald juega un rol crítico en estas elucidaciones.

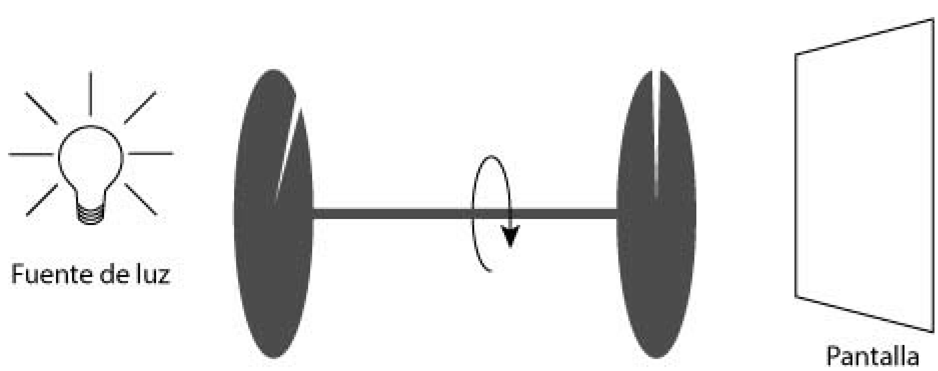

FIGURA V.4

Examinemos el siguiente aparato, que tiene cierta semejanza con uno utilizado por Hippolyte Fizeau para determinar la velocidad de la luz. Dos discos están conectados a una barra que gira muy velozmente. Cada uno de los discos tiene un corte angular radial. La luz tiene que pasar por ambos cortes para atravesar todo el aparato. Puesto que los discos giran, los cortes tienen que emparejarse para que la luz pueda pasar a través de los dos, y esto depende de la velocidad con que giran, de la distancia entre ambos y del tiempo que le tome a la luz llegar de un disco al otro (figura V.4).

Podríamos intentar usar el aparato que aparece en la figura V.4 para determinar la velocidad de la luz mediante la medición de la distancia entre los discos, la velocidad de su rotación y el ángulo entre los cortes, pero esto involucraría el complejo asunto de medir la distancia y el tiempo que arriba mencionamos. De forma alternativa, simplemente podríamos *afinar* el aparato, ajustando la velocidad de la rotación o el ángulo entre los cortes o la distancia entre los discos, hasta que la luz pudiera atravesar ambos cortes y llegar a la pantalla. Una vez afinado el aparato de esta forma, serían varios los experimentos que se podrían realizar.

Podríamos, en primer lugar, variar la fuente de luz, incluyendo, por ejemplo, la utilización de fuentes de luz en movimiento relativo hacia adelante o hacia atrás del aparato. Si la ley de la luz es correcta, la luz habría de atravesarlo sin que importara el estado de la fuente.

Podríamos colocar el aparato en una base giratoria y hacerlo girar para que apuntara en diferentes direcciones. Esto es similar a lo que Albert Michelson y Edward Morley hicieron en su famoso experimento, donde el aparato estaba colocado en un bloque de mármol que flotaba en un recipiente lleno de mercurio. Michelson y Morley utilizaron la interferometría en vez de nuestra sencilla técnica mecánica, pero la idea era la misma: el instrumento no era capaz de determinar el valor de la velocidad de la luz (relativa al aparato), pero resultaba extremadamente sensible a los cambios de la velocidad (relativa al aparato). Pero sin que tuviera importancia la orientación del instrumento o la época del año, el resultado jamás variaba. La época del año hubiera sido significativa si la órbita de la Tierra cambiara su velocidad en relación con el supuesto éter luminóforo.

Finalmente, podríamos poner el aparato en el barco de Galileo y probar si la luz producida en el barco y si la luz producida en tierra firme podrían atravesarlo. Este tipo de experimentos naturalmente se podrían describir (recurrien do a la intuición newtoniana) como "comprobaciones de si la velocidad de la luz es constante o no", aunque en ellos nunca se pretendiera determinar realmente cuál sería la velocidad de la luz.

A estas alturas sabemos que si *1*) la relatividad especial explica correctamente la geometría del espacio-tiempo, y que si *2*) en las diversas configuraciones el aparato se encuentra en movimiento inercial,[7] y que si *3*) el aparato es un cuerpo rígido, entonces el resultado jamás podría cambiar: una vez debidamente afinado, la luz habría de atravesar el aparato en todas las configuraciones. Todas las condiciones listadas deben existir para que el resultado sea válido. Si pusiéramos el aparato en la cubeta de Newton y lo hiciéramos girar, por ejemplo, las apuestas se retirarían. Y al cambiar la colocación de una en que el aparato se oriente del este al oeste a una en que se oriente del norte al sur, o pasar el aparato de una colocación en que esté en reposo en el laboratorio a otra en que esté en reposo en el barco (en movimiento inercial), se tendría que acelerar el aparato para luego permitir que retornara a su estado de equilibrio. Hemos demostrado que las simetrías en el espacio-tiempo de Minkowski implican que estos estados de equilibrio se relacionan entre sí mediante las transformaciones lorentzianas. Y por lo tanto *cualquier rayo de luz que atraviese el aparato tendrá las mismas relaciones geométricas con las partes del instrumento en cualquiera de las configuraciones experimentales*. Si volvemos a la figura V.3, veremos que los rayos de luz que aparecen en los diagramas tienen las mismas relaciones geométricas con S_{EQ} que las que tienen con S', porque el cono de luz mismo es intrínseco en la geometría del espacio-tiempo. Por lo tanto, si un rayo de luz logra atravesar un aparato en S_{EQ}, necesariamente lo podrá hacer también a través de un aparato en S'. Tal cosa le da una obvia importancia empírica a la aseveración de que "la velocidad de la luz es constante".

No obstante, siempre debemos recordar las condiciones en que estas predicciones empíricas pueden realizarse. El aparato únicamente debe responder a la *trayectoria* de un rayo de luz y no, digamos, a su *color*. Y el aparato debe encontrarse en movimiento inercial cuando se utilice. Y debe estar construido con materiales rígidos y no debe acelerarse demasiado cuando se le cambie de una trayectoria inercial a otra. Si se encuentra sujeto a estas condiciones, la relatividad especial predice que los resultados de estos experimentos serán siempre los mismos, sin que tenga importancia el tipo de construcción del aparato. En este sentido experimental, la relatividad especial predice, y explica, la constancia de la velocidad de la luz.

Explicaciones más profundas de los principios físicos

Para extraer predicciones experimentales de una explicación de la geometría del espacio-tiempo se requieren ciertos principios que conecten esa geometría con el comportamiento observable de los entes físicos. Hemos utilizado tres principios de este tipo en nuestra discusión: la ley de la luz, la ley relativista de la inercia y la hipótesis del reloj. Los primeros dos, a diferencia del último, tienen la forma correcta como para ser leyes fundamentales de la naturaleza. Pero es posible que también sean consecuencias de leyes más profundas formuladas en términos de conceptos diferentes.

Por ejemplo, la electrodinámica de Maxwell se podría formular en términos del espacio-tiempo de Minkowski: es por esta razón que las consideraciones sobre los fenómenos electromagnéticos desembocaron para empezar en la relatividad. Así que podríamos postular la electrodinámica max-weliana como una ley física fundamentalmente relativista y luego identificar los rayos de luz con las ondas electromagnéticas. Podríamos entonces *derivar*, en vez de postular, a partir de tal situación que la trayectoria de un rayo de luz en un vacío estará en el cono de luz.

También es posible aplicar consideraciones similares a la ley relativista de la inercia. Ésta se basa en la primera ley del movimiento de Newton y tiene la forma apropiada como para ser una ley natural fundamental. Pero podría no serlo. Podríamos imaginar una explicación de las trayectorias de los cuerpos masivos cuyas consecuencias determinaran que tales cuerpos siempre, acaso sólo de manera aproximada, siguen líneas rectas a través del espacio-tiempo cuando se encuentran libres de fuerzas externas. Si estas leyes más fundamentales se formularan en términos de la geometría de Minkowski, también se vindicaría la relatividad especial. La cuestión básica consiste sencillamente en saber si la estructura del espacio-tiempo de Minkowski constituye toda la estructura del espacio-tiempo que se requiere en la física, y si las leyes físicas articuladas en términos de esa estructura pueden hacer predicciones precisas. La teoría de Maxwell demostró que tal cosa se pudo hacer respecto al electromagnetismo, y las investigaciones posteriores integraron las fuerzas nucleares débiles y fuertes en la cuestión. Pero el intento de producir una física compatible con la relatividad especial se tropezó con la gravedad. La resolución de esta dificultad produjo la teoría general de la relatividad.

[1] A. Einstein, *op. cit.*, pp. 59-61. [Versión en español: *Notas autobiográficas*, pp. 61-62.]

[2] C. Misner, K. Thorne y J. Wheeler insisten en este punto en su clásico *Gravitation*,. W. H. Freeman, San Francisco, 1983, p. 393.

[3] A. Einstein, *op. cit.*, p. 55. [Versión en español: *Notas autobiográficas*, pp. 58-59.]

[4] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2ª ed., Cambridge University Press, Cambridge, 2008, p. 68.

[5] *Ibid.*, cap. 9, "How to Teach Special Relativity".

[6] Nuestra presentación de la contracción física de Lorentz-FitzGerald, la cual explica el rompimiento del hilo en el ejemplo de Bell, fundamenta el efecto en tres circunstancias: las simetrías geométricas del espacio-tiempo de Minkowski, la necesidad de especificar las leyes dinámicas en función de la estructura minkowskiana y la constitución física que hace del hilo un cuerpo rígido (en el sentido apropiado). Es similar la explicación de por qué un reloj de luz con una barra rígida retorna a su ritmo de funcionamiento correcto después de un impulso físico. Harvey Brown cuestiona si los datos respecto a la geometría del espacio-tiempo pueden jugar semejante rol explicativo, y además rechaza la idea de apelar a la geometría *per se* del espacio-tiempo en las explicaciones físicas. Sin embargo, en el punto crítico de su libro, donde Brown intenta poner en duda la corrección de recurrir a la geometría del espacio-tiempo (y al espacio-tiempo mismo) en cuanto base física de una explicación, su impugnación se reduce a esta figura retórica: "¿Pero no es esta manera de razonar circular?" (H. Brown, *Physical Relativity*, Oxford University Press, Oxford, 2006, p. 139). Brown no especifica exactamente en qué consiste dicha circularidad y por lo tanto es difícil responder a su pregunta. Pienso que mi explicación muestra que no hay ninguna circularidad involucrada.

[7] Es más exacto decir que si el aparato se afina cuando está sujeto a cierta aceleración constante (incluyendo cero), dará el mismo resultado cuando esté sujeto a la misma aceleración.

VI. LA RELATIVIDAD GENERAL

EL ESPACIO CURVO Y EL ESPACIO-TIEMPO CURVO

La relatividad general se desarrolló como una teoría de la gravedad que incorpora la explicación cualitativa de la estructura del espacio-tiempo que se encuentra en la relatividad especial. Es necesario disipar unos cuantos mitos comunes en torno a la relatividad general antes de exponer la teoría. Con frecuencia se ha dicho que la relatividad general es una extensión de la relatividad especial, con las siguientes características: en la relatividad especial todos los marcos de referencia inerciales (es decir, todos los marcos de Lorentz) son equivalentes, y en la relatividad general todos los marcos de referencia son equivalentes; o en la relatividad especial hay una distinción física entre el movimiento acelerado y el movimiento no acelerado, pero en la relatividad general no la hay en absoluto; o en la relatividad especial el espacio-tiempo posee una estructura inercial que no es una función de la distribución de la masa, pero en la relatividad general sí es una función de la distribución de la masa (lo cual es una vindicación de Mach). Todas estas aseveraciones son falsas. La relatividad general le atribuye una estructura geométrica objetiva, intrínseca, al espacio-tiempo, exactamente de la misma manera que lo hace la relatividad especial, y esa estructura geométrica incluye una estructura afín que distingue entre el movimiento acelerado y el movimiento no acelerado. La explicación en relatividad general del fenómeno de las esferas de Newton tiene exactamente la misma forma general que la explicación proveniente de la relatividad especial o, de hecho, de la explicación que se da en el marco del espacio-tiempo galileano: existe una tensión en la cuerda que conecta los globos justo cuando los globos aceleran (es decir, giran). La aceleración se define con relación a la estructura intrínseca del espacio-tiempo: los globos podrían ser los únicos objetos materiales en existencia. Es cierto que en la relatividad general la distribución de la materia *influye* en la geometría del espacio-tiempo, pero la distribución de la materia no *determina* la geometría del espacio-tiempo. Por ejemplo, en la relatividad general existen muchas diferentes soluciones en vacío, en cada una de las cuales el espacio-tiempo se encuentra vacío de todo tipo de materia y de energía. Una de estas soluciones en vacío es el espacio-tiempo de Minkowski.

También se dice a veces que la búsqueda de una teoría relativista de la gravedad representaba un reto peculiar porque la fuerza gravitacional newtoniana es instantánea, mientras que en la relatividad ya no existe una noción de la simultaneidad mediante la cual se pueda definir una acción de tipo instantáneo. Todo esto es erróneo por diversos motivos. Primero, no es muy probable que Newton haya pensado realmente que la fuerza gravitacional sea instantánea: Newton pensaba que en la fuerza tenía que interceder un cierto tipo de partícula a la cual le hubiera tomado cierto tiempo llegar, por ejemplo, del Sol a la Tierra. Claro, fue exactamente respecto a este punto que Newton declaró: *Hypotheses non fingo*. Pero es más crucial aún que no haya nada en la forma general de la ley gravitacional de Newton que implique dificultades para una versión relativista: la ley de la electroestática de Coulomb es una ley cuadrática inversa de la fuerza, precisamente como la ley de la gravedad newtoniana, y la electrodinámica de Maxwell se reduce a la ley coulombiana en el límite apropiado. Pero la electrodinámica maxweliana es cabalmente relativista. De hecho, poco después de que Newton hubiera propuesto la relatividad especial en 1905, muchas y muy diferentes teorías de la gravedad compatibles con la relatividad especial fueron desarrolladas. El mismo Einstein investigó varias, aunque rechazándolas porque no lograban satisfacer ni el principio de equivalencia débil ni el principio de equivalencia fuerte, como pronto veremos.[1] Únicamente con la relatividad general en 1915 se satisficieron los criterios de Einstein, pero el punto crucial no fue la acción instantánea.

Hay otra aseveración común respecto a la relatividad general que no es totalmente correcta, aunque se acerque a la verdad. Ésta es la afirmación de que en la relatividad general se explican los efectos gravitacionales mediante la sustitución del espacio euclidiano de la física newtoniana con el espacio "curvo" no euclidiano. La afirmación certera no consistiría en que el espacio plano euclidiano se sustituya con el espacio curvo, sino que el espacio-tiempo plano minkowskiano se sustituya con el espacio-tiempo curvo no minkowskiano. De hecho, no queda claro, aun en la relatividad especial, qué es lo que se quiere dar a entender con la geometría *espacial* del mundo: de la misma manera que no hay noción objetiva o única o privilegiada de la simultaneidad de los eventos en la relatividad especial, tampoco existe una noción objetiva o única o privilegiada de la "geometría del espacio". Un sencillo ejemplo puede demostrar esto.

Supongamos que el espacio-tiempo tenga una estructura minkowskiana. ¿Qué podríamos decir sobre "la estructura geométrica del espacio"? Puesto que el espacio es tridimensional, la pregunta sólo se puede hacer una vez que hayamos seleccionado algún tipo de subespacio tridimensional en el espacio-tiempo minkowskiano. Semejante subespacio no podría contener eventos relacionados con el tiempo o con la luz. Sin embargo, hay muchos subespacios relacionados con el espacio que es posible extraer del espacio-tiempo de Minkowski, y todos muestran estructuras geométricas distintas.

El tipo más obvio de subespacio topológico "espacial" en el espacio-tiempo de Minkowski se relaciona con las coordenadas de Lorentz. Dado cualquier sistema de coordenadas de Lorentz (t, x, y, z), examinemos todos los eventos que tienen el mismo valor t; es decir, examinemos una "rebanada de simultaneidad" de ese sistema coordenado. El intervalo entre los eventos p y q es proporcional a

$$\sqrt{(t_p - t_q)^2 - (x_p - x_q)^2 - (y_p - y_q)^2 - (z_p - z_q)^2}.$$

Por lo tanto, si todos los valores t de los eventos en el subespacio son iguales, el intervalo entre esos eventos será proporcional a

$$\sqrt{-(x_p - x_q)^2 - (y_p - y_q)^2 - (z_p - z_q)^2} =$$
$$i\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}.$$

Con excepción del factor de i, esto tiene exactamente la misma forma que la función de la distancia en el espacio euclidiano, tal como se expresa en términos de las coordenadas cartesianas, y el factor de i se diluye porque solamente nos interesan las proporciones entre esos números. De modo y manera que en sentido riguroso la geometría intrínseca de una rebanada de t equivalente en un sistema de coordenadas de Lorentz es tridimensional y euclidiana.

Pero el espacio-tiempo de Minkowski contiene muchas otras superficies de tipo espacio. Por ejemplo, el hiperboloide de la figura IV.5 es otro conjunto de eventos en el espacio-tiempo de Minkowski al cual se le podría denominar "espacio", pero su geometría intrínseca no es euclidiana. Si se quisiera identificar este conjunto de eventos con el "espacio", entonces se tendría que decir que en la relatividad especial el "espacio" es hiperbólico, más que euclidiano.

La investigación de la geometría espacial no euclidiana constituye una buena preparación para la comprensión cabal de la relatividad general, y la geometría no euclidiana se capta más fácilmente con relación a los espacios de dos dimensiones. Nosotros sabemos que únicamente el espacio euclidiano admite las coordenadas cartesianas, de manera que en la investigación del espacio no euclidiano se requiere que dejemos de lado la comodidad de las coordenadas sencillas. El mejor método desecha las coordenadas totalmente, y es así como hemos de proceder.

El espacio euclidiano tienen muchas simetrías: es homogéneo, isotrópico e invariante en escala. Los espacios no euclidianos más sencillos son homogéneos e isotrópicos y es posible describirlos mediante sencillos axiomas. El más conocido es el espacio de curvatura positiva constante: es decir, la superficie de una esfera. En la superficie de una esfera, la línea más recta entre dos puntos es siempre un arco de un gran círculo, y estas líneas también satisfacen la noción intuitiva de la rectitud: si se nos diera la orden de caminar hacia adelante en línea recta en la superficie de la Tierra, naturalmente trazaríamos un gran círculo. Si identificamos los grandes círculos con las líneas rectas en la esfera y medimos las distancias de la manera usual, le adscribimos una geometría intrínseca no euclidiana a la superficie.

Existen formas matemáticas precisas para caracterizar esta geometría, pero para nuestros propósitos es suficiente enfocar algunas de las propiedades genéricas. Califiquemos de "localmente paralelas" un par de líneas rectas en un espacio de dos dimensiones, si es que ambas intersecan en ángulo recto una cierta línea recta. En el espacio euclidiano, las líneas paralelas locales son en realidad paralelas: jamás se intersecan. Pero en la superficie de una esfera, las líneas localmente paralelas se acercan la una a la otra y finalmente se intersecan, lo cual se muestra por las líneas de las longitudes en un globo terráqueo. Todas éstas intersecan el ecuador en ángulo recto, y por lo tanto son localmente paralelas, pero todas se reúnen en los polos norte y sur (figura VI.1).

