

LA LÓGICA Y EL DILEMA DE BENACERRAF

RAÚL ORAYEN

Instituto de Investigaciones Filosóficas
UNAM

En "Mathematical Truth", Paul Benacerraf les plantea a los filósofos de la matemática un dilema muy difícil de resolver, que puede presentarse de la manera siguiente: o bien postulamos que el lenguaje matemático se refiere a objetos abstractos (como clases, números, etc.) o nos negamos a tal postulación; en el primer caso, se dificulta enormemente explicar el conocimiento matemático, y en el segundo perdemos la explicación semántica estándar de la noción de verdad matemática. A muchos nos resultan convincentes, *prima facie*, los argumentos del autor acerca de lo que sucede en cada caso. No está claro en el artículo cuál de las dos alternativas problemáticas abandonaría más fácilmente el autor; pero un trabajo suyo anterior, "What Numbers Could not Be" sugiere que Benacerraf estaría más tentado a elegir el segundo cuerno del dilema (es lo que hace en el último artículo citado respecto de la aritmética, pues según él "there are no such things as numbers"; véase p. 294). Como Benacerraf mismo muestra en "Mathematical Truth", esta posición conduce a la dificultad de que exige abandonar la explicación semántica estándar del lenguaje matemático. Está claro, y así se reconoce, el hecho de que el segundo cuerno del dilema tiene esa consecuencia. Se ha reparado menos, en cambio, en las revisiones que hay que hacer en nuestras ideas sobre la lógica, en caso de adoptar la posición teórica mencio-

nada. En el presente trabajo deseo hacer notar tres consecuencias acerca de nuestras concepciones lógicas que se siguen del abandono de la postulación de entidades matemáticas abstractas.¹

1. *Sobre la aplicabilidad de la lógica*

La teoría lógica actual más difundida y utilizada es la teoría cuantificacional de orden uno con identidad a la que denotaré en lo que sigue con la abreviatura " TC^1 ". La aplicabilidad de TC^1 al lenguaje ordinario y a las teorías físicas es muy controvertida. Los condicionales del lenguaje ordinario no se comportan como el condicional material, las oraciones coloquiales presentan lagunas veritativas que desaparecen en sus traducciones a fórmulas de TC^1 , muchas inferencias cotidianas no pueden fundamentarse con tal lógica porque envuelven giros no extensionales, etc., etc. En menor medida, muchas de estas dificultades aparecen también respecto del lenguaje de las ciencias físicas. ¿Qué aplicaciones exitosas pueden citarse en favor de TC^1 ? A pesar de todas las dificultades en otros campos, hay una aplicación *tan* exitosa que bastaría para justificar la existencia de la lógica estándar: su uso en la reconstrucción de la inferencia matemática correcta. En este terreno, la adecuación y utilidad de TC^1 es notable.² Pero si relacionamos este hecho con el dilema de Benacerraf surgen algunas sorpresas.

Consideremos la aplicación de la lógica a la aritmética elemental. Para efectuar tal aplicación debemos traducir los

¹ Habitualmente, incluyo en mis trabajos partes expositivas que permiten ubicar el contexto teórico en el que pretendo hacer algún aporte; no lo hago en el presente ensayo porque otros artículos publicados en este mismo número de *Crítica* suministran con holgura el conocimiento de la ideas de Benacerraf que presupongo aquí (véanse especialmente las contribuciones de Consuegra y Hart).

² Sobre esto, véase mi libro *Lógica, significado y ontología*, cap. IV, sec. 8.

enunciados aritméticos a fórmulas de TC^1 . Pero al hacerse la traducción se presupone que el lenguaje aritmético es referencial en un sentido que es negado por la posición que estamos analizando (que es la defendida por Benacerraf en "What Numbers Could not Be"). Si formalizamos

(a) 2 es par

mediante

(b) Fa

presuponemos que "2" es una constante individual y que denota un objeto. Similarmente, si representamos enunciados generales de la aritmética mediante fórmulas cuantificadas presuponemos que en la aritmética se habla de un dominio de objetos que se adopta como el dominio de las variables utilizadas en las fórmulas. Pero el cuerno elegido del dilema de Benacerraf niega exactamente las presuposiciones que hemos hecho en estos casos.

