Desde la paradoja de Russell y el problema de la parada hasta los teoremas de Gödel, Cantor o Arrow, hay un patrón común: cuando un sistema es lo suficientemente expresivo como para referirse a sí mismo y negarse a la vez, surgen límites inevitables para la lógica, la computación o la formalización del conocimiento. A partir de ahí, quizá no todo lo que sabemos o experimentamos puede capturarse en estructuras lógicas. En ese territorio fronterizo —donde la razón encuentra sus límites—, el arte y la poesía aparecen como formas alternativas de acercarse a lo inalcanzable.

[ILUSTRACIÓN: MARESSA-ANDRIOLI/ ISTOCK]

A veces, basta una frase para poner en jaque a la lógica: “Esta frase es falsa”. Si es verdadera, es falsa; si es falsa, es verdadera. Ese nudo conceptual –la combinación de autorreferencia y negación– ha acompañado al pensamiento occidental durante siglos y reaparece, una y otra vez, en algunos de los descubrimientos más profundos de la ciencia y la filosofía.

Basta una frase para poner en jaque a la lógica: “Esta frase es falsa”. Si es verdadera, es falsa; si es falsa, es verdadera

Por ejemplo, si digo “Soy una mentirosa”, podemos concluir que estoy mintiendo si y solo si no estoy mintiendo. No podemos asignar un valor de verdad coherente a mi oración –mi oración trasciende el conjunto de frases verdaderas o falsas–. Si intentamos solucionarlo añadiendo un tercer valor de verdad, como “semicierta” o “indefinida”, podré construir una nueva paradoja que escapará de la nueva clasificación. Y así sucesivamente ad infinitum.

¿Qué me permite construir esta oración? Un lenguaje donde pueda formular la referencia a mí misma, así como la negación. “Soy honesta” o “Esta frase es verdadera” contienen autorreferencia sin negación, y no resultan problemáticas. “Esto no es una manzana” contiene negación sin autorreferencia, por eso, no plantea retos lógicos. La autorreferencia unida a la negación conforman un ataque desconcertante y poderoso, que nos deja sin lugar donde refugiarnos de inconsistencias. Como un uróboro que muerde su propia cola venenosa, y pone en cuestión su propia existencia.

El alcance de la paradoja de autorreferencia y negación es espectacular. Primero, por la cantidad de escenarios donde se puede formular: se puede reencarnar, esencialmente, en toda actividad intelectual que se refleje en la lógica, si la última es lo bastante expresiva. Segundo, por su persistencia: no hay forma sencilla o plenamente satisfactoria de resolver la paradoja. Su honda huella se encuentra en el corazón de algunos de los resultados más profundos en matemáticas, informática, semántica, teoría de la elección social e, incluso, la filosofía.

Por ejemplo, en teoría de conjuntos, Bertrand Russell nos urge a considerar la colección de todos los conjuntos que no se contienen a sí mismos. Resulta que esta colección se contiene a sí misma si y solo si no lo hace. Inadmisible. Resulta que no para toda propiedad existe el conjunto de elementos con aquella propiedad. Esta es la paradoja de Russell.

Las muchas versiones de la paradoja de la mentirosa

En la teoría de la computación, no puede existir una máquina de Turing que resuelva el problema de la parada. Es decir, no existe un programa maestro que pueda analizar el código de otro programa y predecir correctamente si se detendrá o entrará en un bucle infinito. Si existiera, le daríamos su propia descripción, modificaríamos el output y resultaría que el algoritmo pararía sí y solo si no parara.

Por su parte, el primer teorema de incompletitud de Gödel considera un sistema formal que construye la frase "yo no soy demostrable". Tras enredarnos en la contradicción de si la frase es demostrable o no, escapamos concluyendo que es indecidible. De este modo, los sistemas formales no pueden ser consistentes y completos a la vez, es decir, demostrar todas y solo las frases ciertas, si son lo bastante expresivos para formular la problemática frase. Esto limita el poder de la deducción.