Gracias a la figura VI.1 es evidente que la geometría intrínseca de la superficie de una esfera no es euclidiana: la suma de los ángulos internos del triángulo formado por dos líneas de longitud y el segmento del ecuador debe ser mayor a dos ángulos rectos, puesto que los ángulos en el ecuador ya son dos ángulos rectos. La suma de los ángulos interiores de un triángulo en este espacio es siempre mayor que la de dos ángulos rectos, siendo mayor el exceso en proporción al área del triángulo.

En el espacio hiperbólico, es decir, un espacio de curvatura negativa constante, las líneas localmente paralelas se comportan de manera opuesta: son divergentes en vez de convergentes. Una buena aproximación a la figura de este espacio se da por una superficie de dos dimensiones con la forma de una silla de montar en el espacio tridimensional euclidiano (figura VI.2). En este caso, la suma de los ángulos internos de cualquier triángulo es menor que la suma de dos ángulos rectos.

Además de estos espacios especialmente sencillos de curvatura constante, existen los espacios de curvatura variable. Algunas de las partes de estos espacios variables pueden ser planas, otras positivamente curvas, y aun otras más negativamente curvas. También puede ocurrir que el espacio siempre tenga el mismo tipo de curvatura, pero de diferente dimensión en diferentes lugares. Riemann desarrolló los métodos para describir esta especie de espacios utilizando funciones métricas generales que pueden variar de lugar en lugar. El examen de estos diversos espacios riemannianos nos puede proporcionar una analogía que puede ser útil en la comprensión de la relatividad general. La analogía es básicamente ésta: los espacios de curvatura variable riemannianos se relacionan con el espacio plano euclidiano del mismo modo que los espacio-tiempos de la relatividad general se relacionan con el espacio-tiempo plano de Minkowski. Pero para entender cómo podemos utilizar esta analogía, tenemos que examinar primero algunos de los aspectos más notables de la gravedad.

ELIMINACIÓN DE LA GRAVEDAD MEDIANTE LA GEOMETRÍA

Si el experimento más importante en la historia de la física del espacio-tiempo fue el experimento de la cubeta newtoniana, el que ocupa el segundo lugar en importancia fue realizado por Galileo. En *Diálogos acerca de dos nuevas ciencias*, Galileo se propone refutar la propuesta aristotélica de que la velocidad con que caen los cuerpos es proporcional a su peso. Puesto que los cuerpos más pesados experimentan más la fuerza de la gravedad que los más ligeros, naturalmente se puede pensar que los más pesados caen con mayor velocidad, pero Galileo comprobó que la diferencia en la velocidad del descenso es insignificante. Si Aristóteles hubiera tenido la razón al respecto, un cuerpo 10 veces más pesado caería 10 veces más rápido, pero en el diálogo Sagredo dice:

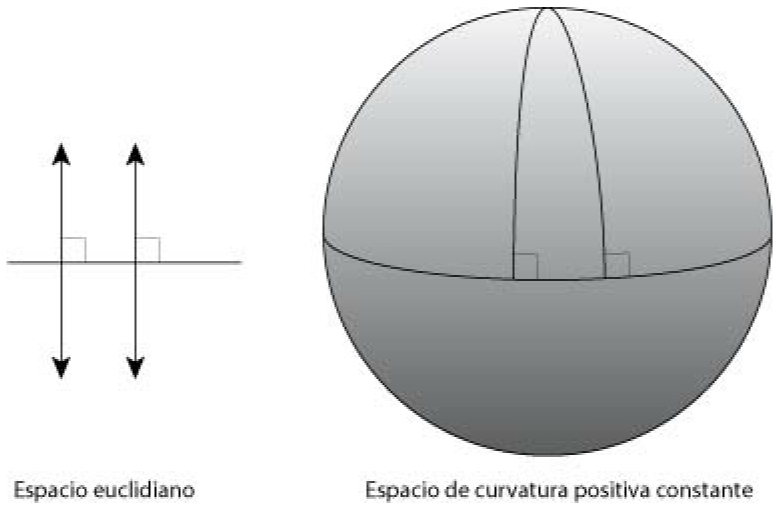

FIGURA VI.1

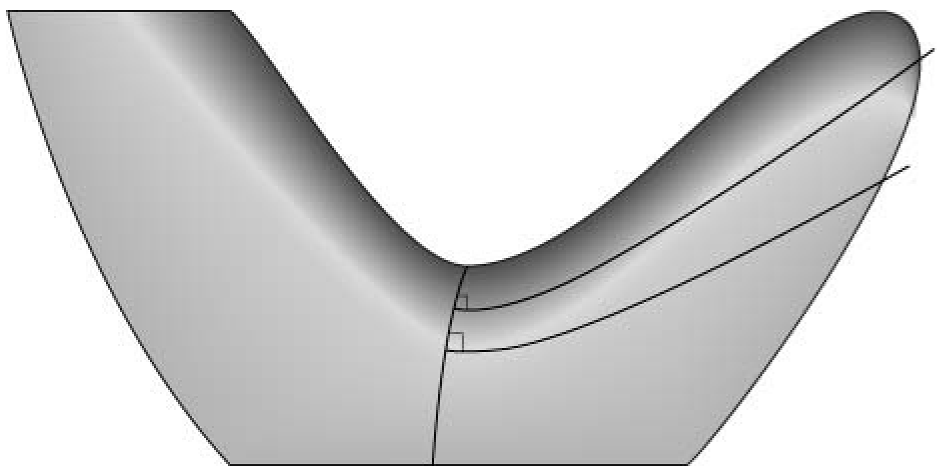

FIGURA VI.2

> Pero, Simplicio, yo que he llevado a cabo la prueba, puedo asegurarte que una bala de cañón que pesa cuarenta y cinco kilogramos (o noventa, o incluso más kilogramos) no se anticipa ni siquiera por un palmo a la llegada al suelo de una bala de mosquetón de no más de diez gramos, cuando ambos caen desde una altura de doscientos brazos [aproximadamente 90 metros].[2]

Galileo supuso correctamente que la pequeña diferencia en la caída se debía a la resistencia del aire y que los cuerpos con diferentes masas que se sometieran a prueba caerían exactamente con la misma rapidez si sólo estuvieran sujetos a la influencia de la fuerza de la gravedad. Este principio con el tiempo se denominó *principio de equivalencia débil*.

La teoría gravitacional de Newton supone el principio de la equivalencia débil, puesto que plantea que la fuerza de gravedad que un cuerpo experimenta es proporcional a su masa. La masa, en la explicación newtoniana, juega tres roles físicos distintos: es una medida de la resistencia natural de un cuerpo a experimentar la aceleración producida por una fuerza (la masa inercial), una medida de cuánta fuerza gravitacional un cuerpo genera en otros cuerpos (la masa gravitacional activa) y una medida de cuánto le afecta a un cuerpo la gravedad de otros cuerpos (la masa gravitacional pasiva). El principio de equivalencia débil, en este escenario, afirma que la masa inercial de un cuerpo es proporcional a su masa gravitacional pasiva. Si esto es válido, entonces las fuerzas gravitacionales en dos cuerpos experimentales serán proporcionales a sus masas inerciales, de modo que habrán de experimentar la misma aceleración (si las únicas fuerzas son las gravitacionales). Por lo tanto, la bala de cañón y la bala de mosquetón caerán al mismo tiempo.

Newton puso a prueba el principio de equivalencia débil mediante la utilización de péndulos hechos de diferentes materiales y encontró que era válido en la proporción de una parte a varios miles de partes. Los experimentos modernos lo han verificado en la proporción de una parte a 10^{12} partes. Desde una perspectiva newtoniana, la explicación obvia del principio de equivalencia débil es que la masa inercial *únicamente es* la masa gravitacional pasiva, de modo que la proporcionalidad tiene que ser exacta.

Ningún principio que se parezca incluso ligeramente al principio de equivalencia débil es aplicable a cualquier tipo de fuerzas, salvo la fuerza de la gravedad. Por ejemplo, dos partículas de prueba que se sujetan sólo a las fuerzas electromagnéticas aceleran de manera completamente distinta si tienen cargas opuestas, o si una tiene carga y la otra es neutral. En este sentido, el principio de equivalencia débil establece que todos los cuerpos masivos reaccionan ante la gravedad exactamente del mismo modo, a diferencia de otras fuerzas, cuyos efectos en un cuerpo dependen de su carga.

La figura VI.3 es un diagrama de espacio-tiempo que muestra el experimento de Galileo tal como Newton lo entendió. Si se descarta el movimiento de la Tierra, Galileo, en la cima de la torre, se mueve inercialmente. Siente una fuerza gravitacional descendiente que se compensa exactamente por la fuerza opuesta y equivalente de la torre bajo sus pies. Desde el momento en que suelta las bolas, éstas tan sólo experimentan la fuerza gravitacional. La bola más pesada acelera con más dificultad, debido a su masa inercial mayor, pero la fuerza gravitacional que siente es correspondientemente mayor, de manera que cae junto con el objeto más ligero. Ambas caen lado a lado. La figura VI.3 también indica la línea de mundo de la parte superior de la cabeza de Galileo. Puesto que la fuerza neta en su cabeza equivale a cero, se trata de una trayectoria inercial recta, según Newton.

No hay un problema de tipo empírico en la explicación newtoniana del principio de equivalencia débil. Sin embargo, Einstein no dejó de detectar un punto conceptual endeble en la explicación de Newton. En cierto sentido, podemos hacer remontar la insatisfacción de Einstein a las observaciones del propio Newton. El experimento de la cubeta y el ejemplo de las esferas relacionan la fuerza neta que experimentó un objeto con la línea de mundo del objeto. Y las fuerzas que experimenta un objeto pueden tener fuentes observables, tales como la tensión en la cuerda que une los globos. De hecho, una forma de construir un "acelerómetro" es haciendo que la fuente de una fuerza sea visible. Por ejemplo, consideremos el caso de un peso sencillo que pueda moverse libremente dentro de un tubo y que esté conectado mediante resortes a los extremos del tubo.[3] Si ponemos este aparato en un automóvil con el tubo orientado hacia adelante y pisamos el acelerador, el peso “se moverá hacia atrás” en el tubo. El resorte de enfrente se estirará y el resorte de atrás se comprimirá hasta que se produzca una fuerza en el peso que sea suficiente para acelerarlo junto con el auto. *Si suponemos que únicamente los resortes producen la fuerza dirigida hacia adelante que experimenta el peso*, la comprensión y la tensión de los resortes medirán la aceleración del vehículo. Es éste el tipo de fenómeno en que Clarke pensaba al mencionar el *shock* que un cuerpo experimenta cuando se lleva desde un estado de movimiento uniforme absoluto hasta un estado de reposo absoluto, es decir, cuando se acelera.

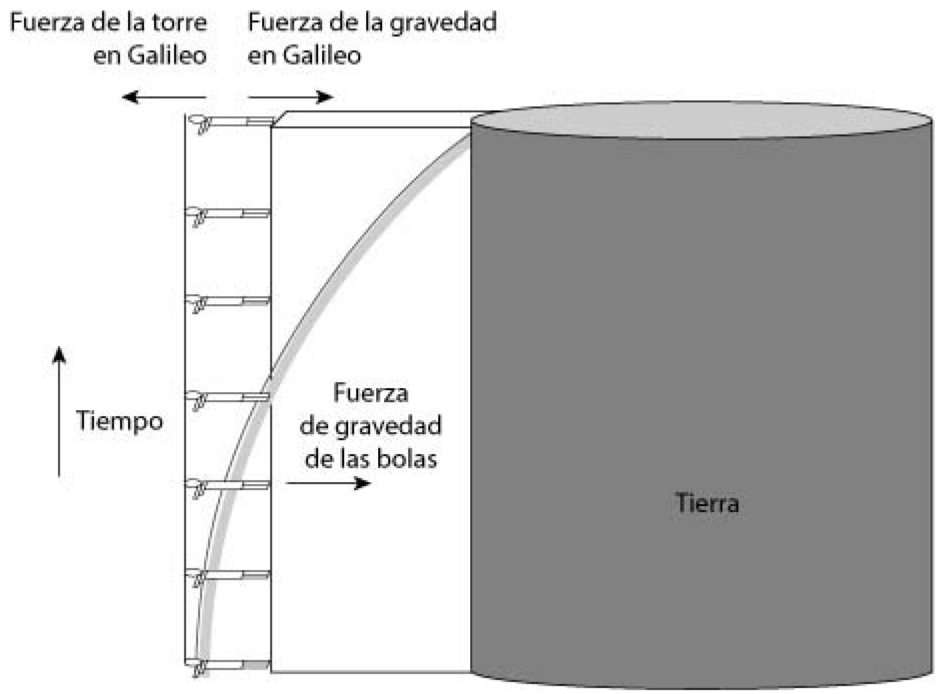

FIGURA VI.3

Supongamos que Galileo se encuentra en la cima de la torre con un acelerómetro de este tipo que apunta hacia abajo. El peso caerá dentro del tubo, estirando el resorte superior y comprimiendo el inferior. Según Newton, tal cosa no indica ningún tipo de aceleración del aparato: más bien, los resortes proporcionan una fuerza que es exactamente suficiente para contrarrestar la gravedad y de este modo mantener el peso en movimiento inercial. Puesto que el funcionamiento correcto del acelerómetro exige que sólo los resortes ejerzan fuerza en la dirección paralela al tubo, Galileo estaría utilizando el aparato de manera incorrecta. Y si Galileo lo soltara y lo dejara caer, los resortes harían que el peso regresara a la posición neutral. Puesto que el revestimiento del acelerómetro caería con la misma velocidad que el peso que se encuentra dentro, no sería necesaria una fuerza adicional proveniente de los resortes para mantenerlos juntos a los dos. El acelerómetro en su caída no mostraría ningún efecto visible de la aceleración: se comportaría como si estuviera moviéndose inercialmente en un espacio vacío donde no hubiera gravedad. Según Newton, esto representa nuevamente un mal uso del aparato.

Para llegar hasta los conceptos centrales de la relatividad general, todo lo que tenemos que hacer es considerar este aparente fenómeno —la incapacidad del acelerómetro de registrar al caer alguna aceleración— tal como se muestra. Regresemos a la figura VI.3 y supongamos que simplemente eliminamos la "fuerza de la gravedad" en el diagrama. Entonces no habrá ningún tipo de fuerzas en las bolas que "caen", pero habrá una *fuerza desequilibrada* sobre Galileo en dirección opuesta al suelo al pie de la torre. Puesto que la versión del espacio-tiempo de la ley de la inercia dice que los objetos que no experimentan fuerzas tienen trayectorias rectas a través del espacio-tiempo, entonces las trayectorias de las bolas serían rectas y la trayectoria de la cabeza de Galileo sería curva. El fenómeno que interpretamos diciendo que mostraba la "igualdad de la masa gravitacional pasiva y de la masa gravitacional inercial" ahora tiene un sentido completamente diferente. Las dos bolas caen lado a lado porque siguen esencialmente la misma trayectoria recta a través del espacio-tiempo. *En la relatividad general no existe una "fuerza de la gravedad":* adscribimos erróneamente una fuerza de este tipo a los objetos porque nosotros, al igual que Galileo, nos encontramos en trayectorias aceleradas cuando nos encontramos "en reposo" en la Tierra. Un marco de referencia conectado a la superficie de la Tierra no es un marco inercial, y por lo tanto la "fuerza de gravedad" que se utiliza en un marco de este tipo es ficticia.

Si aceptamos este escenario, entonces obtenemos gratis el principio de equivalencia débil: los cuerpos que "solamente" están sujetos a la gravedad en realidad no están sujetos a ninguna fuerza y por lo tanto sus trayectorias serán iguales. Pero además del principio de equivalencia débil, también obtenemos un resultado más fuerte: el principio de equivalencia fuerte. Éste afirma que el resultado de cualquier experimento que se realice "en caída libre" en un campo gravitacional uniforme tendrá el mismo resultado que uno que tenga lugar en un laboratorio inercial en el espacio vacío, y que cualquier experimento que se haga "en reposo" en un campo gravitacional uniforme tendrá el mismo resultado que uno que se realice en un laboratorio uniformemente acelerado en el espacio vacío. Pues una vez que eliminamos la fuerza de la gravedad en la figura VI.3, vemos que Galileo, en la cima de la torre, *está* en una condición de aceleración constante, de manera que cualquier experimento que haga tiene que reflejar este hecho.

El principio de equivalencia fuerte es algo difícil de manejar. En la Tierra, el campo gravitacional (tal como Newton lo entiende) no es perfectamente uniforme: los cuerpos no tienden a caer en líneas paralelas, sino que lo hacen en una ligera pendiente hacia el centro de la Tierra. No caen con una velocidad constante, sino que lo hacen más rápidamente mientras más cerca se encuentren de la superficie de la Tierra. Por lo tanto, el principio de equivalencia fuerte no implica simplemente que sea posible distinguir entre un laboratorio "en reposo" en la Tierra, y un laboratorio uniformemente acelerado en el espacio vacío. En el segundo caso, los cuerpos "caerían" (con relación a un marco de referencia conectado al laboratorio) a una velocidad constante en líneas perfectamente paralelas. Pero en la medida que sea posible descartar estas ligeras desigualdades, los resultados de todos los experimentos deberán ser los mismos porque la situación física en los dos casos es la misma: ambos laboratorios se encuentran objetivamente acelerados. Para dar otro ejemplo, las trayectorias de las dos bolas en la figura VI.3 no son perfectamente paralelas: al final las bolas chocarán entre sí al aproximarse ambas al centro de la Tierra. Estos "efectos de marea" pueden ser importantes, pero por el momento no los tomaremos en cuenta.

Una razón por la cual el principio de equivalencia fuerte es más fuerte que el principio de equivalencia débil consiste en que influye en *todo* tipo de experimentos. Por ejemplo: se envía un rayo de luz en una trayectoria inicialmente horizontal de un lado a otro de un laboratorio en la Tierra. ¿Se "doblará" el rayo de luz, es decir, terminará más cerca del suelo cuando llegue a la pared lejana que cuando salió? Con base en el principio de equivalencia débil, no hay manera de saberlo: el resultado dependería de *lo que la luz es*. Si la luz es una especie de partícula con una masa inercial (y por lo tanto gravitacional), entonces será atraída hacia abajo. Si es una especie de fenómeno de ondas, o una partícula sin masa, entonces la cuestión está abierta a la especulación.

Pero si el principio de equivalencia fuerte es válido, el rayo de luz tiene que "doblarse hacia abajo". La razón es que en un laboratorio de aceleración en el espacio vacío el piso del laboratorio "se acelerará hacia arriba" mientras la luz se encuentra en su trayectoria, de manera que ésta terminará más cerca del suelo. En resumen, el principio de equivalencia fuerte predice el "doblamiento de la luz por un campo gravitacional" simplemente porque la luz se encuentra en una trayectoria de la geometría intrínseca del espacio-tiempo. La luz no "se dobla" en absoluto: su trayectoria es recta. Más bien, son las trayectorias del equipo del laboratorio que se encuentran "en reposo" en la Tierra las que están dobladas en virtud de las fuerzas tangibles que el piso del laboratorio, las mesas, etc., imprimen en las trayectorias.

Una de las predicciones capitales de la relatividad general —una predicción que la diferencia de las diversas teorías de relatividad especial en torno a la gravedad que estuvieron vigentes hasta 1915— es el "doblamiento" de la luz que pasa cerca del Sol. La investigación del eclipse realizada por Arthur Eddington en 1919 confirmó la existencia de ese efecto; a partir de entonces la relatividad general ha sido la explicación preponderante de la gravedad.

