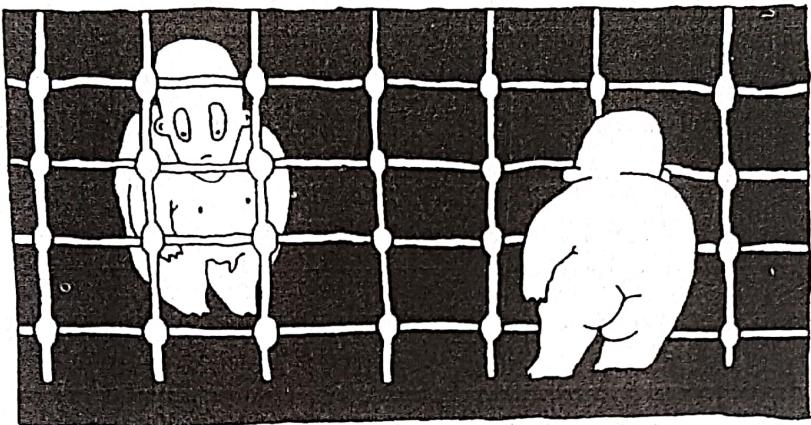


CAPÍTULO

3

El dilema del prisionero



Si usted ve barrotes delante suyo, no quiere decir que sea un prisionero.
Puede que el que está fuera de la celda sea usted.

El dilema del prisionero es el «chiclé» de la teoría de juegos: se lo puede masticar indefinidamente. Miles de matemáticos, psicólogos, científicos sociales, filósofos y economistas han reflexionado sobre esta cuestión, intentando descubrir una solución. Y, a pesar de todo, el dilema sigue siendo tan misterioso y chocante como en 1950, cuando lo propusieron por vez primera Merrill Flood y Melvin Drescher. El nombre de *el dilema del prisionero* se lo puso Albert W. Tucker, quien en 1951 escribió el primer artículo sobre el tema. Tucker presentaba el problema bajo la forma de una historia detectivesca. Aquí tenemos una versión:

La policía atrapa a dos criminales que han cometido un delito grave. No existen pruebas fehacientes de que estos dos hombres hayan cometido ese crimen y lo único que puede demostrar la policía es que conducían con exceso de velocidad. Al fiscal le gustaría mucho cerrar el caso, así que les hace la siguiente propuesta a los dos prisioneros, a los que ha hecho encerrar en celdas distintas:

«Éste es el trato: si usted confiesa su crimen e implica a su cómplice, ayudándonos así a resolver el caso, le pondré en libertad y nos olvidaremos del asunto del exceso de velocidad. En este caso, su cómplice irá a la cárcel diez años y éste será un caso cerrado para siempre. Sin embargo, esta oferta

sólo es válida si su cómplice no confiesa su autoría y, por tanto, no nos ayuda a resolver el asunto. Si él también confiesa, entonces, claro está, la confesión que usted haga no tendrá mucho valor, dado que nos enteraremos de la verdad incluso si usted no habla. En este caso, los dos pasarán cinco años en prisión. Si ninguno de los dos confiesa, lamentablemente no podremos condenar a ninguno de los dos, pero seremos bastante duros con el incidente del exceso de velocidad y ambos pasarán un año en la cárcel. Por último, debo informarle de que le he hecho esta misma oferta a su cómplice. Espero su respuesta mañana a las diez. Y piénselo bien: ¡podría estar libre a las once!»

Resumamos esta situación en una tabla:

		<i>el otro cómplice</i>	
		confiesa	no confiesa
<i>el primer cómplice</i>	confiesa	-5, -5	0, -10
	no confiesa	-10, 0	-1, -1

El primer número de cada celda, el que está en negrita, indica qué es lo que gana el primer cómplice, mientras que el segundo indica lo que gana el segundo. Dado que es peor estar encerrado diez años que cinco, se debe considerar que los años de cárcel son una ganancia negativa y de ahí que se indiquen con signos de menos. En esa situación, el mejor resultado que se pueda obtener es cero, quedar en libertad.

Dos SOLUCIONES LÓGICAS

Asumamos que los dos cómplices no sienten lealtad el uno por el otro y que su colaboración fue totalmente accidental. El único objetivo de cada uno es salir tan bien parado como pueda. ¿Cuál es la acción más lógica, confesar o no confesar?

Imagínese en la piel de esos dos hombres. Si mi cómplice confiesa, las alternativas son: si yo también confieso, estaré en la cárcel cinco años. Pero si no confieso, pasaré diez años encerrado. Así que, si mi cómplice confiesa, sería mejor que yo también lo hiciera. Ahora bien, si mi cómplice no confiesa, vuelve a haber dos alternativas: si confieso, mañana seré libre. Pero si no confieso, pasaré un año en la cárcel. Así, si mi cómplice no confiesa, sería mejor que yo sí lo hiciera.