La paradoja también limita el poder de la inducción mediante el argumento diagonal de Putnam, que excluye la existencia de una máquina inductiva universal. Como se imaginan, el argumento asume que tal máquina existe, la describe, la niega y llega a una contradicción. Esto implica que es imposible automatizar completamente el descubrimiento y debe haber espacio para la libertad.

De igual forma, en las teorías de la verdad no puede existir un predicado que dictamine si cada frase es verdad, pues le daríamos su propia descripción y la negaríamos, como afirma el teorema de Tarski.

El problema de los infinitos, el voto y los viajes en el tiempo

Esta paradoja también tiene consecuencias descomunales para los infinitos. El teorema de Cantor dice que no hay ningún conjunto que pueda abarcar el conjunto de sus partes, llamado "conjunto potencia". Para demostrarlo, asumimos que, por algún conjunto, existe una función cuya imagen es todo el conjunto potencia. Apareamos el conjunto original con el conjunto potencia en una tabla, consideramos su diagonal (autoreferencia), la modificamos (negación) y encontramos un elemento del conjunto potencia que no está en la imagen de la función –contradicción–. Así, el infinito de los números naturales es inferior al de los reales, que a su vez es inferior al siguiente… hasta concluir que ¡hay infinitos tipos de infinito!

Otro ejemplo es la teoría de la elección social, donde deseamos agregar las voluntades de votantes de forma que se respeten sus preferencias, formulado en tres condiciones muy razonables. La primera se llama unanimidad: si todos los votantes prefieren una opción respecto a otra, entonces el agregado de voluntades también. La segunda condición se llama inependencia de alternativas irrelevantes, y dice que el agregado de voluntades de las alternativas a y b solo depende de las voluntades de los votantes sobre a y b. La tercera condición se llama no-dictadura, y pide que no haya ningún votante cuyo voluntad se respete siempre en el agregado independientemente de la voluntad de los otros votantes. Pero el teorema de imposibilidad de Arrow dice que no se pueden satisfacer, ya que no cumple sus tres condiciones.

Incluso los viajes en el tiempo también pueden verse como una forma de autoreferencia y negación en el tiempo. Si pudiera viajar al pasado y matarme a mí misma antes de empezar a escribir este artículo, ¿lo habría escrito o no? No se puede resolver (lo estoy escribiendo, pero me lo estoy planteando…).

El teorema del punto fijo

Una de las cosas más emocionantes que he aprendido en los últimos años es que todos estos resultados son ejemplos de un teorema relativamente obscuro, formulado en teoría de categorías en los años 1960, llamado el teorema del punto fijo de Lawvere. Se demuestra mediante un diagrama que captura de forma transparente el problema de la autorreferencia y la negación. En esencia, el teorema dice que, si tienes un conjunto de cosas que quieres representar mediante un elemento universal, entonces debe haber alguna limitación para asegurar la consistencia. Porque la cosa puede representarse directamente o a través del elemento universal,

y ambas representaciones deben ser iguales. La limitación es que toda función debe tener un punto fijo, es decir, un valor que la función envía a sí mismo. La función más simple sin punto fijo es la negación, que transforma 0 en 1, y 1 en 0 –es decir, la operación que transforma verdadero en falso y falso en verdadero. Por tanto no hay ningún valor que quede igual, es decir, no tiene punto fijo–. Lo que es problemático, pues, es la autorreferencia seguida de una función sin punto fijo.

El lugar de la filosofía

Los tentáculos de la paradoja también se extienden en la filosofía. Uno de mis libros favoritos, *Beyond the limits of thought*, de Graham Priest, identifica el problema de intentar alcanzar más allá de lo alcanzable, sea en el pensamiento, en la cognición... Tantos pensadores han caído en la trampa lógica de hablar de lo que no se puede hablar, concebir lo que no se puede concebir. Un caso famoso es Parménides, quien nos dice que el no ser no se puede pensar. Sin embargo, aquí estamos, pensándolo. Kant sostiene que la cosa-en-sí-misma no admite ningún atributo, pero esto ya es decir algo de esa cosa, le estamos asignando un atributo de forma inadvertida. Y muchos más. Arquitas, un pitagórico, supone que el universo tiene un final, que imaginó como una gran membrana. Entonces esa membrana debería estar tocando algo, pues está separando el Universo del resto. Pero, por construcción, no puede existir nada más allá del Universo, no hay tal "resto". La clave es que imaginar tal membrana requiere concebir el lado que no es, cosa que está prohibida. En sí mismo, concebir el fin del Universo ya es problemático.