¿Cómo explica la relatividad general los fenómenos gravitacionales en función de la geometría del espacio-tiempo? Una vez que aceptamos que las trayectorias de las partículas en caída libre son líneas rectas, es evidente que el espacio-tiempo no puede ser minkowskiano. Por ejemplo, consideremos los dos satélites en órbitas opuestas que aparecen en la figura VI.4. De acuerdo con la relatividad general, las trayectorias de los satélites son rectas, pero también se encuentran una y otra vez.

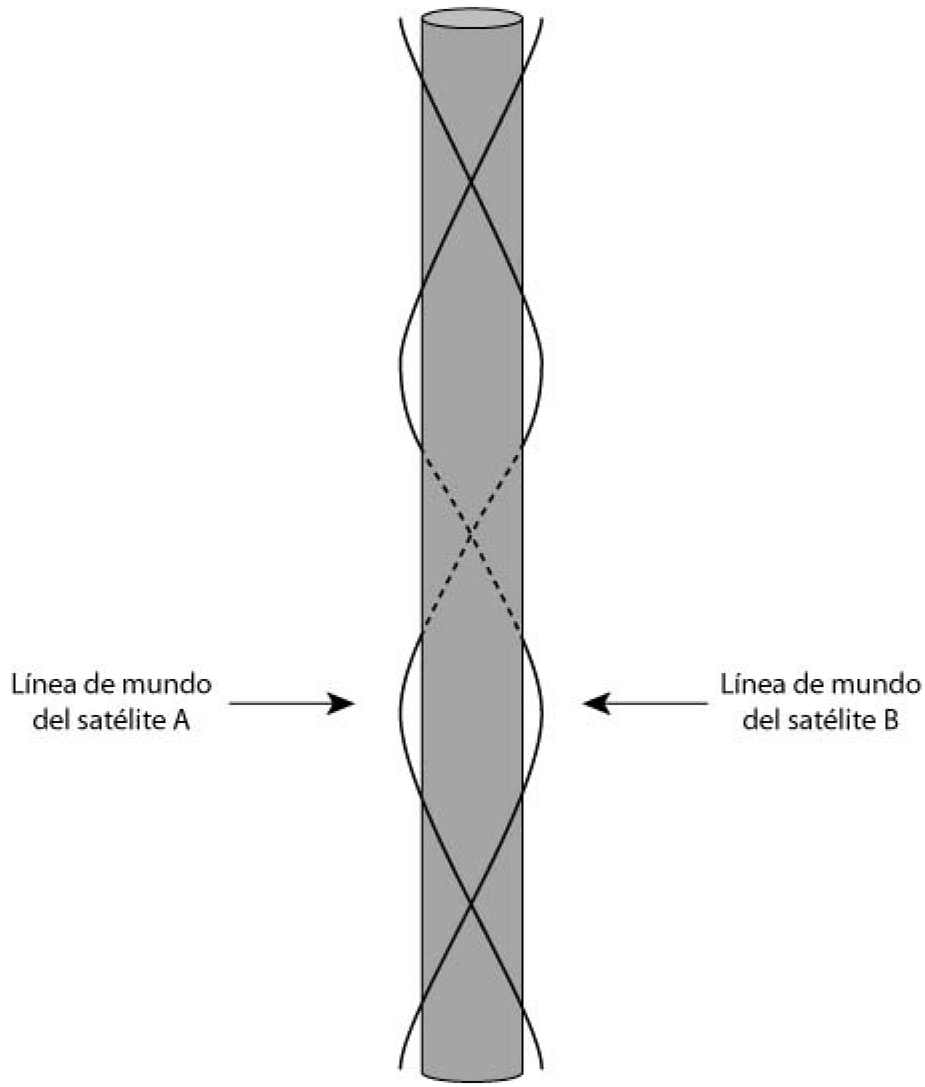

FIGURA VI.4

Puesto que ningún par de líneas rectas en el espacio-tiempo minkowskiano jamás se encuentra más de una vez, la relatividad general tiene que sustituir la geometría plana de Minkowski con una alternativa curva. Y exactamente por la misma razón nuestros diagramas de espacio-tiempo en la relatividad general tienen que leerse con muchísimo más cuidado: ya no podemos utilizar líneas rectas en el diagrama únicamente para representar trayectorias rectas en el espacio-tiempo. Ni la estructura afín ni la estructura métrica del espacio-tiempo curvo representado pueden leerse con facilidad.

El carácter recto de las líneas de mundo satelitales es una consecuencia de la decisión de eliminar cualquier fuerza de gravedad, pero también se corresponde perfectamente con las propiedades de las trayectorias rectas de tipo tiempo en la relatividad especial. En el espacio-tiempo de Minkowski, las curvas rectas de tipo tiempo tienen una longitud máxima: un reloj en una trayectoria de ese tipo habrá de mostrar más tiempo transcurrido entre un par de eventos que cualquier otro reloj. La paradoja de los gemelos nos da un ejemplo del caso. En la relatividad general, las trayectorias en caída libre son localmente máximas: cualquier desviación, por pequeña que sea, habrá de acortar su longitud. Una trayectoria recta no necesariamente tiene que ser de una longitud global máxima, pero un reloj en una trayectoria de ese tipo registrará más tiempo que los relojes acelerados cercanos a él.

Richard Feynman nos proporciona un ejemplo maravilloso de las implicaciones físicas de todos estos principios.[4] Supongamos que hay un reloj "en reposo" en el suelo al pie de la torre de Galileo en la figura VI.3. Puesto que el reloj se encuentra en una trayectoria curva en el espacio-tiempo, no deberá estar registrando el máximo tiempo posible entre los eventos en su línea de mundo. Veamos, por ejemplo, el evento en que el reloj registra las 12:00 y el evento en que registra 12:01. Tendría que haber una manera de que hubiera un reloj que empezara en el primer evento registrando 12:00 también y que terminara en el segundo evento habiendo registrado un tiempo transcurrido mayor. ¿Cómo podemos maximizar el tiempo que transcurre en semejante reloj?

Si los relojes en trayectorias rectas registran (localmente) tiempos transcurridos máximos, y si los objetos en caída libre caen en trayectorias rectas, entonces la respuesta es fácil: a las 12:00 el reloj deberá ser lanzado hacia arriba en línea recta con la cantidad exacta de fuerza que sea necesaria para que regrese a la base de la torre cuando el reloj que ha quedado abajo muestre las 12:01. La segunda parte de la trayectoria de este reloj se parecerá a la trayectoria de las bolas en la figura VI.3 y, a pesar de las apariencias en el diagrama, la trayectoria habrá de ser recta. Ésta es una sencilla predicción empírica de la relatividad general —el hecho de que el reloj lanzado hacia arriba mostrará más tiempo transcurrido que cualquier otro reloj que registre el inicio y el final de estos eventos— y es una predicción que resulta ser correcta.

Los detalles matemáticos exactos de la curvatura del espacio-tiempo rebasan el alcance de nuestro texto, pero la formulación de la ecuación de campo de Einstein puede darnos al menos el tono de la teoría.[5] La ecuación tiene la forma siguiente:

$$G_{ab} = R_{ab} - \tfrac{1}{2} R g_{ab} = 8\pi T_{ab}.$$

La cantidad G_{ab} es el llamado *tensor de curvatura de Einstein*. La primera parte de la ecuación proporciona la definición de este tensor en función de la métrica del espacio-tiempo, del tensor de Ricci R_{ab} y la curvatura escalar R. Sea suficiente decir que todas éstas son cantidades puramente geométricas. Físicamente, el tensor de curvatura de Einstein en p determina en qué forma cambia el volumen de cualquier enjambre esférico pequeño de partículas en caída libre en p, inicialmente en reposo (unas respecto a las otras). Si el tensor de curvatura de Einstein es cero, entonces el volumen del grupo permanece constante. (Todo esto constituye, evidentemente, un comentario sobre cómo se comportan las geodésicas de tipo tiempo en la cercanía de p: las partículas en caída libre siguen tales geodésicas.)

Es esencial diferenciar el tensor de curvatura de Einstein, la métrica y el tensor de curvatura de Riemann R_{abcd} completo. Si este último es cero en todas partes, entonces el espacio-tiempo es plano, y las geodésicas localmente paralelas no son ni convergentes ni divergentes. Pero un tensor de Einstein nulo no garantiza que el espacio-tiempo sea plano: el volumen del enjambre esférico de partículas puede seguir siendo el mismo, aun cuando algunas de las partículas converjan, con la condición de que otras partículas diverjan. Este tipo de comportamiento distorsiona la esfera, y la convierte en un elipsoide, haciéndola más estrecha en una dirección y estirándola en una dirección ortogonal. Puesto que este tipo de distorsión de los océanos en la Tierra se deriva de la influencia gravitacional de la Luna que causa las mareas, se le califica como *efecto de marea*.

En el lado derecho de la ecuación, T_{ab} es el tensor de estrés-energía. Representa la distribución de la materia y la energía en el espacio. Por lo tanto, la ecuación de campo esencialmente expresa que la forma en que el volumen de las pequeñas bolas de partículas de prueba en caída libre se comporta en una región se determina por la cantidad de materia y energía en esa región. Mientras haya más materia y más energía, mayor será la curvatura de Einstein y más se encogerá el volumen de la bola.

Es importante señalar que el tensor de curvatura de Einstein no describe la geometría detallada, completa, del espacio-tiempo. Esta tarea se realiza por la métrica g_{ab} (o R_{abcd}, el cual es interdefinible con g_{ab}) de la que se derivan R_{ab} y R. Por lo tanto, la completa distribución de la materia y la energía, T_{ab}, no determina la geometría del espacio-tiempo, sino que la constriñe. Ya hemos visto un ejemplo de tal cosa: el hecho de que T_{ab} sea uniformemente cero es *compatible* con el hecho de que espacio-tiempo sea minkowskiano (es decir, plano), pero también compatible con la existencia de las ondas gravitacionales que producen los efectos de marea. La distribución de la materia y la energía *constriñe* la geometría del espacio-tiempo, pero no la *determina*.

Por lo anterior, en un nivel conceptual de gran amplitud, es posible decir que la relatividad general echa por tierra algunos de los aspectos más centrales de la relatividad especial. La relatividad especial propone una sola estructura plana del espacio-tiempo que es posible especificar totalmente mediante la existencia de los sistemas de coordenadas globales de Lorentz. La relatividad general plantea que la geometría espacio-temporal del universo depende de la distribución de la materia y de la energía, y además de otras condiciones de frontera. En la relatividad general no existen las coordenadas globales de Lorentz con relación a ningún universo material o incluso en relación con la mayoría de los universos vacíos. Uno podría sospechar que toda la física de la relatividad especial —la teoría electromagnética, la teoría de las fuerzas nucleares, etc.— tendrían que modificarse fundamentalmente para que fueran coherentes con la relatividad general.

Pero tal cosa resulta no ser verdad. De la misma manera que la superficie de la Tierra parece euclidiana en las pequeñas parcelas —es decir, las desviaciones respecto a la geometría euclidiana se hacen pequeñas si el área en consideración es pequeña—, los espacio-tiempos de la relatividad general se parecen al espacio-tiempo de Minkowski en regiones suficientemente pequeñas. De manera que si, al igual que las ideas políticas de Tip O'Neill, toda la física es local (es decir, que la física de una pequeña región del espacio-tiempo se determina completamente por lo que sucede sólo en esa pequeña región, por las leyes que hacen referencia únicamente a lo que existe en esa pequeña región), entonces las teorías de la relatividad especial pueden utilizarse en un escenario de la relatividad general.[6] Las desviaciones respecto a la relatividad especial sólo se manifiestan a gran escala, de manera que la apreciación del cambio de la relatividad especial a la relatividad general se hace más fácil cuando se trata de la astrofísica y la cosmología. La siguiente sección examina dos ejemplos de esto.

LOS AGUJEROS NEGROS Y EL *BIG BANG*

Los efectos gravitacionales son intrínsecamente mucho más tenues que los efectos de otras fuerzas de la naturaleza: las fuerzas nucleares fuertes y débiles, y el electromagnetismo. Nosotros sólo podemos darnos cuenta de la presencia de la gravedad porque sólo funciona de una manera: toda la materia y toda la energía producen una curvatura positiva einsteiniana, haciendo que las trayectorias localmente paralelas en caída libre converjan. Los grandes conglomerados de materia, tales como los planetas y las estrellas, generan grandes curvaturas einsteinianas. Esto no es verdad en el electromagnetismo, por ejemplo, en el que las influencias opuestas de las cargas negativa y positiva pueden hacer que los grandes conglomerados de materia sean efectivamente neutros.

La dinámica natural de la gravedad, por lo tanto, puede causar que la materia y la energía se apelmacen. El efecto de enfoque incrementa la densidad de la materia, fenómeno que a su vez incrementa la curvatura einsteiniana. La materia se vuelve más densa hasta que alguna fuerza contraria, tal como la presión o como la repulsión electrodinámica entre los átomos, produzca el equilibrio. Es esto precisamente lo que impide que toda la materia en la Tierra siga en trayectorias inerciales hasta el centro de la Tierra. Si la materia acumulada llegara a ser lo suficientemente densa, el electromagnetismo ya no podría impedir que los átomos se aceleraran y salieran de sus trayectorias inerciales, y que por lo tanto la materia se colapsara todavía más. Algunos efectos sutiles de la mecánica cuántica pueden contrarrestar el incremento de la densidad en cierto grado, y por lo tanto las estrellas colapsantes con cierto rango de masa se vuelven nuevamente estables como estrellas de neutrones. Pero como hemos de ver, más allá de un cierto nivel de densidad, la misma geometría del espacio-tiempo dicta el resultado: nada puede evitar el colapso. Semejante situación produce un agujero negro.

La geometría espacio-temporal de un agujero negro nos puede sugerir nociones sobre la situación más moderada de los planetas, las estrellas y las estrellas de neutrones: un agujero negro se les parece cuando se encuentra lo suficientemente lejos del centro. Aquí no tenemos los recursos matemáticos para describir con detalle la geometría de los agujeros negros, de manera que un diagrama cualitativo del espacio-tiempo tendrá que ser suficiente al respecto. Pero todas nuestras precauciones con relación a la interpretación de estos diagramas deberán redoblarse. Ya no podremos proyectar el diagrama de manera que las trayectorias rectas se correspondan con líneas rectas, y en él los intervalos no tendrán una relación patente con las distancias. De hecho, la única estructura geométrica que indicaremos con algún detalle es la estructura del cono de luz. E incluso esto requiere elegir arbitrariamente las convenciones pictóricas.

Adoptaremos el siguiente método:[7] la trayectoria de la luz que se dirige directamente hacia el centro del agujero negro siempre se representará por una línea a 45° respecto a la vertical, precisamente como hemos proyectado todos los rayos de luz en nuestros diagramas minkowskianos. Las líneas que representan los rayos de luz que se dirigen en otras direcciones divergen de 45°. En consecuencia, los conos de luz aparentemente se “inclinan” hacia el centro en los diagramas y también se “angostan”. Éste es simplemente un efecto de las convenciones representacionales: los conos de luz “angostos” e “inclinados” son inherentemente iguales a cualquier otro tipo de cono de luz. Las aparentes inclinación y angostura se parecen bastante a la aparente dimensión enorme de Groenlandia en una proyección de Mercator: las dimensiones relativas del mapa no son proporcionales a las áreas reales que se representan.

La figura VI.5 representa la geometría. Si hubiera un planeta o una estrella en el centro, los conos de luz gradualmente volverían a la vertical en el centro, pero el agujero negro tiene una geometría mucho más dramática todavía.

Se muestran las líneas de mundo de cuatro cohetes: un observador externo que permanezca a una distancia constante del agujero negro, un explorador que parte desde el observador externo en p y se interna en el agujero negro, y nuestros dos viejos amigos, el gemelo A y el gemelo B. El cilindro en el centro está formado por los rayos de luz verticales en el diagrama y representa el *horizonte de eventos* del agujero negro. La importancia física del horizonte de eventos es particularmente clara en el diagrama: puesto que los conos de luz de forma progresiva “se inclinan” al irse acercando (en el diagrama) a la singularidad, ninguna luz que se origine en el horizonte de eventos podrá escaparse de éste jamás. La luz puede viajar “verticalmente” a lo largo del horizonte de eventos, pero cualquier rayo de luz que se encuentre en realidad dentro del cilindro inevitablemente alcanzará la singularidad.

El comportamiento de la luz puede leerse directamente en el diagrama, puesto que los conos de luz se indican allí. Por ejemplo, el observador externo emite un rayo de luz en q, apuntando directamente al agujero negro. El explorador recibe el rayo de luz en s, lo cual significa que aún puede ver al observador externo cuando éste se vuelve y mira en dirección opuesta al agujero negro. Pero cuando el explorador intenta devolver la señal luminosa, dirigiendo desde s un rayo de luz hacia el todavía visible observador, sabemos que no lo va a lograr. Habiendo cruzado el horizonte de eventos, no es posible en absoluto que su rayo de luz pueda salir, y el rayo de luz termina en la singularidad.

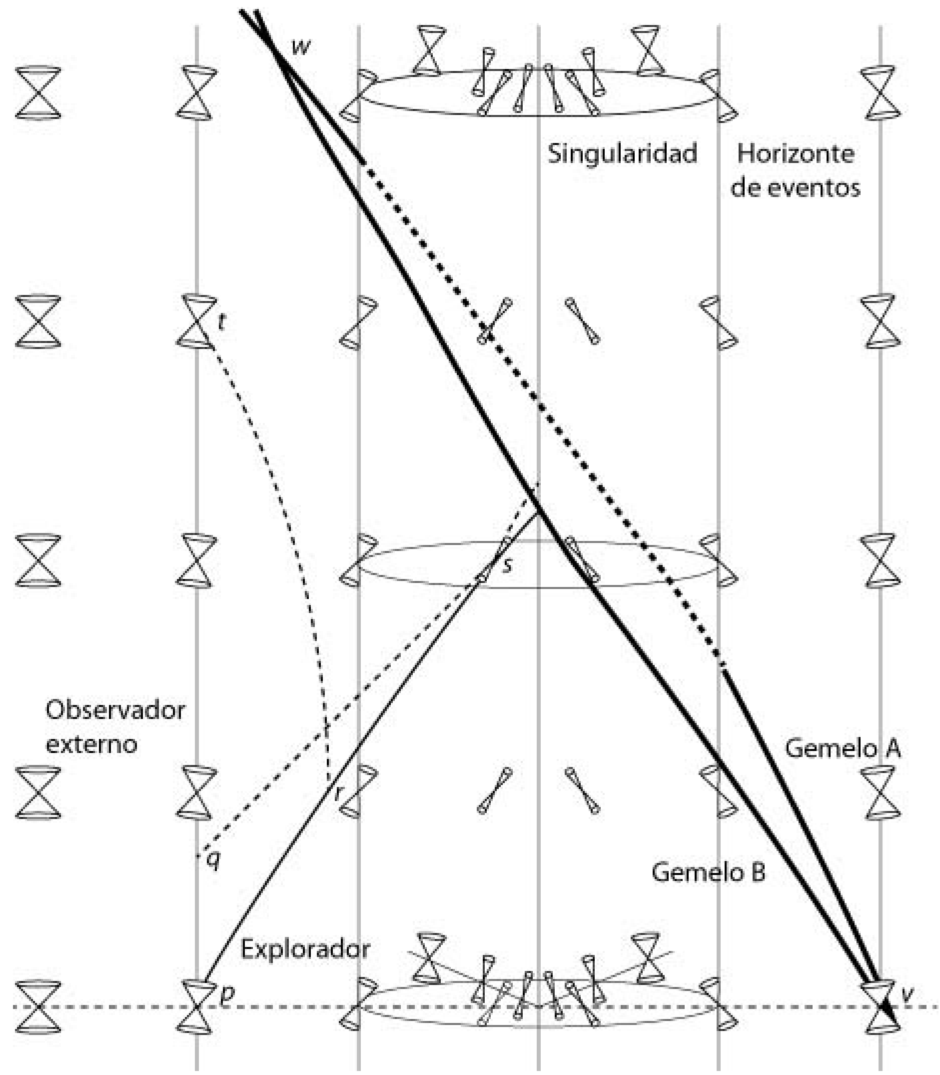

FIGURA VI.5

Los intentos que hace el explorador para comunicarse con el observador siguen siendo eficaces en *r*, como se indica en el diagrama. Sin embargo, puesto que los conos de luz se han "inclinado", la trayectoria de la luz tiene que iniciarse de manera casi vertical en el diagrama y tiene que esforzarse por llegar lentamente hasta el observador, quien recibe el mensaje en *t*. Aunque no es posible para nosotros leer directamente las longitudes de las líneas de mundo —y por lo tanto de los tiempos transcurridos— en el diagrama, la impresión general que podemos obtener es correcta: al irse acercando el explorador al horizonte de eventos, el observador tiene que esperar más para recibir los mensajes del explorador. Claro, un mensaje que se envíe desde el horizonte, apuntando hacia atrás, se volverá "vertical" en el diagrama y por lo tanto jamás se recibirá.