Sin importar lo que haga mi cómplice, saldré ganando si confieso. No hay otra opción. La lógica me dicta que confiese.

La lógica le dicta lo mismo al otro prisionero. Así, como ambos son seres lógicos, ambos confesarán, recibiendo cada uno una sentencia de cinco años, aunque hubieran pasado un solo año en la cárcel de no haber confesado ninguno de los dos. Éste es el *dilema del prisionero*. La pregunta es si la lógica está equivocada. En otras palabras, ¿acaso la lógica excluye la cooperación racional de ambos prisioneros?

El siguiente razonamiento es tan lógico como el anterior: a mí no es que me caiga muy bien mi cómplice y no tengo relación con él. Sin embargo, por despreciable que sea, sé que es una persona tan inteligente y lógica como yo. Si no lo fuera, no me habría juntado con él. También sé que está en una situación idéntica a la mía. Él tampoco me tiene simpatía y, además, ha recibido la misma oferta que yo. Como yo, basará su decisión en sus propios intereses y en la lógica.

Se nos garantiza que la lógica siempre lleva a la misma conclusión. Dos y dos son cuatro, independientemente del método con que se realice el cálculo correcto. Así que mi cómplice tomará la misma decisión que yo, sin importar cuál sea ésta. Por consiguiente, si decido confesar, estoy seguro de que él llegará a la misma conclusión. Si mi forma de razonar no me lleva a confesar, su forma de pensar le llevará al mismo resultado.

Si decido confesar (y él también lo hará), estaré condenado a cinco años de cárcel (y él también, aunque eso me da igual, ¡mala suerte!). Si decido no confesar (y él también) pasaré un año en prisión. Es mejor pasar un año en la cárcel que cinco, de modo que no confesaré.

Esta forma de razonar parece tan lógica como la anterior. Pero ¿cómo es que dos argumentos igualmente lógicos conducen a resultados distintos? ¿Es que uno de ellos es falso? ¿O hay algún problema con la lógica?

SOBRE LA NATURALEZA DE LA LÓGICA

Una cuidadosa comparación entre los dos razonamientos demostrará que el segundo incluye un paso que está ausente en el primero. Asume que dos razonamientos complementarios conducen necesariamente a la misma conclusión. Éste es el único paso que podría causar la contradicción.

Aquí tenemos la solución. Lo que realmente se desprende de los dos razonamientos, *considerados juntos*, es que no pueden existir situaciones como la del dilema del prisionero. Para ser más exactos, esto se desprende

de la afirmación inicial del segundo argumento. El primer argumento se limitaba a aplicar unos pasos que también se permiten en el segundo. Por consiguiente, podemos continuar el segundo argumento simplemente añadiéndole el primero. De este modo llegamos a una contradicción. Así pues, según las reglas de la lógica, debe ser imposible la hipótesis de que exista un dilema del prisionero.

Si los dilemas del prisionero no existen, podremos extraer de ellos cualquier cosa (si imaginamos que existen), dado que una de las reglas de la lógica es que de una afirmación falsa se puede extraer cualquier conclusión. Si las brujas no existen, entonces las afirmaciones «Todas las brujas vuelan en escobas» y «Ninguna bruja vuela en escoba» son, desde el punto de vista lógico, igualmente ciertas. Puede que esto suene raro, pero es mera y absurdamente cierto afirmar que todas las brujas que no existen vuelan en escoba. Por otra parte, si no construimos la lógica de este modo, aceptando afirmaciones tan absurdas como si fueran ciertas, la propia lógica será incoherente y, por tanto, inútil. Así es como funciona la lógica.

Si no existen los dilemas del prisionero, entonces, en una situación propia de tal dilema, es lógico tanto confesar como no confesar. Nuestro primer argumento llegó a la primera conclusión y el segundo, a la otra posible.

Como demostró el teorema de Gödel (1931), no existe un sistema de determinada complejidad en el que todas las verdades que se afirman en el mismo se puedan demostrar dentro de sus límites. El primer razonamiento sólo utilizaba las reglas normales de la lógica. Esto no quiere decir que la condición adicional del segundo argumento (que dos argumentos lógicos deben conducir a la misma solución) sea falsa. Tampoco excluye su falsedad. Dentro de este sistema no se puede deducir ninguna de las dos cosas.

La lógica, bajo su forma usual, no incluye esta condición adicional. Así, este sistema en realidad excluye la cooperación de los dos prisioneros dentro del dilema del prisionero; pero no excluye la existencia de un mundo en el que no existan dilemas de este tipo. Por el momento, ésta es sólo una posibilidad abstracta, pero en el capítulo 4 veremos ejemplos concretos de los *mecanismos* con los que se podría obtener un mundo exento de dilemas del prisionero. En semejante mundo, el fiscal no podría colocar a ambos reos en la misma situación. Incluso si pudiera, los dos reos no podrían percibir esa situación como un dilema del prisionero. Consideremos un ejemplo.