Tantos pensadores han caído en la trampa lógica de hablar de lo que no se puede hablar, concebir lo que no se puede concebir

En palabras de Wittgenstein, para poder trazar un límite del pensamiento, ambos lados del límite deben ser pensables. En palabras de Hegel, determinar algo como una limitación implica que esa limitación ya ha sido trascendida. En la misma línea, la exhaustiva exploración de Priest ilumina las muchas reencarnaciones de este círculo vicioso.

¿Cómo debemos responder a estas paradojas?

La respuesta más habitual es negar la existencia de un objeto que reúne todas las propiedades deseadas. Para la computabilidad, negamos que exista una máquina de Turing universal que calcule funciones totales, es decir, que su respuesta esté definida para todos los inputs. Para sistemas formales, negamos que sean completos y consistentes a la vez. Para el teorema de imposibilidad de Arrow, negamos que exista una función de bienestar social que satisfaga las tres condiciones deseadas, etcétera.

Esto resulta en sistemas más pobres de lo deseado, encima de los cuales se suele construir una jerarquía que cuidadosamente aumenta en expresividad sin permitir la autorreferencia. Una respuesta menos popular consiste en abrazar las paradojas y aceptar algunas contradicciones como verdaderas.

Pero la opción que más me intriga y más me cuesta articular es la posibilidad de que haya formas de acercarse al conocimiento o a la experiencia imposibles de reflejar en una estructura lógica. Quizá no todo pueda formalizarse. Ciertos poemas me evocan sensaciones e intuiciones que siento como muy ciertas y, sin embargo, no sé reflejar en la lógica. Y, si lo pudiera hacer, dudo que fuesen coherentes.

El arte como camino



Para mí, aquí es donde entra el arte. A menudo, tengo la impresión de que el arte se acerca a los sitios prohibidos por la lógica sin miedo, con coraje, como si se lanzara de cabeza a una piscina sin agua. Como en la película *8 1/2* de Fellini, donde Mastroianni entra y sale de la película como director de la misma. En el arte visual, también encontramos ejemplos de estas paradojas. Consideren el cuadro de Magritte que representa una pipa y se lee “Esto no es una pipa”, o las manos que se dibujan a sí mismas de Escher. De hecho, Hofstadter ya explicó en *Gödel, Escher, Bach* que la estructura del *strange loop* del teorema de Gödel es análoga a la que aparece en algunos cuadros de Escher y Magritte. En esencia, representan superficies no orientables, como una cinta de Möbius o una botella de Klein.

El arte se acerca a los sitios prohibidos por la lógica sin miedo, con coraje, como si se lanzara de cabeza a una piscina sin agua

A mi entender, Heidegger sugiere algo similar. Tras enredarse en las contradicciones de definir el ser (pues no se puede decir nada sobre el ser, ya que para decir algo sobre él habría que decir algo de la forma “el ser es…” y así tratarlo como un objeto), subraya que el arte y, en particular, la poesía, cuando se entienden del modo adecuado, pueden revelar qué es el ser, llevándonos más allá de la lógica.

Pero hacer precisa esta idea parece requerir definir el arte, una tarea difícil, si es que resulta posible. Y, aún más, ¡la propia idea parece paradójica! Pues definir el arte como aquello no formalizable hace que la definición sea víctima de su propio ataque de autorreferencia y negación. Quizá debo conformarme con la intuición y no intentar formalizarla, si no quiero quedar atrapada en las arenas movedizas de la paradoja.