¿Qué le sucede al explorador? La representación de su línea de mundo termina en la singularidad del diagrama. El tiempo total transcurrido para él desde el punto *p* hasta la terminación es finito: sus relojes (si es que continúan en funcionamiento) jamás registran más que una cierta cantidad de tiempo. Según el diagrama, al parecer el explorador "se encuentra” con la singularidad, pero aquí la imagen puede ser engañosa. De hecho, la llamada “singularidad” en el diagrama es una línea que no se corresponde con parte ninguna del espacio representado. Esto no debería sorprender a nadie: si hubiera eventos físicos reales en la singularidad, tendrían que tener conos de luz futuros, pero puesto que los conos de luz que se angostan convergen todos desde diferentes direcciones, no sería posible ningún tipo de evolución suave hasta un solo cono de luz en el centro.

De hecho, la representación de la singularidad en el diagrama puede ser mucho muy engañosa. Para el observador incauto, la línea en el centro del horizonte de eventos se parece bastante a la línea de mundo de un objeto, como si la singularidad fuera un cierto tipo de *cosa* que yace en el centro mismo de un agujero negro. Sería incluso tentador considerarla como un punto infinitamente denso en el que toda la masa que contribuyó a formar el agujero negro se ha acumulado. Pero hay que darse cuenta de que la singularidad no es para nada una línea de *tipo tiempo*: atraviesa la parte superior de los conos de luz. Cualquier objeto masivo que se acerque a la singularidad llega a tener líneas de mundo inclinadas a 45° en el diagrama. Es el horizonte de eventos, todas las “líneas verticales” en el diagrama son de tipo espacio más que de tipo tiempo. Por lo tanto, “chocar con la singularidad” no es para nada semejante a chocar con algún objeto, y no se podría “evitar la singularidad” esquivándola, en la forma que se podría evitar el choque con un árbol. Más bien, la singularidad es un borde del espacio-tiempo mismo donde las curvas de tipo tiempo simplemente no pueden continuar. Como ha observado Sean Carroll, uno no podría evitar el choque con la singularidad más de lo que podría evitar el choque con el día de mañana: cualquier curva de tipo tiempo extendida al máximo dentro del horizonte tiene que terminar allí.

Al aproximarse el explorador a la singularidad, los efectos de marea gravitacionales se incrementan. A pesar de que la curvatura einsteiniana es esencialmente cero a lo largo de la línea de mundo, las trayectorias inerciales localmente paralelas se enfocan con intensidad en una dirección y se desenfocan en otras direcciones. El explorador se irá alargando y adelgazando en la dirección del agujero negro, y finalmente se desgarrará. Este efecto aumenta ilimitadamente al tiempo que la curvatura del espacio-tiempo crece sin medida en direcciones específicas. Y puesto que la ecuación de campo de Einstein no admite la curvatura infinita, no puede haber ninguna solución en absoluto en los puntos marcados como “singularidad” en el diagrama. Es decir, en la relatividad general clásica, estos puntos en el diagrama no representan eventos para nada. El espacio-tiempo mismo ha llegado a su fin.

Además del observador externo y del explorador, hemos delineado en la figura VI.5 al gemelo A y al gemelo B. Aquí rehabilitamos una aseveración que hicimos en el capítulo iv: el fenómeno de los gemelos puede tener lugar en la relatividad general sin que haya aceleración en absoluto. El gemelo A y el gemelo B sincronizan su reloj en el evento *v*. A pesar de las apariencias en el diagrama, sus sendas trayectorias son perfectamente rectas: el “doblamiento” se debe a la curvatura del espacio-tiempo mismo. Gracias al efecto gravitacional, los gemelos se reúnen nuevamente en el evento *w*, y uno de ellos típicamente será más viejo que el otro. Las longitudes de las dos trayectorias rectas de *v* a *w* generalmente no serán iguales: cuál será la más larga dependerá de los detalles exactos. Este ejemplo deberá eliminar la idea de que el fenómeno de los gemelos se relaciona de alguna manera con la aceleración.

La idea de que el espacio-tiempo finaliza con la singularidad en un agujero negro naturalmente sugiere la posibilidad contraria: el espacio-tiempo también podría empezar con una singularidad. Y efectivamente, si se hace una retrodicción con base en la expansión actual del universo, se llega a un universo temprano cuya densidad aumenta en la medida en que el tiempo se remonta. Si asumimos que la relatividad general es válida en todos los niveles de energía, el resultado será el *Big Bang:* una singularidad inicial en la historia del universo.

Obviamente, postular singularidades asociadas con agujeros negros y el *Big Bang* depende de una adhesión absoluta a la ecuación de campo de Einstein. Si la relatividad general deja de funcionar en el nivel de alta energía o de gran curvatura, entonces todas las apuestas deben retirarse. Además, estos modelos postulan que es posible representar exactamente la distribución de la materia y la energía en el universo mediante un tensor de estrés-energía. No resulta claro cómo esta suposición puede tener coherencia con la explicación mecánico-cuántica de la materia.

El problema de incorporar la explicación mecánicocuántica de la materia en una teoría como la relatividad general es particularmente agudo, porque si es cierto que el tensor de estrés-energía representa la distribución local de la materia y la energía en el espacio-tiempo, no es para nada claro qué tipo de distribución de la materia o energía localmente definidas existen en una teoría mecánico-cuántica. Esto plantea problemas conceptuales en la comprensión de la teoría cuántica que son mucho más profundos que su mera reconciliación con la relatividad. Examinaremos estos problemas en el segundo volumen de esta obra.

La idea de que el espacio-tiempo en sí mismo pueda tener un inicio o un fin es el corolario de la relatividad general más intrigante desde el punto de vista filosófico, aunque también el más insustancial. Muchos físicos trabajan en teorías según las cuales el *Big Bang* tan sólo es la fase más antigua en nuestro rincón local del universo, no el principio de toda la existencia universal. Especulaciones similares sugieren que el viaje del explorador no tiene que terminar necesariamente en una singularidad, aunque parece ser probable que los efectos de marea fueran fatales en todos los casos. Pero cualquier tipo de planteamiento riguroso de estos temas requeriría una física más avanzada de la que actualmente poseemos.

EL ARGUMENTO DEL AGUJERO

En los capítulos II y III de la presente obra examinamos un par de argumentos que Leibniz planteó para impugnar el espacio absoluto y el tiempo absoluto newtonianos. Ambos argumentos se basan en una cierta simetría que existe en la física newtoniana. El llamado argumento del *desplazamiento estático* utiliza la simetría traslacional y rotacional del espacio euclidiano con el fin de establecer que, para cada posible distribución de la materia, la distribución que se genera desplazando todo en una dirección o rotando todo en un ángulo determinado en torno a un eje es también una distribución posible. Si la distribución original en sí misma es incapaz de mostrar la simetría, entonces la nueva distribución será un estado distinto, físicamente posible, coherente con todas las leyes de Newton. De forma similar, el *desplazamiento cinemático* cambia por una cantidad fija las velocidades absolutas de toda la materia, nuevamente generando una posibilidad física distinta. La aceptación de estos pares como posibilidades distintas supuestamente violaba tanto el principio de la identidad de los indiscernibles como el principio de la razón suficiente, este último en relación con la decisión de Dios respecto a la creación de una u otra de las posibilidades. Nosotros señalamos en aquellos capítulos que el principio de la identidad de los indiscernibles nada tiene que lo recomiende y que la inyección de consideraciones teológicas en un argumento físico es inaceptable.

El desplazamiento cinemático, pero no el estático, también indica que Newton se había entregado a ciertas cuestiones respecto a los hechos físicos que no era posible solucionar mediante la observación. En particular, la exacta magnitud y dirección de la velocidad absoluta de cualquier porción de materia no podía determinarse experimentalmente, tal como se demuestra en el caso del barco de Galileo. Pero una vez que abandonamos el espacio y tiempo absolutos de Newton por el espacio-tiempo galileano o el espacio-tiempo de la relatividad especial, esa dificultad desaparece: ahora ya no habrá velocidades absolutas cuyas magnitudes y direcciones sean inaccesibles para nosotros.

Puesto que tanto el desplazamiento estático como el cinemático recurren a simetrías del espacio-tiempo, se podría esperar que estos tipos de cuestiones no se dieran en la relatividad general. El espacio-tiempo de la relatividad general simplemente no posee semejantes simetrías. Puesto que la curvatura del espacio-tiempo varía de lugar en lugar con la densidad de la energía, una solución genérica de la ecuación de campo de Einstein no puede ser ni geométricamente homogénea ni isotrópica. No es posible definir las operaciones del desplazamiento estático y del desplazamiento cinemático.

No obstante, John Earman y John Norton han propuesto un argumento (basado en un conjunto de consideraciones diferentes del mismo Einstein) de que un problema relacionado y más radical contamina a la relatividad general. Ellos lo han llamado el *argumento del agujero*.[8] Con el propósito de valorar el argumento, recordemos los tres tipos de transformaciones que aparecen en la figura I.1. Los argumentos del desplazamiento de Leibniz utilizan isometrías: las distancias relativas, y por ende velocidades relativas, de los objetos se preservan en la transformación, mientras que la estructura métrica del espacio-tiempo permanece inalterable. El argumento del agujero, en contraste, propone una transformación topológica. En específico, plantea un *difeomorfismo del agujero:* una transformación topológica suave que constituye el mapa de la identidad en el exterior de alguna región (el “agujero”) pero no la identidad en el interior. La figura VI.6 muestra un difeomorfismo de agujero semejante. Debemos entender que los dos diagramas se proyectan en *exactamente el mismo* fondo, por lo que las ubicaciones marcadas como *p* y *q* en los dos diagramas se supone que representan *exactamente los mismos eventos*. Lo único que ha cambiado es la forma en que la materia y los campos se distribuyen en esos eventos.

A primera vista, podría parecer que si el estado *S* es una solución a las ecuaciones de campo einsteinianas, el estado S_{CURVO} no podría serlo. Por ejemplo, sabemos que un rayo de luz en un vacío debe tener una línea de mundo recta, pero el rayo de luz en S_{CURVO} parece extremadamente curvo. El artificio en el difeomorfismo del agujero consiste en que cuando movemos las cosas de acuerdo con la transformación topológica, no sólo movemos la materia y la luz, sino *la métrica del espacio-tiempo también*. La métrica del espacio-tiempo se manipula como si fuera un campo electromagnético y por lo tanto se transporta por el difeomorfismo. Si la línea punteada en *S* es recta, concordando con la métrica original, entonces la línea punteada en S_{CURVO} es recta, concordando con la métrica transformada. Puesto que la ecuación de campo relaciona el comportamiento de la materia con la métrica del espacio-tiempo, cuando transformamos conjuntamente la materia y la métrica el nuevo objeto sigue siendo una solución a la ecuación.

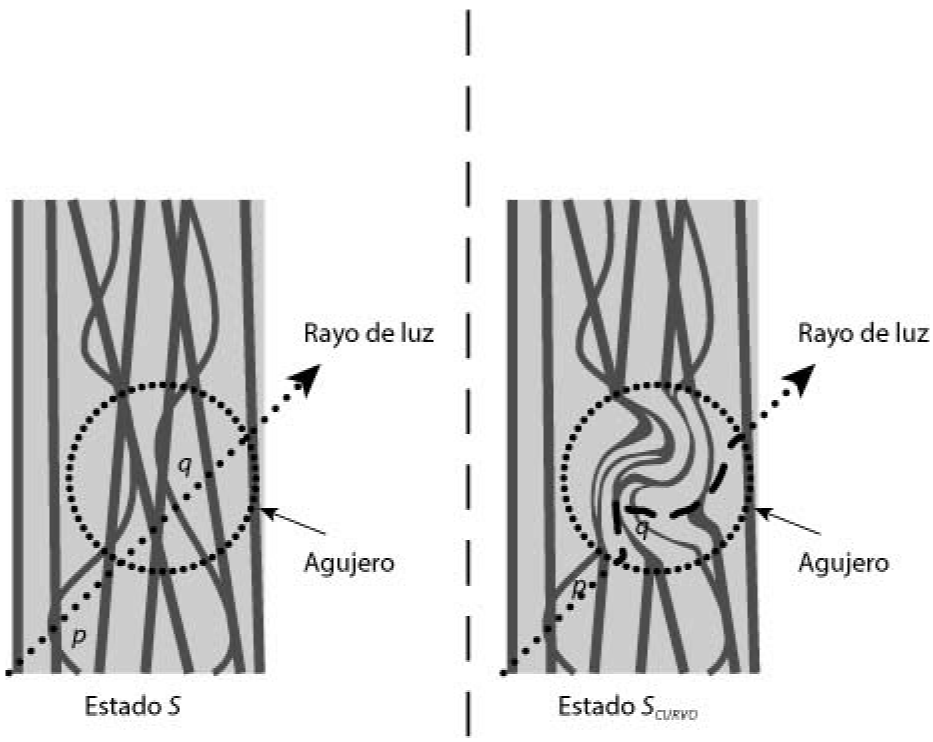

FIGURA VI.6

Para que *S* y S_{CURVO} representen diferentes situaciones físicas posibles, tiene que haber alguna diferencia física en lo que cada una representa. Es aparentemente posible especificar semejante diferencia si aceptamos que *p* y *q* representan exactamente los mismos eventos en cada uno de los diagramas. Es que en *S, p y q* tienen una separación de tipo luz, mientras que en S_{CURVO} tienen una separación de tipo espacio (el rayo de luz los separa uno del otro en S_{CURVO} pero no en *S*). Es esencial en el argumento del agujero que el espacio topológico de fondo se considere como "fijo" cuando tanto la materia como la métrica "se muevan de un lado a otro" entre ambas situaciones. El argumento del agujero se deriva de la propuesta de que tanto *S* como S_{CURVO} representan una situación físicamente posible, una solución a las leyes físicas fundamentales, y además de que estas dos situaciones son físicamente distintas.

Si aceptamos esta propuesta, entonces el argumento toma una forma diferente de cualquiera de los argumentos del desplazamiento de Leibniz. Ya hemos visto que la consecuencia más perjudicial que resulta de un argumento del desplazamiento de Leibniz consiste en la existencia de hechos físicos empíricamente inaccesibles, como la velocidad absoluta de un objeto. El argumento del agujero no parece involucrar este tipo de hechos empíricamente inaccesibles. Por ejemplo, no se puede decir que es empíricamente inaccesible el hecho de que los eventos denominados *p* y *q* tengan, *en efecto*, una separación de tipo luz o una separación de tipo espacio: sólo hay que enviar luz desde *p* hasta *q* y ver si llega a *q*. Pues los términos singulares *p* y *q* no denotan mágicamente eventos específicos: para aplicar tales etiquetas a eventos del espacio-tiempo específicos, necesitaríamos algún modo de diferenciar tales eventos de otros al designarlos. Y si podemos elegir *p* y *q* para empezar, es que podemos distinguir (en principio) si tienen una separación de tipo tiempo, de tipo espacio o de tipo luz.

La fuerza del argumento del agujero no tiene nada que ver con los hechos empíricamente inaccesibles: más bien, es un planteamiento sobre el *determinismo*. Si aceptamos que *S* y S_{CURVO} representan situaciones peculiares que son físicamente posibles pero ontológicamente diferentes, entonces el problema surge porque *en el exterior del agujero las situaciones físicas son idénticas*. Resulta crucial en el argumento que el difeomorfismo de agujero no modifique el exterior del agujero. En este sentido, el difeomorfismo del agujero no es similar a los desplazamientos de Leibniz: todos éstos modifican la ordenación de la materia a través del tiempo. Pero si el exterior del agujero es exactamente el mismo en ambos casos, y si tanto *S* como S_{CURVO} son soluciones físicamente posibles de las ecuaciones de campo, entonces las ecuaciones de campo mismas tienen que ser indeterministas radicales: la situación física en su totalidad en el exterior del agujero, junto con las leyes de la física, no determinan la situación física en el interior del agujero. Puesto que el agujero puede ser de cualquier tamaño y de cualquier forma, y puesto que la comba puede asumir muchas y muy diferentes formas, esto significa un indeterminismo de la especie más radical.