Una vez, Merrill Flood, uno de los descubridores del dilema del prisionero, le hizo la siguiente proposición a una secretaria de su instituto: que le

daría 100 dólares o, alternativamente, 150, si ella y su colega se ponían de acuerdo sobre cómo dividir esta segunda cantidad entre ellas. Flood no le ofreció dinero a la otra secretaria. Lo que le interesaba era saber de qué manera y basándose en qué dividirían las secretarias los 50 dólares extra. Para gran sorpresa suya, las dos enseguida optaron por recibir 75 dólares cada una, aunque la primera podía haber obtenido 100 si no le hubiera mencionado la oferta a su compañera. Sin embargo, las secretarias percibieron la situación como un juego cooperativo, eludiendo por completo competir por el dinero extra.

Para los que piensan de este modo, en realidad los dilemas del prisionero no existen, incluso si hay un fiscal que intente crearlos. Aunque la lógica prohíbe que los prisioneros cooperasen si existieran los dilemas, no excluye un mundo en el que no *pueda haber* dilemas del prisionero. Quizá resultaría agradable vivir en un mundo así. Sin embargo, sabemos por experiencia que los dilemas del prisionero sí existen en nuestro mundo.

LOS DILEMAS DEL PRISIONERO EN LA VIDA COTIDIANA

Hay dos gasolineras situadas una al lado de otra. Al principio de cada mes, cada propietario debe fijar el precio de la gasolina para el mes que comienza porque la ley de ese estado prohíbe el cambio de precio durante ese período de tiempo. El nuevo precio debe fijarse exactamente a las doce de la noche del primer día del mes.

El propietario de una de las gasolineras puede pensar esto: «El mes pasado obtuve ciertos beneficios, pero no muy grandes. Si la otra gasolinera se quedara sin clientes, mis ventas se duplicarían y entonces sí que obtendría grandes ingresos, dado que los costes de mantenimiento de mi gasolinera no aumentarían mucho. Valdría la pena sacrificarse un poco para quitarle clientes a la competencia. ¿Qué tal si bajo un poco el precio? Aunque obtendría un poco menos de beneficio por cada litro de gasolina, mis ventas casi se duplicarían y eso bien valdría la pena».

Tras una serie de complicados cálculos, el propietario de la gasolinera llega a la conclusión de que, si rebajara un poco los precios y, por consiguiente, captara la mitad de los clientes de la competencia, entonces su unidad de beneficio se multiplicaría por cuatro. Pero entonces le asalta la duda: «¿Y qué pasa si el otro propietario también ha pensado en esto y rebaja el precio? ¡Entonces mis ventas no subirán nada!». Preocupado, hace nuevos cálculos y decide que el mes próximo, si rebajara los precios, no ob-

¿qué pasaría si mantuviera su precio habitual, más elevado, y el otro propietario redujera el suyo? ¡El resultado sería catastrófico! Los costes de mantenimiento son tan elevados que, con sólo la mitad del volumen de ventas, su déficit sería de tres unidades incluso aunque mantuviera los precios altos.

Se acerca la medianoche y, si quiere fijar un nuevo precio, debe hacerlo cuando el minutero llegue a las doce. Por si acaso, ha preparado una pizarra con un precio más bajo. Si ve que su vecino baja los precios, él podrá hacer lo mismo rápidamente. Lentamente, se acerca a su pizarra mientras suenan las primeras campanadas y ve que el otro propietario también se acerca a la suya, con aspecto preocupado, aferrando un pizarrín. Justo cuando están a punto de entablar conversación, ven que se acerca el temido Inspector de Ordenanzas Municipales, que quiere comprobar qué pasará con los precios a medianoche. No hay tiempo para negociaciones, que, dicho sea de paso, son ilegales. Ambos propietarios deben decidir inmediatamente si cambiarán o no el precio. Se acerca el momento crítico, la última campanada, y ninguno de los dos puede ver qué ha hecho el otro.

Una vez más, podemos resumir esta situación en una tabla:

	<i>el segundo propietario</i>	
	baja el precio	no baja el precio
<i>el primer propietario</i>	baja el precio	0, 0
	no baja el precio	4, -3
	-3, 4	1, 1

La lógica es exactamente la misma del dilema del prisionero: sin importar lo que haga el segundo propietario, lo mejor que puede hacer el primero es bajar el precio. Si el segundo también lo baja, el primero podrá evitar una pérdida. Si el segundo no baja el precio, el primero podrá cuadruplicar sus beneficios. Así, esa medianoche, la avaricia y el temor a las pérdidas exigen una disminución del precio, pero si los dos hombres lo rebajan, ambos perderán toda su ventaja.

La compra y venta habituales pueden conducir también a un dilema del prisionero, sobre todo en el mercado negro, donde no existen garantías de que volvamos a encontrarnos con el vendedor al día siguiente. No disponemos de mucho tiempo para examinar la calidad: yo puedo pagar con dinero falso y el