El meollo del argumento del agujero es evidentemente una propuesta metafísica a profundidad, a saber, que es posible suponer que *S* y S_{CURVO} representan dos situaciones metafísicamente distintas pero físicamente posibles. Puesto que las diferencias entre *S* y S_{CURVO} sólo pueden expresarse en función de los eventos nombrados, específicos (en vez de los aspectos cualitativos, como la existencia de un par de partículas que colisionan, o de un agujero negro), debe abordarse la cuestión metafísica de la identidad de los individuos en diferentes "mundos posibles". Jeremy Butterfield[9] responde al argumento del agujero diciendo que S y S_{CURVO} (que son meras representaciones matemáticas) no pueden representar situaciones *metafísicamente distintas*, puesto que un evento en S sólo puede identificarse con un evento en S_{CURVO} mediante alguna relación de equivalencia. Pero cualquier relación de equivalencia equipara el evento p en S con el evento que el difeomorfismo marca como p en S_{CURVO}, y lo mismo ocurre con q. De manera que si p y q tienen una relación de tipo luz en S, sus equivalentes tienen una relación de tipo luz en S_{CURVO}. Según la semántica de las locuciones modales de David Lewis, esto significa que si p y q realmente tienen una relación de tipo luz, no es posible que haya un mundo donde no tengan una relación de tipo luz.[10]

Si se adopta el punto de vista de Saul Kripke sobre la semántica modal, emerge una solución diferente. Kripke insiste en que no es necesario determinar mediante relaciones de equivalencia las identidades que pasan de un mundo a otro: es posible estipularlas directamente al describir una situación.[11] Así, un punto específico en S_{CURVO} representa el evento real$p$ simplemente porque al construir S_{CURVO} yo digo que lo representa. Supongamos que S representa de forma correcta el mundo real, y que en el mundo real yo califico a dos eventos de p y q, respectivamente. Entonces S_{CURVO} en efecto representa los eventos p y q con una relación de tipo espacio en vez de una relación de tipo luz (como en realidad son). A diferencia del planteamiento de Butterfield, S y S_{CURVO} representan situaciones *distintas*. Pero no es para nada obvio que S_{CURVO} represente una situación física o metafísicamente *posible*. Yo puedo decir “Si Nixon fuera un sándwich de jamón…” y de este modo producir una representación según la cual Nixon pudiera ser un sándwich de jamón, pero de ninguna manera se deriva de mi frase que sería*posible* que Nixon hubiera sido alguna vez un sándwich de jamón. Es contrario a la naturaleza esencial del Nixon real, el hecho de que hubiera sido un sándwich de jamón, a pesar de mis estipulaciones lingüísticas. De manera similar, podemos decir que si los eventos reales p y q en particular tienen una relación de tipo luz, entonces no es metafísica o físicamente posible que esos mismos eventos hayan tenido una relación de tipo espacio: las relaciones espacio-temporales entre los eventos del espacio-tiempo son esenciales en su identidad. El planteamiento de esta respuesta se explaya en Maudlin.[12]

Una respuesta más, la tercera, ha sido propuesta por Carl Hoefer y Nancy Cartwright.[13] Ésta acepta el indeterminismo radical que Earman y Norton proponen, aunque insiste en que este indeterminismo es inofensivo y nada problemático. Aunque S y S_{CURVO} sean diferentes situaciones físicas posibles, les parecerán exactamente iguales a cualquier criatura que pueda existir allí. Ningún experimento que sea posible idear puede tener en S un resultado observable distinto del que pueda tener en S_{CURVO}. De la misma manera que en el caso del desplazamiento estático, ninguna pregunta comprensible sobre el mundo real puede carecer de respuesta por motivo de la inaccesibilidad empírica. Earman y Norton piensan que el indeterminismo sigue siendo inaceptable, aun cuando no tenga consecuencias empíricas, porque no se ha derivado de ninguna fuente físicamente plausible. Pudiera ser que el determinismo fuera falso, dicen, pero al menos debería dársele una oportunidad de defenderse. Y en efecto, debería ser evidente que la versión básica del argumento del agujero podría aplicarse con tanta certeza en el espacio-tiempo galileano y newtoniano como en la relatividad general: si los eventos del espacio-tiempo sólo tienen una estructura geométrica propia de manera contingente, entonces el argumento del agujero tiene que funcionar en cualquier espacio-tiempo.

Por lo tanto, existen muchas vías de respuesta al argumento del agujero.[14] Aun así, el argumento podría ser la inspiración de una explicación radicalmente nueva del espacio y el tiempo que sea inmune a las impugnaciones. Hasta la fecha, ninguna alternativa semejante ha surgido. El tiempo se pronunciará al respecto.

LECTURAS RECOMENDADAS SOBRE LA RELATIVIDAD GENERAL

La matemática de la relatividad general es compleja, y este capítulo tan sólo ofrece una visión de conjunto vaga y cualitativa. Existen textos en todos los niveles que la exponen con más detalle. En el libro de R. Geroch se proporciona una discusión con un nivel técnico muy moderado que puede servir como introducción al tema. Entre los manuales de física que enfatizan los fundamentos geométricos se encuentran el de S. Carroll y las referencias cibernéticas de J. C. Baez y E. F. Bunn, así como las secciones iniciales de la obra de C. Misner, K. Thorne y J. Wheeler. Otros textos atractivos son el de R. Wald y, en un nivel en un nivel profundamente matemático, el de S. W. Hawkins y G. F. R. Ellis. Entre los filósofos que presentan y analizan la teoría detalladamente se encuentran M. Friedman y J. Earman.[15]

[1] Para una explicación extremadamente clara de algunos de estos intentos, véase J. Norton, “Einstein, Nordstrom and the Early Demise of Lorentz Covariant, Scalar Theories of Gravitation”, *Archive for History of Exact Sciences*, 45: 17-94, 1992.

[2] G. Galilei, *Two New Sciences*, University of Wisconsin Press, Madison, 1974, p. 66. Si las cifras de Galileo son correctas, no puede estar describiendo unos experimentos que se llevaran a cabo en la Torre Inclinada de Pisa, puesto que ésta sólo tiene la mitad de la altura mencionada.

[3] Este ejemplo es discutido por N. Huggett en *Space from Zeno to Einstein*, MIT Press, Cambridge, 1999, p. 136.

[4] Véase, por ejemplo, R. Feynman, R. Leighton y M. Sands, “Motion in Curved Space-Time”, en *The Feynman Lectures on Physics*, vol. 2, Addison-Wesley, Reading, Massachusetts, 1975, pp. 42-12 y ss.

[5] Para los textos que dan explicaciones apropiadas y rigurosas de la relatividad general, véanse las lecturas recomendadas al final de este capítulo. Esta explicación del tensor de la curvatura de Einstein se encuentra en J. C. Baez y E. F. Bunn, “The Meaning of Einstein’s Equation”, University of California, Riverside, 2006.

[6] La mecánica cuántica parece sugerir que no toda la física *es* local, como veremos en el segundo volumen. Además, es posible hacer observaciones técnicas más detalladas con relación a cómo una teoría de la relatividad especial puede adaptarse a un escenario de la relatividad general. Véase S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*, Addison-Wesley, Boston, 2004, sec. 4.7.

[7] Esta convención se utiliza, por ejemplo, en R. Geroch, *Relativity from A to B*, The University of Chicago Press, Chicago, 1978, cap. 8. La nuestra es una versión resumida del mismo.

[8] J. Earman y J. Norton, “What Price Space-Time Substantivalism? The Hole Story”, *British Journal for the Philosophy of Science*, 38: 515-525, 1987; J. Earman, *World Enough and Space-Time*, MIT Press, Cambridge, 1989, cap. 9.

[9] J. Butterfield, “The Hole Truth”, *British Journalfor the Philosophy of Science*, 40(1): 1-28, 1989.

[10] Véase, por ejemplo, D. Lewis, *On the Plurality ofWorlds*, Basil Black-well, Oxford, 1986, cap. 4.

[11] S. Kripke, *Naming and Necessity*, Harvard University Press, Cambridge, 1980, pp. 14 y 44.

[12] T. Maudlin, “The Essence of Space-Time”, en A Fine y J. Leplin (comps.), *Procedings of the Philosophy of Science Association Meetings*, vol. 2, Philosophy of Science Association, East Lansing, 1989.

[13] T. C. Hoefer y N. Cartwright, “Substantivalism and the Hole Argument”, en J. Earman *et al*. (comps.), *Philosophical Problems of the Internal and External Worlds*, University of Pittsburgh Press, Pittsburgh, 1994.

[14] Para una discusión más completa de la bibliografía respectiva, véase J. Norton, “The Hole Argument”, en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, Stanford University, Palo Alto, 2008.

[15] R. Geroch, *General Relativity from A to B*, The University of Chicago Press, Chicago, 1978; S. Carroll, *op. cit*., J. C. Baez y E. F. Bunn, *op. cit*., C. Misner, K. Thorne y J. Wheeler, *op. cit*., R. Wald, *General Relativity*, University of Chicago Press, Chicago, 1984; S. W. Hawking y G. F. R. Ellis, *The Large-Scale Structure of Space-Time*, Cambridge, University Press, 1973; M. Friedman, *Foundation of Space-Time Theories*, Princeton University Press, Princeton, 1986; J. Earman, *World Enough and Space-Time, op. cit*., y J. Earman, *Bangs, Crunches, Whimpers, and Shrieks*, Oxford University Press, Oxford, 1995.

VII. DIRECCIÓN Y TOPOLOGÍA DEL TIEMPO

LA GEOMETRÍA DEL TIEMPO

El tiempo es unidimensional. La relevancia exacta de esta aparente trivialidad se entiende fácilmente cuando se acepta la explicación newtoniana del tiempo absoluto. El tiempo absoluto está hecho con instantes o momentos, cada uno de los cuales contiene infinitamente múltiples eventos que tienen lugar de manera simultánea. En un ensayo denominado "Sobre la gravedad y el equilibrio de los cuerpos", al hacer una analogía, Newton utilizó una descripción sorprendente de un instante del tiempo y de la relación entre el tiempo y el espacio.[1]

> Pues nosotros no adscribimos varias duraciones a las diferentes partes del espacio, sino que decimos que todas perduran juntas. El momento de duración es el mismo en Roma y en Londres, en la Tierra y en las estrellas, y a través de todos los firmamentos. Y de la misma manera que entendemos que cualquier momento de duración se encuentra difundido a través de todos los espacios, de acuerdo con su especie, sin pensar en absoluto en sus partes, no resulta contradictorio que pensemos que también la mente, de acuerdo con su especie, pueda encontrarse a través del espacio sin pensar para nada en sus partes.

Para Newton, un solo momento del tiempo abarca todo el espacio, y el tiempo absoluto es la totalidad de los momentos. Decir que según Newton el tiempo absoluto es unidimensional, entonces es decir simplemente que estos momentos tienen una geometría unidimensional: una "línea temporal". Lo que caracteriza a la geometría unidimensional, en contraste con la geometría de más dimensiones, es que en cualquier punto únicamente hay (a lo sumo) dos direcciones: izquierda y derecha, digamos, o arriba y abajo, o adelante y atrás. De manera similar, en cualquier momento dado del tiempo absoluto newtoniano, sólo existen dos direcciones temporales a partir de ese momento: hacia el futuro y hacia el pasado.

Decir que una cosa es unidimensional en este sentido no determina toda su geometría. El ecuador es unidimensional, y en cualquiera de sus puntos sólo puede haber dos direcciones en que nos podríamos mover si permaneciéramos en él: hacia el este y hacia el oeste. Pero en el ecuador existe además la característica de que si viajáramos lo suficientemente lejos hacia el este o hacia el oeste, a la larga retornaríamos al punto de partida. Dado cualquier punto p en el ecuador, los demás puntos no se dividen entre los que se encuentran en el este y los que se encuentran en el oeste de p: todos los puntos (incluyendo el mismo p) pueden alcanzarse mediante un viaje que se realice en cualquiera de las dos direcciones. No es éste un hecho sobre el carácter dimensional del ecuador, sino sobre su topología.

Newton hubo de asumir, sin comentarlo, que la topología del tiempo es diferente: si uno viaja indefinidamente hacia el futuro, jamás habrá de regresar al tiempo presente. El pasado, decimos, ha terminado definitivamente: jamás podrá regresar. La topología del tiempo, según Newton y conforme al sentido común, no es como la topología del ecuador: cualquier evento que no esté sucediendo ahora, se encuentra ya sea en el pasado o ya sea en el futuro, pero no en ambos al mismo tiempo. La línea temporal del universo, para Newton, no es un círculo.

La explicación newtoniana del tiempo absoluto se basa en la idea de la simultaneidad absoluta. Los instantes del tiempo "se encuentran difundidos a través de todos los espacios" de manera exacta porque un par de eventos distintos pueden ocurrir objetivamente "al mismo tiempo". Pero en la relatividad no existe la simultaneidad absoluta, ningún momento se difunde a través del espacio. La totalidad del tiempo no es el conjunto de todos los instantes universales. Más bien, la estructura temporal se define sobre la base de los eventos individuales en el espacio-tiempo. Por lo tanto, tenemos que retornar a ciertas cuestiones sobre el carácter dimensional y topológico del tiempo.

Incluso sin la simultaneidad absoluta, el tiempo en la relatividad puede caracterizarse como localmente unidimensional. Consideremos, por ejemplo, el caso del evento *p* en la figura VI.5. En un sentido, hay un número infinito de formas de moverse continuamente en el tiempo desde *p*: cualquier trayectoria de tipo tiempo o de tipo luz a través de *p* representa un conjunto continuo de eventos con un orden temporal definido. Pero la estructura del cono de luz en *p* divide estas trayectorias en dos tipos distintos que se corresponden con los dos conos de luz. Cualquier dirección a partir de *p* hacia el interior de un cono de luz o en uno de los conos de luz puede desplazarse continuamente en cualquier otra dirección hacia el interior o en ese cono de luz: imaginemos que torcemos una flecha que apunta hacia "arriba" desde *p* para convertirla en otra flecha semejante. Pero no podríamos transformar fácilmente una flecha que apunta hacia arriba en una flecha que apunta hacia abajo sin pasar a través de algunas direcciones de tipo espacio. Así que de la misma manera que sólo sería posible moverse hacia el este o el oeste en el ecuador, el conjunto de direcciones de tipo tiempo y de tipo luz desde *p* se divide en exactamente dos clases, las cuales describimos como la que se dirige hacia el futuro y la que se dirige hacia el pasado.

El cambio del tiempo absoluto, que se compone de instantes, a la estructura relativista del cono de luz genera posibilidades extrañas. O por lo menos es posible construir *matemáticamente* algunos muy curiosos objetos que localmente parecen ser iguales al espacio-tiempo relativista. Algunos de estos casos matemáticos no parecen representar con claridad posibilidades físicas reales del espacio-tiempo, mientras que otros son motivo de feroces debates. Empecemos con uno tan extremo que nadie considera que pueda ser una posibilidad física.

Todos los eventos en el espacio-tiempo relativista poseen un cono de luz que divide las direcciones de tipo tiempo y de tipo luz en dos clases. Por lo general, a éstas se les llama cono de luz del pasado y cono de luz del futuro, respectivamente. Y en cualquier modelo matemático que los físicos puedan tomar en serio, esta diferencia es *global:* habiéndose seleccionado el cono de luz futuro en cualquier evento en particular, existe una forma específica de determinar el cono de luz futuro en cualquier otro evento en particular. Volvamos de nuevo a la figura VI.5 y observemos que el lóbulo superior del cono de luz en *p* es el cono de luz futuro. ¿Pero cuál es la situación del evento *t*? ¿El hecho de determinar cuál será el cono de luz futuro en*p* puede determinar lo mismo respecto a *t* también?

Debería ser obvio que sí lo podría determinar. Imaginemos, por ejemplo, seguir la línea de mundo del observador externo continuamente desde *p* a *t*. Si seleccionamos cualquier flecha dirigida hacia el futuro en *p* y la llevamos a lo largo de la línea de mundo, sin permitir jamás que asuma una dirección de tipo espacio en el trayecto, entonces habrá de apuntar hacia el interior del lóbulo superior del cono de luz cuando llegue a *t*. Y lo mismo es cierto si tomamos la ruta de *p* a *r* y luego de *r* a *r*, como se muestra en el diagrama, sin permitir jamás que la dirección de la flecha se convierta en una de tipo espacio. Para cualquier ruta de *p* a *t* (incluso una de tipo espacio) el resultado será el mismo: cualquier dirección que apunte hacia el lóbulo superior en *p* continuamente seguirá una dirección que apunte hacia el lóbulo superior en *t*, siempre que no se permita que la dirección se convierta en una de tipo espacio en el trayecto. De manera que denominar cono de luz *futuro* a un lóbulo en *p* y cono de luz *pasado* al otro también determina la distinción entre las futuras direcciones y las pasadas direcciones en todos los otros puntos del espaciotiempo. Se califica de *temporalmente orientable* a semejante espacio-tiempo.

Es posible describir los espacio-tiempos no orientables como un ejercicio puramente matemático. La superficie no orientable más conocida es la cinta de Möbius, y ésta puede utilizarse para modelar un espacio-tiempo no orientable de dos dimensiones. Hay que dibujar conos de luz de forma que un circuito de éstos a lo largo de la cinta resulte en que el cono de luz "se dé la vuelta": al retornar al punto de partida *p*, la dirección que originalmente se denominaba "futuro" ahora apunta hacia el lóbulo contrario (figura VII.1).

De manera local, en cualquier porción suficientemente pequeña, la estructura que se muestra en la figura VII.1 es isomorfa a un espacio-tiempo de dos dimensiones estándar (y el mismo truco puede hacerse en el caso de un espacio-tiempo de cuatro dimensiones). La estructura parece extraña sólo en relación con su topología global. Pero este aspecto de todos modos resulta fatal: ningún físico vería en la figura VII.1 una representación de un espacio-tiempo posible. La división en conos de luz pasados y futuros debe ser permanente y global. La figura VII.1 no es más que una curiosidad matemática sin importancia física.

Ésta es una lección crucial: incluso si pudiéramos describir una estructura matemática que en todos aspectos pareciera como si fuera *localmente* una estructura del espacio-tiempo posible, no significaría que el objeto en su totalidad correspondiera a una posibilidad física. Hay formas menos radicales de manipular la topología global de una representación matemática, y no se puede saber cuál de éstas describiría configuraciones reales del espacio-tiempo. La figura VII.2 muestra otras dos formas de "empaquetar" un espacio-tiempo minkowskiano de dos dimensiones: uno lo cierra en la dirección de tipo espacio y el otro en la dirección de tipo tiempo.

El espacio-tiempo de la izquierda en la figura VII.2 ha sido enrollado —el término oficial es "compactado"— en una dirección de tipo espacio. Las secciones transversales circulares del cilindro forman "curvas cerradas de tipo espacio" que de manera abstracta se corresponden con el ecuador en la Tierra. De la misma manera que es posible circunnavegar la Tierra dirigiéndose constantemente hacia el este o hacia el oeste, es posible circunnavegar este universo siguiendo una trayectoria inercial recta de tipo tiempo. Recordemos que estos dos espacio-tiempos se parecen localmente al espacio-tiempo de Minkowski: no hay una sola curvatura espacio-temporal en ninguna parte. (El hecho de que un cilindro normal no posee una curvatura intrínseca es fácil de comprobar, puesto que uno puede enrollar una hoja de papel plana para hacer un cilindro sin estirarla ni deformarla.) Así, el espacio-tiempo minkowskiano compactado de tipo espacio nos permite nuevamente crear una situación del estilo de la paradoja de los gemelos sin ninguna aceleración. Tanto el gemelo A como el gemelo B en la figura VII.2 tienen trayectorias no aceleradas, pero el gemelo A registrará menos tiempo transcurrido entre sus sucesivos encuentros que el gemelo B.

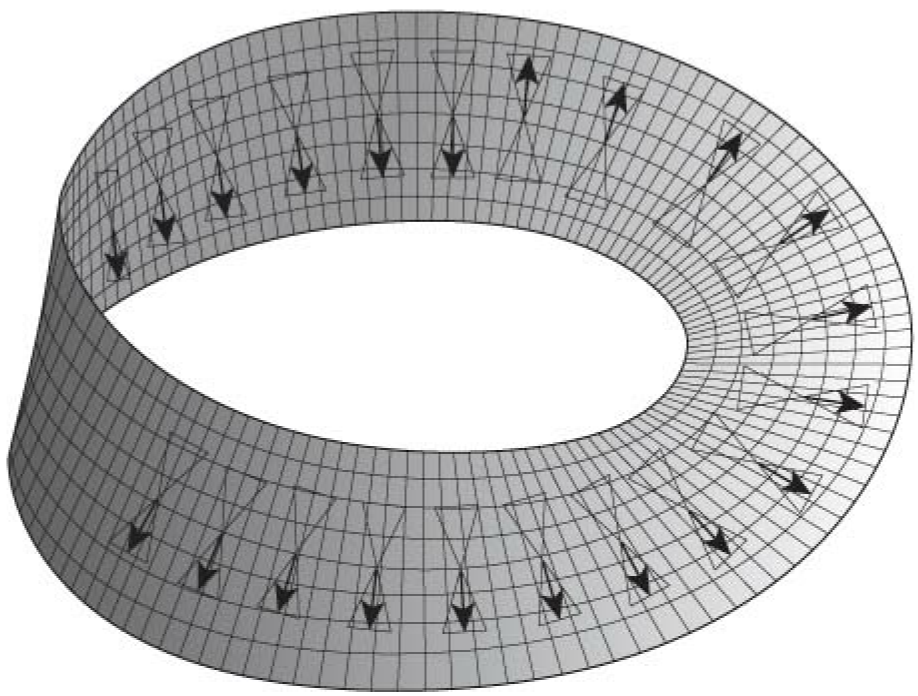

FIGURA VII.1

En un sentido obvio, el espacio-tiempo compactado de tipo espacio es "espacialmente finito": los gemelos A y B tan sólo pueden ocupar una cierta cantidad de espacio. Por mucho que lo intenten, jamás se alejarán uno del otro más que la mitad de la circunferencia del cilindro, y al añadirse más observadores el espacio se llenará más.

El espacio-tiempo de la izquierda en la figura VII.2 posee algunas propiedades interesantes, pero el espacio-tiempo de la derecha es profundamente extraño. En este último caso, el espacio-tiempo minkowskiano se ha compactado en una dirección de tipo tiempo y los cortes transversales circulares del cilindro forman *curvas cerradas de tipo* *tiempo* (CCT). Este espacio-tiempo, al igual que el espacio-tiempo compactado de tipo espacio, es temporalmente orientable: habiéndose determinado cuál de los lóbulos de un solo cono de luz es el cono de luz pasado y cuál el cono de luz futuro, se da una división única de todos los conos de luz en conos de luz del pasado y del futuro. Pero en el espacio-tiempo compactado de tipo tiempo es posible regresar al mismo evento tras haber seguido una trayectoria continua hacia adelante en el tiempo. La figura VII.2 muestra la línea de mundo de un solo reloj. Cuando éste marca 1:00, se encuentra en movimiento inercial hacia la derecha en el diagrama y continúa moviéndose hasta chocar con un objeto cuando marca 1:01 y súbitamente acelera (la línea de mundo se dobla). Pero el objeto contra el cual choca es *el mismo reloj exactamente*. Si usted fuera un observador con el reloj, la secuencia de los eventos se desarrollaría de la siguiente manera: a la 1:00 usted estaría en movimiento inercial (aparentemente "en reposo") al darse cuenta de que un objeto se aproxima. La colisión ocurre a la 1:01, impulsando el objeto (que resulta ser un reloj con un observador) hacia adelante. Mirando hacia atrás (es decir, en la dirección contraria al lugar donde apareció el objeto), usted entonces vería acercarse "desde atrás" otro reloj con un observador. A la 1:03 tiene lugar otra colisión, impulsándolo a usted "hacia adelante" y dejando al reloj que chocó detrás.

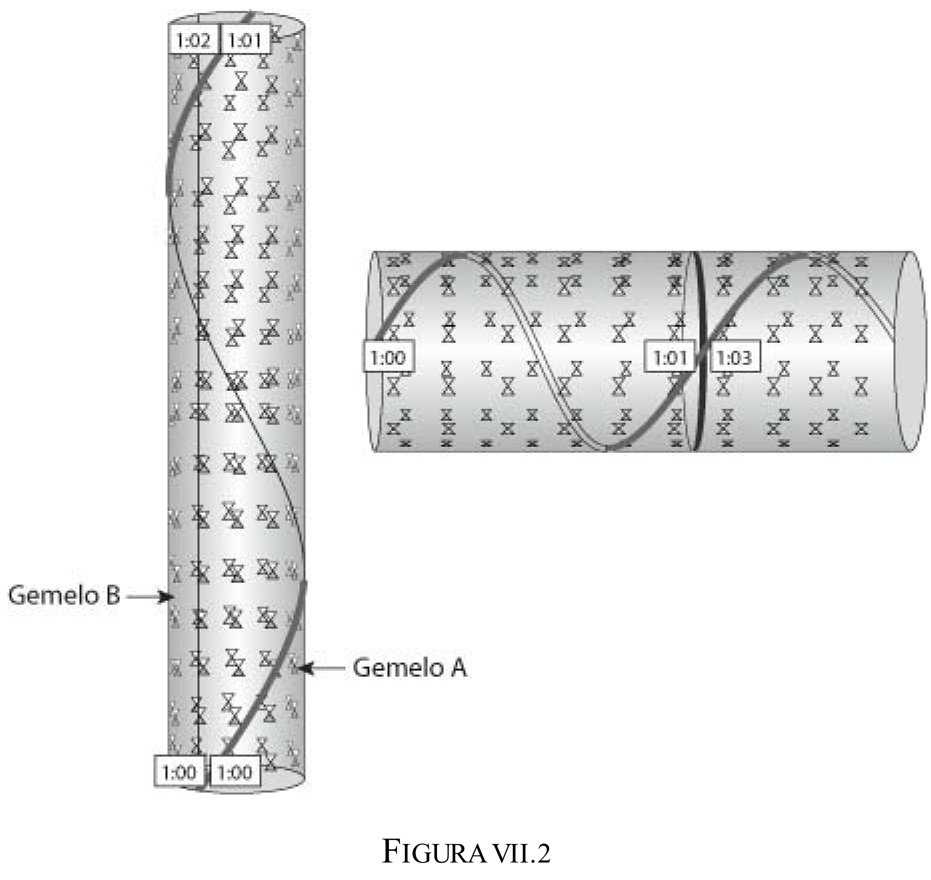

FIGURA VII.2

Es obvio que en realidad tan sólo hay una colisión en la historia de este universo: el choque del reloj en su etapa "anterior" con el reloj en su etapa "posterior". Y en este universo hay tan sólo un objeto material: un reloj. Sin embargo, la estructura temporal del universo se halla cerrada de tal forma que ningún evento pertenece exclusivamente al "pasado" o al "futuro" de cualquier otro evento: todo evento se encuentra tanto en el cono de luz pasado como en el cono de luz futuro de todos los demás eventos, incluso en los conos de luz futuro y pasado de sí mismo. A diferencia del espacio-tiempo compactado de tipo espacio, ya no se puede saber objetivamente cuál evento es anterior o posterior a otro. (Es divertido trazar los rayos de luz en este universo: resulta fácil comprobar que nuestro observador siempre será capaz de ver tanto las etapas pasadas como las futuras de sí mismo, refiriéndose el "pasado" y el "futuro" aquí al tiempo que marca el reloj.)

En resumen, la modificación de la topología del espacio en esta forma permite en cierto sentido viajar en el tiempo: un observador que siempre "va hacia adelante" en el tiempo puede regresar a un evento que ya ha experimentado. En cierta forma, esto es similar a la experiencia de un viajero que al moverse hacia el este continuamente en el ecuador al final regresa a casa, pero en un sentido más profundo es diferente: el viajero en el ecuador "vuelve" a un conjunto de eventos que son distintos a los que había conocido y su casa pudo haber sufrido cambios en el ínterin. El viajero en el tiempo regresa al mismo evento exactamente. Claro es que no hay posibilidad ninguna de "cambiar el pasado": la colisión en la figura VII.2 sólo sucede una vez, en una forma específica, aun cuando el observador experimente el evento dos veces, jugando un rol diferente cada vez. Esta circunstancia genera ciertas paradojas que examinaremos en la próxima sección.

Un enigma relacionado con los espacio-tiempos que tienen esta peculiar topología temporal se muestra en la figura VII.2. Obsérvese que el reloj que marca 1:00 bien podía haber continuado eternamente —dada la estructura del espacio-tiempo y dada la circunstancia de que es el único objeto material que existe— sin colisionar con nada, siguiendo una trayectoria helicoidal ininterrumpida. O podría haber más colisiones de las que aparecen en la figura VII.2. Así que hay una curiosa especie de subdeterminación respecto a lo que le pueda suceder al reloj, aun cuando la física pueda ser totalmente determinista en todas las regiones pequeñas del espacio-tiempo.

Nuestro concepto inicial de la estructura temporal del universo, entronizado en el tiempo absoluto de Newton, es que el universo entero se encuentra en cualquier momento dado en un estado específico, y su estado subsecuente se determina únicamente por el estado precedente (si las leyes de la naturaleza son deterministas). Incluso el abandono de la noción de la simultaneidad absoluta en la relatividad no socava una parte de este escenario fundamental. El sustituto relativista de un momento de tiempo absoluto se conoce como una *superficie de Cauchy*. Una superficie de Cauchy es una porción de un espacio-tiempo relativista que toda curva inextensible de tipo tiempo interseca una vez exactamente. Por lo tanto, en la figura VII.2, cualquier sección transversal circular u oblicua (cortada a menos de 45°) del espacio-tiempo compactado de tipo espacio es una superficie de Cauchy: cualquier línea de tipo tiempo, tal como la línea de mundo del gemelo A o del gemelo B, llegará hasta la sección transversal tan sólo cuando se extienda lo suficientemente lejos. Un espacio-tiempo relativista en que sea posible una superficie de Cauchy se califica de *globalmente hiperbólico*. Y bajo las suposiciones usuales sobre los espacio-tiempos relativistas, cualquier espacio-tiempo globalmente hiperbólico no sólo posibilita una sola superficie de Cauchy: permite una foliación de un conjunto de superficies de Cauchy. Por ejemplo, nuestro espacio-tiempo compactado de tipo espacio puede cortarse para formar una pila de secciones circulares o de secciones cortadas oblicuamente. Si las leyes de la naturaleza son deterministas y relativistamente locales, entonces este modo de "rebanar" un espacio-tiempo puede jugar el mismo rol que el del seccionamiento del espacio-tiempo que forma las rebanadas de simultaneidad: el estado físico total en cualquiera de estas rebanadas, junto con las leyes de la naturaleza, determina el estado físico total de la totalidad del espacio-tiempo.[2] Cada uno de estos posibles seccionamientos del espacio-tiempo en superficies de Cauchy proporciona un modo de ver el estado total del universo en el avance de su evolución en el tiempo.

Pero el espacio-tiempo compactado de tipo tiempo de la figura VII.2 no admite la posibilidad de las superficies de Cauchy. Puesto que allí hay curvas cerradas de tipo tiempo (CCT), una sola trayectoria de tipo tiempo puede intersecar cualquier región más de una vez, de la misma forma que la trayectoria de nuestro reloj interseca la región de la colisión más de una vez. De manera que no es posible pensar que este universo en su totalidad de alguna manera avanza en el tiempo desde un estado dado. Ésta es una idea más radicalmente revisionista que el simple abandono de la simultaneidad absoluta.

Nuestro ejemplo del espacio-tiempo no orientable de Möbius ya constituyó una especie de advertencia: no todo objeto matemático bien definido representa forzosamente una posibilidad física. El espacio-tiempo del "viaje en el tiempo" de la figura VII.2 está bien definido matemáticamente, y sobre la base de nuestros principios generales podemos describir en qué forma los relojes y los rayos de luz se comportarían en semejante situación. Pero nada de esto es una prueba de que el espacio-tiempo físico real admita semejante posibilidad. Aquellos que piensen que el tiempo esencialmente involucra un ordenamiento asimétrico de los eventos, que ningún evento puede ser a la vez anterior y posterior a sí mismo (en el sentido de que una curva de tipo tiempo recurre al mismo evento), se pueden permitir el rechazo de la posibilidad física de un espacio-tiempo con curvas cerradas de tipo tiempo. A diferencia de los agujeros negros, los cuales indudablemente se crean (de acuerdo con la teoría general) cuando una cantidad suficiente de materia se encierra en un volumen suficientemente pequeño, no existe una condición física conocida que determine la creación de las curvas cerradas de tipo tiempo.[3] La compactación de los modelos en la figura VII.2 se realizó a mano: tienen la topología que muestran porque nosotros la decidimos, no porque la física la hubiera determinado. De manera similar, los físicos pueden construir modelos que contienen curvas cerradas de tipo tiempo, pero éstas no surgen naturalmente a partir de condiciones físicas conocidas.[4] Así que aunque resulte divertido pensar en los problemas de los viajes en el tiempo, es necesario hacerlo con un cierto escepticismo: no existe una razón física o empírica real que nos permita pensar que sea físicamente posible viajar hacia nuestro pasado local.[5]

Puesto que la coherencia matemática del modelo no demuestra que los viajes en el tiempo representan una posibilidad física real, entonces acaso la cuestión puede determinarse de otra manera mediante la lógica. Desde hace mucho tiempo se ha pensado que los viajes en el tiempo, en el sentido de las curvas cerradas de tipo tiempo, tendrían que generar paradojas o contraindicaciones lógicas. Pero si una contradicción no puede ser verdad, cualquier suposición que implique una contradicción no puede ser verdadera. ¿Es así la suposición de la posibilidad de los viajes en el tiempo?

La paradoja estándar de los viajes en el tiempo consiste en viajar al pasado y hacer algo que pruebe que el viajero jamás pudo haber viajado al pasado, lo cual es una contradicción patente. Un ejemplo dramático sería el hecho de asesinar a mi abuelo antes de que él tuviera hijos. Si mi abuelo no hubiera tenido hijos a la sazón, entonces yo obviamente no hubiera jamás nacido: uno de mis progenitores jamás hubiera existido. O, en la misma línea, yo podría viajar el pasado y matar al niño que yo fui, de manera que yo jamás hubiera crecido para convertirme en un viajero del pasado. Es claro que estos tipos de ficciones son incoherentes, y por lo tanto imposibles. Si viajar en el tiempo de algún modo implicara este tipo de historias, entonces los viajes en el tiempo también serían imposibles.

Pero una vez que examinamos con detalle el meollo de la cuestión, resulta muy difícil describir cómo alguien podría emprender un viaje en el tiempo que tan sólo produjera resultados contradictorios o paradójicos. En el caso del autoasesinato, por ejemplo, es posible razonar: si voy al pasado y mato al niño que fui, entonces no estaría aquí ahora dispuesto a viajar al pasado, lo cual es una contradicción. Pero si no viajo al pasado y me mato, entonces no hay nada que me impida hacerlo, de manera que lo haré: otra contradicción. Ahora bien, a un nivel pedestre, la segunda afirmación es obviamente muy cuestionable: el hecho de que no haya regresado al pasado para matar al niño que fui, no significa para nada que nada me pudiera impedir hacerlo si quisiera intentarlo. Todo tipo de eventos podría impedir que llevara a cabo con éxito el intento del asesinato de mí mismo: un arrepentimiento, la heroica intervención de mis padres, etc. Todas éstas pueden calificarse de soluciones *lógicamente* posibles, tal como la solución donde un arrepentimiento me impide, para empezar, viajar en el tiempo, o donde la heroica intervención de mis (hoy ancianos) padres me impide asesinarme a mí mismo. Sin embargo, este tipo de escenarios donde aparece *deus ex machina* son patentemente arbitrarios: se salva la coherencia lógica, pero no en virtud de ningún tipo de principio rector fundamental.

Una solución mucho más satisfactoria surge de la consideración de que es demostrable que, bajo ciertas limitaciones físicas muy plausibles, siempre puede haber una forma coherente de escenificar el proyecto sin ninguna intervención externa: el sistema autointeractivo mismo proporciona sus propias posibilidades coherentes. Por ejemplo: yo viajo al pasado con la firme intención de asesinarme a mí mismo en la infancia. Emprenderé el viaje en un cierto día de mi niñez en el cual sé que mis padres estarán ausentes; me lleno de rabia para estar seguro de que no me voy a arrepentir, etc. Remonto el tiempo y me veo como niño en la cuna, indefenso. Alzó mi revólver para disparar… pero mi mano tiembla y en vez de matar al niño sólo lo hiero en el hombro. De hecho, el daño a los músculos y nervios causado por ese disparo me sigue afectando durante toda la vida y en particular es la causa de ese tremor en el día fatal. Tanto la paradoja como la incoherencia se resuelven, pero sin apelar a la intervención de un factor externo al sistema autointeractivo. La física de ese sistema por sí misma permite una historia consistente.

Este círculo causal cerrado,

tremor → disparo fallido → daño neural → tremor

podría parecer tan milagroso como la intervención de los padres, pero en cuanto asunto puramente matemático es posible probar con base en suposiciones generales que siempre existe una evolución coherente de este tipo en un sistema autointeractivo cerrado.[6] En general, únicamente necesitamos asumir que la física es determinista y continua, y que el espacio de los estados físicos posibles en un sistema posee una forma geométrica sencilla. Los físicos han estudiado este mismo problema, en relación con bolas de billar que se envían al pasado a través de agujeros de gusanos con el fin de interactuar con su estado precedente, y han encontrado exactamente el mismo resultado.[7] De hecho, lo que se encuentra sobre la base de suposiciones plausibles es que, lejos de que no haya *ninguna* solución a las ecuaciones físicas en tales casos, existen típicamente *muchas* soluciones. La situación es exactamente como la de nuestro reloj en la figura VII.2: conforme a las leyes, el reloj podría chocar consigo mismo cuando marca 1:01, tal como se muestra, o podría chocar consigo mismo de la misma manera anterior o posterior, o simplemente no chocar.

Hay que subrayar de nuevo que no es posible que la trama de un viaje en el tiempo proporcione la posibilidad de ir al pasado y *modificarlo*, es decir, hacer diferente al pasado de lo que en realidad era. Tal cosa es sencillamente imposible. Puesto que sólo tuve una infancia, en ella sufrí un ataque casi fatal que me dejó herido o no lo experimenté. Si en efecto lo viví, y si viajar en el tiempo es posible, entonces acaso el atacante era una versión de mí mismo de más edad. Si no lo viví, entonces no lo viví, punto, aun cuando, por causa de un viaje en el tiempo, una versión de mí mismo de más edad estuvo presente en mi niñez. El pasado —al igual que el presente y el futuro— tan sólo existe una vez y sólo de una cierta manera. Éste no es un profundo principio metafísico. Decir que un evento sucedió de una cierta manera “la primera vez” y de otra cierta manera “la segunda vez” no tiene sentido, puesto que no hay ni una primera vez ni una segunda vez.

Este trivial detalle es difícil de apreciar porque muchas de las historias ficticias sobre viajes en el tiempo son incoherentes y por lo tanto imposibles. Marty McFly, en el filme *Back to the Future* [*Volver al futuro]*, no puede *cambiar* la forma en que sus padres se conocieron accidentalmente, por la sencilla razón de que se conocieron como lo hicieron.

Si es capaz de viajar al pasado, de jugar un rol en su encuentro, puede *influir* en la forma en que sucedió. Pero no es posible que hubiera un conjunto de eventos en los que él no influyó y luego un segundo conjunto de eventos en los que sí influyó, puesto que sólo hubo un conjunto de eventos.

Las ficciones inconsistentes sobre los viajes en el tiempo usualmente sólo se preocupan por la *contingencia* radical en la vida: el hecho de que la forma en que la vida de una persona se desarrolla depende de muchos pequeños factores que están más allá de nuestra comprensión y control. Esta contingencia radical aparece en el poema de Robert Frost, *The Road Not Taken [El camino no tomado]*, y en películas que nos cuentan cómo *hubiera sido* la vida de alguien si algún hecho en apariencia trivial hubiera sucedido de manera distinta. Ejemplos de este tipo son *Sliding Doors [Si yo hubiera…]* y *Lola rennt [Corre, Lola, corre]*, así como *Przypa-dek [El azar]*, de Krzysztof Kieslowski. Estas películas cuentan la misma historia varias veces, con base en pequeñas variantes que producen resultados distintos. Las ficciones de los viajes en el tiempo con frecuencia sólo se preocupan realmente por este tipo de pregunta contrafáctica: ¿qué hubiera pasado si…? Claro que es una pregunta importante e interesante, pero tratar de contestarla mediante un viaje en el tiempo es un ejercicio que lleva a incoherencias.

Claro que también ha habido historias de viajes en el tiempo consistentes. *Twelve Monkeys* [*Doce mono*] (basada en *La Jetée*, de Chris Marker) es un buen ejemplo de éstas, como también el episodio de “Time’s Arrow” [“La flecha del tiempo”] en *Star Trek: La nueva generación*. Acaso el ejemplo más notable es el del cuento “All You Zombies” [“Todos ustedes, zombis”] de Robert A. Heinlein, en el cual el personaje principal termina por ser tanto su padre como su madre, todo narrado de manera lógicamente impecable. Estas historias no intentan dar una idea de lo que pudiera haber sido, puesto que simplemente narran los detalles específicos de lo que realmente sucedió. Se podría pensar que estas ficciones consistentes de viajes en el tiempo no podían haber sucedido tal como se narran, pero los resultados matemáticos arriba mencionados muestran que no es así.

La mera consistencia lógica de viajar en el tiempo no muestra que sea posible ni física ni metafísicamente en ningún sentido real. “Richard Nixon es un sándwich de jamón” es una afirmación *lógicamente* consistente. Es cierto que es falsa. Pero es más importante todavía que no sea *posible* describir un escenario que le pueda otorgar veracidad. Simplemente *no puede* ser cierta en ningún caso: el mundo no pudo haber sido de tal forma que pudiera contener tanto a Nixon como al hecho de que este hombre fuera un sándwich de jamón. Una “posibilidad lógica” no es en realidad una especie de posibilidad: significa meramente que la lógica por sí sola no prohíbe la veracidad de una afirmación. De manera similar, ni la lógica ni la matemática pueden negar la posibilidad de viajar en el tiempo, como se ve en la figura VII.2. Pero de la misma manera que la naturaleza de Nixon le impide ser un sándwich de jamón, acaso la naturaleza del tiempo le impide al tiempo ser circular. Es difícil saber cómo hay que plantearse una cuestión de este tipo.

El protagonista autogenerado de “Todos ustedes, zombis” nos incomoda de la misma manera que lo hace el círculo cerrado de la mano temblorosa con la pistola. Intuitivamente, la causación es una relación asimétrica y por ende irreflexiva: si A es la causa de B entonces B no puede ser la causa de A, de manera que ninguna cosa puede generarse a sí misma. E intuitivamente, la asimetría de la causación es ella misma un parásito de una asimetría fundamental del tiempo. El futuro surge del pasado. La dirección temporal del pasado al futuro es fundamentalmente disimilar de la dirección del futuro al pasado en una forma que carece de una analogía espacial. Únicamente hay dos direcciones a lo largo del ecuador, pero el ecuador mismo no corre más del este al oeste que del oeste al este. El tiempo, al contrario, pasa del pasado al futuro, y todos inevitablemente nos encaminamos hacia la tumba, alejándonos de la cuna.

La dirección del tiempo se arraiga tan profundamente en nuestros conceptos lingüísticos que es imposible extirparla. La mayoría de los verbos tienen un sentido temporal: la diferencia entre una piedra que cae y esa misma piedra que sube está determinada por cuál de las direcciones del tiempo se relaciona con el pasado y cuál con el futuro. Los vocablos “hacia” y “desde” usualmente se aplican con referencia al tiempo: los procesos se desarrollan desde las etapas anteriores a las posteriores. Recordamos el pasado y anticipamos el futuro. Nuestras acciones actuales pueden tener una influencia en los eventos del futuro, pero son incapaces de influir en los eventos del pasado.

Decir que la dirección temporal del pasado al futuro es intrínsecamente diferente de la dirección del futuro al pasado, no es lo mismo que decir que el futuro tiene que ser cualitativamente diferente del pasado. El físico Fred Hoyle propuso una teoría del “estado estacionario”: el universo en su totalidad se ha estado expandiendo y sigue expandiéndose incesantemente, apartándose las galaxias unas de otras cada vez más. Pero en la teoría de Hoyle la densidad de la materia en promedio no disminuye: la nueva materia se crea en los espacios vacíos en la medida precisa para que la densidad permanezca estable. En este modelo, el tiempo en su totalidad —futuro, pasado y presente— son cualitativamente iguales: de ahí se deriva el “estado estacionario”. Pero sigue habiendo una dirección fundamental del tiempo, en virtud de la cual es correcto decir que el universo se *expande, creando* nueva materia, en vez de decir que se *contrae, destruyendo* la materia existente.

Nuestro mundo está lleno de procesos que muestran una dirección en el tiempo evidente. Si se sumergen cubos de hielo en agua caliente y se aíslan de otras influencias externas, los cubos se derriten y producen agua tibia. El agua tibia, aislada de influencias externas, jamás forma espontáneamente cubos de hielo y agua caliente. Sin embargo, las leyes de la física en apariencia permiten estos últimos procesos tanto como aquéllos. De nuevo, nuestra descripción misma de ambos procesos presupone una dirección en el tiempo: la única diferencia entre el derretimiento y el congelamiento es la dirección del tiempo. De manera similar, nuestra descripción de un agujero negro en la figura VI.5 presupone una dirección en el tiempo: es porque la dirección del futuro se representa por "arriba" en el diagrama en vez de por "abajo", que decimos que las cosas *caen dentro* y *jamás se escapan* del horizonte de eventos, en vez de decir que *se expulsan* y *no pueden permanecer dentro* éste. De la misma manera que las leyes fundamentales de la física aparentemente permiten tanto el derretimiento espontáneo como el congelamiento espontáneo del hielo, la relatividad general permite tanto los agujeros negros como sus reversiones temporales. Y de la misma manera que el hielo con frecuencia se derrite pero jamás se forma en el agua tibia, en el universo aparentemente hay muchos agujeros negros pero sus reversiones temporales no existen.

Por lo tanto, es posible plantear dos preguntas diferentes respecto a la dirección del tiempo. Una consiste en preguntar por qué aparentemente hay procesos que jamás tienen lugar aunque sus reversiones temporales ocurran regularmente. Al plantearse en términos de, digamos, el derretimiento y el congelamiento espontáneos, esta pregunta presupone que hay una diferencia fundamental entre la dirección del pasado al futuro y la dirección del futuro al pasado. Pero algunos filósofos y físicos han extraído una conclusión mucho más sorprendente del hecho de que las leyes de la física al parecer permiten ambos tipos de procesos: han concluido que no existe una diferencia fundamental entre ambas direcciones. Por ejemplo, Paul Horwich afirma: "Yo argumentaría que las evidencias empíricas actuales indican que el tiempo mismo es intrínsecamente simétrico".[8] Huw Price dice: "No se ha aceptado propiamente que no tenemos derecho a asumir que la entropía *aumenta* en vez de *disminuir*, por ejemplo. El hecho objetivo es que existe una pendiente entrópica a lo largo del tiempo, y no que el universo 'recorra' esta pendiente en una dirección en vez de en otra".[9]

Consideremos las secuelas de estas afirmaciones. La diferencia entre *aumentar* y *disminuir* se deriva de la dirección del tiempo: la entropía aumenta si se vuelve más grande hacia el futuro y decrece si se vuelve más grande hacia el pasado. Por tanto, si no existe un hecho objetivo respecto al hecho de que la entropía aumente o disminuya en algún sitio, entonces no existe un hecho objetivo respecto a cualquier afirmación que utilice un verbo con una dirección temporal. En un sentido objetivo profundo, no nos encaminamos "hacia" nuestra muerte ni "nos alejamos" de nuestro nacimiento; en ningún sentido objetivo nos hacemos más viejos en vez de más jóvenes. De hecho, la idea de que el tiempo "pasa" resulta ser, según este punto de vista, una especie de ilusión. Las etapas posteriores del universo no surgen fundamentalmente de las anteriores, y tampoco las anteriores de las posteriores. La noción de la causa y el efecto, o la producción de una cosa a partir de otra, es igualmente ilusoria.

Ésta es una manera de ver lo que está en juego en el debate. Si la dirección temporal del futuro al pasado es en realidad cualitativamente idéntica a la dirección del pasado al futuro, entonces no hay nada obvio y físicamente imposible respecto al espacio-tiempo no orientable que aparece en la figura VII.1. Aparte de la distinción entre las dos direcciones temporales, la estructura del cono de luz del espacio-tiempo es perfectamente lisa por doquier. Pero si las dos direcciones temporales son fundamentalmente distintas, entonces no es posible que semejante estructura no orientable represente un espacio-tiempo posible, puesto que es inconsistente con relación a la naturaleza del tiempo. Ya que ningún físico o filósofo ha sugerido jamás la posibilidad física de un espacio-tiempo temporalmente no orientable, la noción de que las dos direcciones son fundamentalmente diferentes en apariencia se ha aceptado de manera por lo menos tácita.

No es fácil hacer la distinción entre estas dos cuestiones en la literatura. La cuestión de la direccionalidad en sí de los procesos —la razón por la cual los cubos de hielo se derriten de forma espontánea en agua caliente pero el agua tibia no se segrega espontáneamente en cubos de hielo y agua caliente— se plantea en la física estadística, la cual es uno de los temas del segundo volumen de esta obra. La idea de que el tiempo no tenga una direccionalidad intrínseca es más difícil de evaluar. Uno podría caer en la tentación de adoptar esta idea mediante la utilización irreflexiva de los *diagramas* de espacio-tiempo, los cuales son objetos puramente espaciales que carecen de direccionalidad intrínseca. La dirección del pasado al futuro que se representa mediante un diagrama de este tipo debe indicarse mediante una convención adicional, puesto que el medio mismo de la representación no posee semejante asimetría. Pero también la diferencia entre la representación de las direcciones de tipo espacio y tipo tiempo en el diagrama tiene que indicarse mediante una convención. De forma similar, las representaciones puramente *matemáticas* del espacio-tiempo requieren muchas convenciones, para decidir, por ejemplo, si los intervalos de tipo tiempo se deben representar con números reales o con números imaginarios, o si una coordenada del tiempo aumenta o disminuye en la dirección del pasado al futuro. Nada de lo anterior sugiere incluso vagamente que las dos direcciones temporales no sean fundamentalmente diferentes, y tampoco sugiere que las direcciones de tipo tiempo no difieran intrínsecamente de las direcciones de tipo espacio.

Yo discuto algunos de los argumentos sobre el paso del tiempo en “On the Passage of Time”.[10] Pero puesto que no es posible revisar estos temas adecuadamente sin un conocimiento básico de la materia y de las leyes que la gobiernan, así como de la física estadística, una discusión más elaborada de la dirección del tiempo tendrá que esperar la aparición del segundo volumen de esta obra.

[1] El ensayo “De Gravitiatione” puede consultarse en A. R. Hall y M. B. Hall (comps.), *Unpublished Papers of Isaac Newton: A Selection from the Portsmouth Collection in the University Library*, Cambridge University Press, Cambridge, 1962, y se encuentra resumido en N. Huggett, *op. cit.*, p. 113.

[2] Para una discusión exhaustiva del tema, véase J. Earman, *A Primer on Determinism*, Reidel, Dordrecht, 1986.

[3] Véase J. Earman, C. Smeenk y C. Wuthrich, “Do the Laws of Physics Forbid the Operation of a Time Machine?”, *Synthese*, 169 (1): 91-124, 2009.

[4] Aquí estamos obligados a hacer una observación de orden técnico. En un cierto sentido, es obvio que *cualquier* geometría del espacio-tiempo puede hacerse de manera que sea coherente con la ecuación de campo de Einstein. Tomamos una métrica arbitraria, calculamos el tensor de curvatura einsteiniano, y entonces usamos la ecuación de campo para *definir* el tensor de estrés-energía. Así que cuando juzgamos que algunas geometrías “no son físicas” nos basamos en que consideramos que algunos tensores de estrés-energía no son físicos. Una condición estándar, la llamada condición de energía débil, esencialmente exige que en los marcos de Lorentz locales la densidad de la energía local no sea negativa. Es posible formular otras condiciones de energía semejantes (*cf*. S. W. Hawking y G. F. R. Ellis, *op. cit.*, sec. 4.3). Por lo tanto, el intento de encontrar un escenario físico plausible para la formación de las curvas cerradas de tipo tiempo significa encontrar una que satisfaga condiciones de energía razonables, que dé como resultado una solución estable y que no involucre curvas cerradas de tipo tiempo que ya se encuentren en las condiciones iniciales.

[5] El lector podría preguntarse: si el espacio-tiempo de Minkowski puede compactarse en una dirección temporal y en una dirección espacial, ¿se podría compactar en una dirección de género luz, formando curvas cerradas de género luz? Esto también es matemáticamente posible y es fácil de hacer en dos dimensiones. Se lo dejo al lector como un ejercicio. Obsérvese que ninguno de los modelos en la figura vii.2 posee ningún tipo de geodésica cerrada de género luz.

[6] Véase C. J. S. Clarke, “Time in General Relativity”, en J. Earman, C. Glymour y J. Stachel (comps.), *Minnesota Studies in the Philosophy of Science*, vol. 8, University of Minnesota Press, Minneapolis, 1977, y F. Arntzenius y T. Maudlin, “Time Travel and Modern Physics”, en E. N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, Stanford University, Palo Alto, 2013.

[7] F. Echeverría, G. Klinkhammer y K. S. Thorne, “Billiard Ball in Wormhole Spacetimes with Closed Timelike Curves: Classical Theory”, *Physical Review D*, 44 (4): 1077-1099, 1991.

[8] P. Horwich, *Asymmetries in Time*, MIT Press, Cambridge, 1987, p. 38.

[9] H. Price, *Time’s Arrow and Archimedes’ Point*, Oxford University Press, Oxford, 1996, p. 48.

[10] T. Maudlin, *The Metaphysics within Physics*, Oxford University Press, Oxford, 2007, cap. 4, “The Passage of Time”.

APÉNDICE: ALGUNOS PROBLEMAS EN LA FÍSICA DE LA RELATIVIDAD ESPECIAL

HE PRESENTADO LA RELATIVIDAD ESPECIAL en una forma algo atípica, enfocándome primordialmente en la geometría del espacio-tiempo de Minkowski y sólo de forma secundaria en las coordenadas de Lorentz. Las coordenadas tienen un rol de mera conveniencia, a partir del cual se facilita el cálculo del intervalo entre los eventos. Mediante este método es posible solucionar ciertos tipos de problemas sin la necesidad de determinar las transformaciones de Lorentz. También he recurrido mucho al empleo de los diagramas de espacio-tiempo, señalando las diferencias y las similitudes entre el espacio euclidiano y el espacio-tiempo minkowskiano.

La clave para solucionar estos problemas está en la utilización de un marco de Lorentz conveniente. Por un marco conveniente entiendo uno en que los cálculos se hagan con facilidad: podemos utilizar cualquier marco que queramos para obtener los resultados correctos, pero la matemática es más sencilla en algunos de ellos. La selección de un marco conveniente usualmente implica la decisión de considerar que un objeto masivo en movimiento inercial se encuentra "en reposo". La línea de mundo de este objeto será una línea vertical en nuestro diagrama de espacio-tiempo. Si este objeto es un reloj, habrá de marcar los valores coordenados t en este marco lorentziano.

Una vez que hayas decidido cuál objeto se encontrará "en reposo", dibuja un diagrama de espacio-tiempo. No olvides que los objetos en movimiento inercial y que los rayos de luz en un vacío se representan ambos por líneas rectas, y que las trayectorias de los rayos de luz *siempre* tienen un ángulo de 45° en el diagrama. Una vez que traces el diagrama, apunta en él toda la información que tengas respecto a los intervalos entre los eventos. La información es la misma en todos los marcos de Lorentz. Así, por ejemplo, para cualquier rayo de luz muestra que $I = 0$. Y si hay un reloj en el diagrama y se muestra la hora que marca, recuerda que el reloj mide I a lo largo de su trayectoria.

Con todos estos hechos puedes determinar las coordenadas del evento en el marco seleccionado. Tienes la fórmula general "seudopitagórica" para calcular I en cualquier marco de referencia:

$$I(p, q) = \sqrt{(T(p) - T(q))^2 - (X(p) - X(q))^2 - (Y(p) - Y(q))^2 - (Z(p) - Z(q))^2},$$

que será la base de tus ecuaciones. Los siguientes problemas son de dos dimensiones, de modo que sólo necesitarás una coordenada t y una coordenada x para todos los eventos. La diferencia en las coordenadas t en un marco te da el "tiempo transcurrido entre los eventos" en ese marco, y la diferencia en las coordenadas x te da la "distancia entre los eventos" en ese marco. Con base en el "tiempo transcurrido" y en la "distancia" en un marco, puedes calcular la "velocidad" en ese marco. Pero siempre debes tener en cuenta que el "tiempo transcurrido", la "distancia" y la "velocidad" son cantidades *dependendientes del marco*: varían de un marco a otro. El hecho de calcularlas en un marco no te dice cuál será su valor en otro marco. El intervalo, en cambio, es una cantidad independiente del marco: es la misma en todos los marcos lorentzianos.

Puedes aplicar la "fórmula seudopitagórica" $I = \sqrt{\Delta t^2 - \Delta x^2}$ los "triángulos rectos" en el diagrama para calcular la longitud de la hipotenusa del intervalo, pero sólo en los triángulos en que un lado (Δt) es vertical y el otro (Δx) es horizontal. La clave para la solución de muchos problemas está en trazar líneas adicionales con el fin de formar los triángulos rectos adecuados. Cada triángulo proporciona una ecuación.

La trayectoria de cualquier objeto en movimiento inercial se dará mediante una ecuación lineal en un marco de Lorentz. En el caso de un problema de dos dimensiones en que sólo se utilicen t y x, la ecuación será representada por $x = vt + \mathrm{C}$, donde v es la velocidad del objeto en este marco y C es una constante.

1. Los relojes A y B están sincronizados uno junto al otro. Se separan exactamente a las 12:00 (en cada reloj), moviéndose inercialmente en direcciones opuestas. Al marcar las 12:01, el reloj A emite un rayo de luz. Cuando recibe el rayo de luz, el reloj B marca las 12:02. Utilizando solamente la definición del intervalo y los hechos sobre los relojes y los rayos de luz, determinar lo siguiente:

a) En el marco de referencia del reloj A, ¿cuándo llega el rayo de luz al reloj B?

b) En el marco de referencia del reloj A, ¿qué distancia separa a los relojes cuando el rayo de luz llega al reloj B?

c) En el marco de referencia del reloj A, ¿a qué velocidad se mueve el reloj B?

d) En el marco de referencia del reloj A, ¿qué distancia separaba a los relojes cuando el reloj A emitió el rayo de luz?

e) En el marco de referencia del reloj B, ¿cuándo emite el rayo de luz el reloj A?

f) En el marco de referencia del reloj B, ¿qué distancia separaba a los relojes cuando el rayo de luz llegó al reloj B?

g) En el marco de referencia del reloj B, ¿qué distancia separaba a los relojes cuando el reloj A emitió el rayo de luz?

h) Con base en las respuestas anteriores, calcular el retraso de cada reloj según el juicio del otro.

i) La fórmula usual de la dilatación del tiempo dice que un reloj que se mueve a la velocidad v se retrasa por un factor de $\frac{1}{\sqrt{1 - v^2 / c^2}}$, donde c es la velocidad de la luz. Comparar esto con el resultado del inciso *h*. Hay que recordar que en un marco de Lorentz, tal como nosotros lo hemos definido, $c = 1$.

(Es posible solucionar más fácilmente algunos de estos problemas si recuerdas que la velocidad del reloj A con relación al reloj B es igual a la velocidad negativa del reloj B con relación al reloj A.)

2. Dos gemelos, Samuel y Susana, parten al mismo tiempo de la Tierra. Susana sale en una nave espacial y luego viaja inercialmente hasta que sus relojes muestran que ha pasado un año. En ese momento ella revierte su dirección y vuelve a viajar inercialmente. Al llegar de nuevo a la Tierra, dos años más han pasado (en sus relojes), haciendo un total de tres años. A su regreso, Samuel tiene cuatro años más que cuando ella partió.

a) Según el marco de referencia de Samuel, ¿cuándo revierte su dirección Susana?

b) Según el marco de referencia de Samuel, ¿qué distancia separa a los gemelos cuando Susana revierte su dirección?

c) Según el marco de referencia de Samuel, ¿a qué velocidad parte Susana? ¿A qué velocidad regresa?

d) Según el marco inercial de Susana en el momento que sale, ¿qué distancia separa a los gemelos cuando ella revierte su dirección?

e) Según el marco inercial de Susana en el momento de regresar, ¿qué distancia separa a los gemelos cuando ella revierte su dirección?

3. Con el fin de demostrar que la relatividad especial no requiere coordenadas, considera cómo resolver un problema mediante la utilización de sólo una regla recta minkowskiana y un compás minkowskiano. Es decir, supón que es posible trazar líneas rectas en el espacio-tiempo utilizando la regla recta y que el compás marca lugares de igual intervalo invariante desde un punto dado. Si centras el compás en p y seleccionas cierto intervalo, el compás marca un conjunto de puntos que equivalen a ese intervalo desde p. Describe con detalle cómo utilizarías los dos instrumentos para solucionar el siguiente problema.

Como en el problema 2, Susana sale de viaje y luego regresa mientras que Samuel permanece en la Tierra. Las longitudes de sus líneas de mundo se muestran en la figura A.1 En el momento de dar la vuelta para retornar, Susana emite un rayo de luz. Utilizando tus instrumentos, explica cómo construir la longitud de la línea de mundo que representa el tiempo en la Tierra entre la recepción de la señal de Susana y el regreso de Susana.

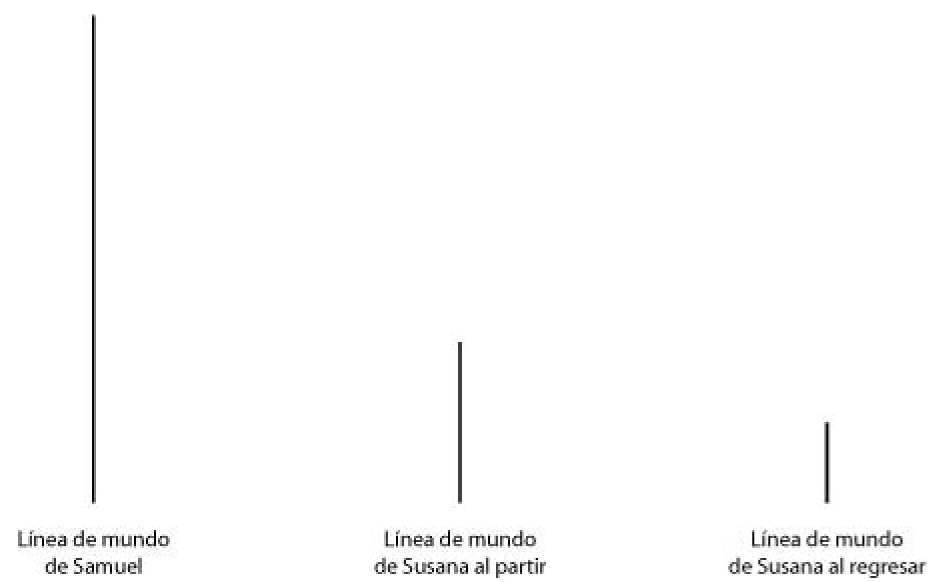

FIGURA A.1

Solución a los incisos a-d del problema 1

Para empezar hay que trazar un diagrama de espacio-tiempo. Puesto que en los incisos a-d se piden cantidades “en el marco de referencia del reloj A”, el reloj A se representará por una línea vertical en el diagrama. El rayo de luz es una línea a 45° y el reloj B es una línea recta angular, puesto que también se encuentra en movimiento inercial. Los registros en el reloj proporcionan los intervalos pertinentes y el intervalo del rayo de luz es 0. Una vez que traces los objetos, tienes que buscar los triángulos rectos. Cuando sea necesario, trazarás líneas adicionales para darle forma a tales triángulos.

El diagrama consiguiente se da en la figura A.2, indicándose la línea adicional con Δx.

Podemos ver dos triángulos rectos en el diagrama. Uno tiene los lados Δx y Δt, siendo la hipotenusa el rayo de luz, y el otro tiene los lados Δx y $1 + \Delta t$, con el reloj B como hipotenusa. El lado vertical del segundo triángulo se puede escribir como $1 + \Delta t$ porque el reloj A mide la coordenada t en este marco, de manera que $I = 1$ en este reloj se traduce directamente como una medida del cambio en la coordenada t a lo largo de esta trayectoria. El reloj B, en cambio, no indica la coordenada t en este marco: tendrá la apariencia de la "dilatación del tiempo". Claro, el reloj B mide la coordenada t en su propio marco de reposo, el cual es pertinente en las partes posteriores del problema.

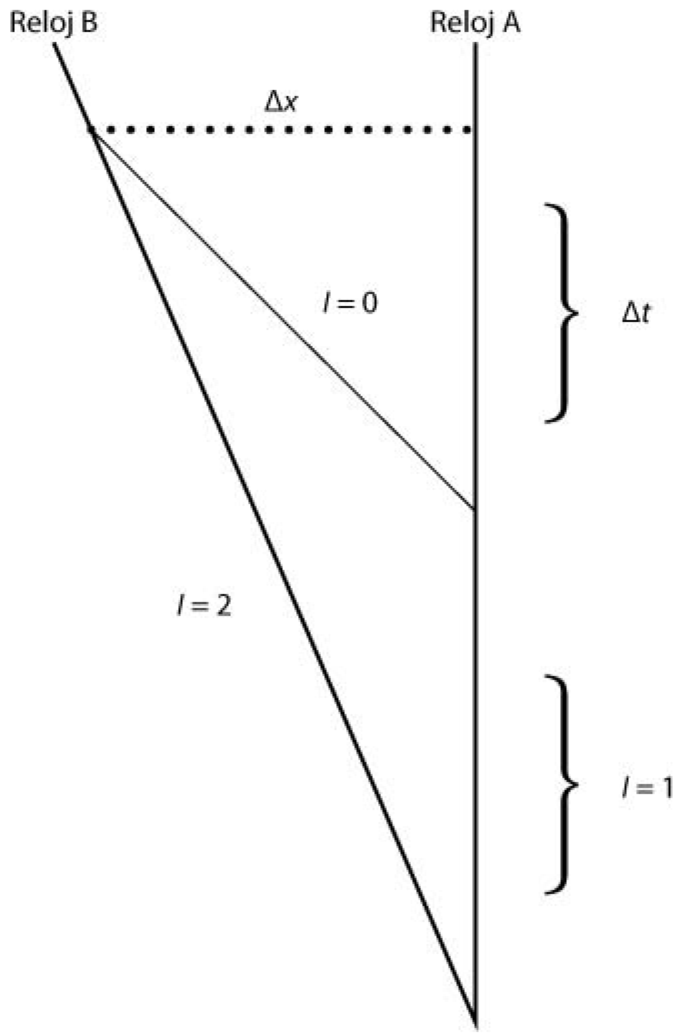

FIGURA A.2

Nuestros dos triángulos rectos producen dos ecuaciones:

$$0 = \sqrt{\Delta t^2 - \Delta x^2}$$
$$2 = \sqrt{(1 + \Delta t)^2 - \Delta x^2}.$$

Elevar al cuadrado ambos lados da

$$0 = \Delta t^2 - \Delta x^2$$
$$4 = (1 + \Delta t)^2 - \Delta x^2.$$

Y la expansión del cuadrado en la segunda ecuación da

$$4 = 1 + 2\Delta t + \Delta t^2 - \Delta x^2$$

Puesto que por la primera ecuación, $\Delta t^2 - \Delta x^2 = 0$, podemos derivar

$$4 = 1 + 2\Delta t$$
$$\Delta t = 3/2.$$

Por lo tanto, la coordenada t, en este marco, del punto donde el rayo de luz llega al reloj B es $1 + 3/2 = 5/2$. Según este marco, el rayo de luz llega al reloj B $2\frac{1}{2}$ minutos después de la separación de los relojes.

Podemos ver ya en este marco el "efecto de la dilatación del tiempo": según el marco, 2½ minutos han transcurrido pero el reloj B sólo registra dos minutos.

La distancia entre los relojes, en este marco, cuando el rayo de luz llega al reloj B es sólo Δx en el diagrama. Pero según nuestra primera ecuación, $\Delta x = \Delta i$. (Esto es válido para cualquier rayo de luz.) De manera que la distancia es $3/2$ los dos relojes tienen una separación de 1½ minutos-luz cuando el rayo de luz llega al reloj B, según el marco del reloj A.

Puesto que según este marco el reloj B ha viajado $1\frac{1}{2}$ minutos-luz en 2½ minutos, su velocidad es $3/2 \,/\, 5/2 = 3/5$ de la velocidad de la luz. Esto indica a qué velocidad el reloj debe moverse en un marco para llegar al nivel de dilatación del tiempo dependiente del marco que hemos calculado.

Una vez que tengamos la velocidad del reloj B en este marco, podemos calcular la distancia entre los relojes cuando el rayo de luz se emitió. Ya que (en este marco) el reloj B se mueve a ⅗ de la velocidad de la luz y que (en este marco) el rayo de luz se emite un minuto después de que los relojes se separan, en este marco se juzga que los relojes tienen una separación de ⅗ minuto-luz en ese momento. Es obvio que todas estas cantidades dependen del marco: la situación es muy diferente en el marco de reposo del reloj B.

BIBLIOGRAFÍA

Alexander, H. G. (comp.), *The Leibniz-Clarke Correspondence*, Barnes & Noble, Nueva York, 1956. [Versión en español: Gottfried Wilhelm Leibniz, *La polémica Leibniz-Clarke*, trad. de Eloy Rada, Taurus, Madrid, 1980.]

Aristóteles, *Física*, Madrid, Gredos, 2008.

Arntzenius, Frank, y Tim Maudlin, "Time Travel and Modern Physics", en Edward N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, Stanford University, Palo Alto, 2013. Disponible en http://plato.stanford.edu/archives/win2013/entries/time-travel-phys/.

Baez, John C., y Emory F. Bunn, "The Meaning of Einstein's Equation", University of California, Riverside, California, 2006. Disponible en http://math.ucr.edu/home/baez/einstein/.

Barbour, Julian, *The End of Time*, Oxford University Press, Nueva York, 2000.

———, *The Discovery of Dynamics*, Oxford University Press, Nueva York, 2001.

———, y Bruno Bertotti, "Mach's Principle and the Structure of Dynamical Theories", *Proceedings of the Royal Society A*, 382 (1783): 295-306, 1982.

Bell, John Stuart, *Speakable and Unspeakable in Quantum Mechanics*, 2ª ed., Cambridge University Press, Cambridge, 2008.

Brown, Harvey, *Physical Relativity*, Oxford University Press, Oxford, 2006.

Butterfield, Jeremy, "The Hole Truth", *British Journal for the Philosophy of Science*, 40 (1): 1-28, 1989.

Carroll, Sean, *Spacetime and Geometry: An Introduction to general Relativity*, Addison-Wesley, Boston, 2004.

Clarke, C. J. S., "Time in General Relativity", en J. Earman, C. Glymour y J. Stachel (comps.), *Minnesota Studies in the Philosophy of Science*, vol. 8, University of Minnesota Press, Minneapolis, 1977.

Descartes, René, *Sobre los principios de la filosofía*, Gredos, Madrid, 1989.

DiSalle, Robert, "Conventionalism and the Origins of the Inertial Frame Concept", en A. Fine, M. Forbes y L. Wessels (comps.), *PAS 1990*, vol. 2, Philosophy of Science Association, East Lansing, 1991.

Earman, John, *A Primer on Determinism*, Reidel, Dordrecht, 1986.

———, *World Enough and Space-Time*, MIT Press, Cambridge, 1989.

———, *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Space-Times*, Oxford University Press, Oxford, 1995.

———, y John Norton, "What Price Space-Time Substantivalism? The Hole Story", *British Journal for the Philosophy of Science*, 38: 515-525, 1987.

———, Christopher Smeenk y Christian Wüthrich, "Do the Laws of Physics Forbid the Operation of a Time Machine?", *Synthese*, 169 (1): 91-124, 2009.

Echeverría, Fernando, Gunnar Klinkhammer y Kip S. Thorne, "Billiard Ball in Wormhole Spacetimes with Closed Timelike Curves: Classical Theory", *Physical Review D*, 44 (4): 1077-1099, 1991.

Einstein, Albert, "Autobiographical Notes", en P. A. Schilpp (comp.), *Albert Einstein: Philosopher-Scientist*, vol. 1, Open Court, LaSalle, 1982. [Versión en español: *Notas autobiográficas*, trad. de Miguel Paredes, Alianza, Madrid, 2003.]

Feynman, Richard, Robert B. Leighton y Matthew Sands, *The Feynman Lectures on Physics*, Addison-Wesley, Reading (Mass.), 1975. [Versión en español: *Lecciones de física*, 3 vols., FCE, México, en prensa.]

Friedman, Michael, *Foundations of Space-Time Theories: Rel-ativistic Physics and Philosophy of Science*, Princeton University Press, Princeton, 1986.

Galilei, Galileo, *Dialogue Concerning the Two Chief World Systems*, trad. de Stillman Drake, University of California Press, Berkeley, 1967. [Versión en español: *Diálogo sobre los dos máximos sistemas del mundo, ptolemaico y copernicano*, trad. de Antonio Beltrán, Alianza, Madrid, 1994.]

———, *Two New Sciences*, trad. de Stillman Drake, University of Wisconsin Press, Madison, 1974. [Versión en español: *Diálogos acerca de dos nuevas ciencias*, Losada, Buenos Aires, 1945.]

Geroch, Robert, *General Relativity from A to B*, University of Chicago Press, Chicago, 1978.

Goldstein, Herbert, *Classical Mechanics*, 2ª ed., Addison-Wesley, Reading (Mass.), 1981. [Versión en español: *Mecánica clásica*, 4ª ed., Reverté, Barcelona y México, 1988.]

Hall, A. Rupert, y Marie Boas Hall (comps.), *Unpublished Papers of Isaac Newton: A Selection from the Portsmouth Collection in the University Library*, Cambridge University Press, Cambridge, 1962.

Hawking, Stephen W., y George F. R. Ellis, *The Large-Scale Structure of Space-Time*, Cambridge University Press, Cambridge, 1973.

Hoefer, Carl, y Nancy Cartwright, "Substantivalism and the Hole Argument", en John Earman, Al Janis, Gerald Massey y Nicholas Recher (comps.), *Philosophical Problems of the Internal and External Worlds*, University of Pittsburgh Press, Pittsburgh, 1994.

Horwich, Paul, *Asymmetries in Time*, MIT Press, Cambridge, 1987.

Huggett, Nick (comp.), *Space from Zeno to Einstein*, MIT Press, Cambridge, 1999.

Kripke, Saul, *Naming and Necessity*, Harvard University Press, Cambridge, 1980. [Versión en español: *El nombrar y la necesidad*, 2ª ed., UNAM, México, 1995.]

Lewis, David, *On the Plurality of Worlds*, Basil Blackwell, Oxford, 1986.

Maudlin, Tim, "The Essence of Space-Time", en Arthur Fine y Jarrett Leplin (comps.), *Proceedings of the Philosophy of Science Association Meetings*, vol. 2, Philosophy of Science Association, East Lansing, 1989.

———, *The Metaphysics Within Physics*, Oxford University Press, Oxford, 2007.

———, "The Geometry of Space-Time", *Aristotelian Society Supplementary*, 84 (1): 63-78, 2010.

Misner, Charles, Kip Thorne y John Wheeler, *Gravitation*, W. H. Freeman, San Francisco, 1983.

Newton, Isaac, *Principia*, trad. de Andrew Motte, rev. de Florian Cajoli, 2 vols., University of California Press, Berkeley, 1934. [Versión en español: *Principios matemáticos de la Filosofía Natural*, 3ª ed., Tecnos, Madrid, 2011.]

Norton, John, "Einstein, Nordström and the Early Demise of Lorentz Covariant, Scalar Theories of Gravitation", *Archive for History of Exact Sciences*, 45: 17-94, 1992.

Norton, John, "The Hole Argument", en Edward N. Zalta (comp.), *The Stanford Encyclopedia of Philosophy*, Stanford University, Palo Alto, 2008. Disponible en http://plato.stanford.edu/archives/win2008/entries/spaceti-me-holearg/.

Price, Huw, *Time's Arrow and Archimedes' Point*, Oxford University Press, Oxford, 1996.

Rindler, Wolfgang, *Essential Relativity*, Springer, Nueva York, 1977.

Rodriguez-Pereyra, Gonzalo, "Leibniz's Argument for the Identity of Indiscernibles in His Correspondence with Clarke", *Australasian Journal of Philosophy*, 77 (4): 429–438, 1999.

Rynasiewicz, Robert, "By Their Properties, Causes and Effects: Newton's Scholium on Time, Space, Place and Motion. Part I: The Text", Studies in History and Philosophy of Science, 26: 133-153, 1995.

———, "By Their Properties, Causes and Effects: Newton's Scholium on Time, Space, Place and Motion. Part II: The Context", *Studies in History and Philosophy of Science*, 26: 295-321, 1995.

Sklar, Lawrence, *Space, Time, and Spacetime*, University of California Press, Berkeley, 1977.

Wald, Robert, *General Relativity*, University of Chicago Press, Chicago, 1984.

ÍNDICE ANALÍTICO

Espacio y tiempo constituyen el escenario en que se desarrolla la historia del universo físico. A pesar de su existencia indudable, su realidad misma escapa a nuestros sentidos: carecen de color, sabor, sonido, aroma o sustancia tangible. Lo que sí parecen tener es una estructura geométrica.

A partir de este indicio, Tim Maudlin recorre la historia de las ideas en torno al espacio-tiempo —desde la física clásica de Aristóteles, Galileo y Newton hasta la teoría de la relatividad de Einstein— para indagar qué es y qué contiene esta estructura.Con destreza, el autor logra guiar al lector a través de las sutilezas matemáticas, físicas y filosóficas de estas teorías, para ofrecerle una explicación clara y concisa del alcance ontológico de las nociones actuales que la física tiene sobre la naturaleza del universo que habitamos.

Tim Maudlin es profesor de filosofía en la Universidad de Nueva York y miembro de la Academia Internacional de Filosofía de las Ciencias. Sus principales líneas de investigación son los fundamentos de la física, la metafísica y la lógica. Es autor de los libros *Quantum Non-Locality and Relativity, The Metaphysics within Physics* y *Truth and Paradox.*

www.fondodeculturaeconomica.com

Índice