¿Por qué razón al aplicar TC^1 a la aritmética debemos presuponer que el lenguaje de esta última es referencial? La respuesta está vinculada al tipo de aplicaciones para las cuales está probada la eficacia de TC^1 . Veamos este tema someramente.³

Entre las fórmulas de TC^1 está definida una relación sintáctica de consecuencia: si A y B son fórmulas de la teoría, B es una consecuencia de A si y sólo si se puede llegar de A a B aplicando sólo axiomas y reglas de inferencia de TC^1 que se han descrito de modo puramente formal. ¿Qué utilidad puede tener esta relación en el estudio de enunciados de la aritmética o algún otro lenguaje? Tal utilidad está mediada por una semántica que se introduce para TC^1 . En esa semántica

³ Mis explicaciones no podrán ser seguidas cabalmente sin ninguna noción previa de la semántica de la teoría cuantificacional. Sobre estas cuestiones puede consultarse Mendelson, *Introduction to Mathematical Logic*, cap. I, sec. 2.

se definen *interpretaciones*, cada una de las cuales consiste, esencialmente, en una asignación de entidades a partes de las fórmulas de TC^1 (se conectan las variables con un conjunto de objetos, cada constante individual con uno de esos objetos, etc.). También se introduce una noción de verdad relativa a una interpretación. Cuando se especifica en qué consiste que una fórmula cerrada sea *verdadera en una interpretación dada*, lo que se hace es dar las condiciones de verdad de esa fórmula en términos de las entidades conectadas con sus partes en esa interpretación. Como ha señalado Hartry Field, esto equivale a decir: “how the truth value of a whole sentence depends on the denotations of its primitive nonlogical parts”.⁴ Regresemos a la noción sintáctica de consecuencia mencionada al comienzo de este párrafo. En la metateoría se prueba la existencia de cierto paralelismo sintáctico-semántico: cuando una fórmula cerrada B es una consecuencia de cierta fórmula cerrada A , no hay ninguna interpretación en que A sea verdadera y B no lo sea (sea falsa). Se puede conectar lo que sucede entre A y B dentro de TC^1 con las relaciones entre dos enunciados matemáticos (o de otro lenguaje cualquiera distinto de TC^1) p y q , si advertimos que hay cierta interpretación de las fórmulas de TC^1 (o de los símbolos que aparecen específicamente en A y B) tal que en esa interpretación A tiene las mismas condiciones de verdad que p y B las mismas de q .⁵ Ahora bien, si B es una consecuencia de A , se sigue (por un paralelismo mencionado antes) que si A es verdadera en una interpretación cualquiera, B también lo será en esa interpretación. Si elegimos aquella interpretación en que A tenía las mismas condiciones de verdad que p y B las mismas que las de q , habremos probado algo interesante acerca de p y q : si p es verdadero q también lo será. En

⁴ Hartry Field, “Tarski’s Theory of Truth”, p. 86.

⁵ O las condiciones de verdad de A en la interpretación y de p en su contexto son lo suficientemente análogas como para que no haya duda razonable, dados ciertos supuestos teóricos, de que A y p tienen el mismo valor de verdad (similarmente respecto de B y q).

otras palabras conociendo las relaciones entre los valores de verdad de A y B (en todas las interpretaciones, si la primera es verdadera, lo es también la segunda) y sabiendo que en cierta interpretación tienen las mismas condiciones de verdad que p y q , respectivamente, habremos inferido una relación entre los valores de verdad de p y q . La inferencia dependió del dato de mismidad de condiciones de verdad entre A , en una interpretación, y p en su significado habitual (o el pertinente en el contexto), y similarmente respecto de B y q . Pero las condiciones de verdad de A y B en una interpretación se formulan en términos de entidades conectadas con sus partes; por esa razón, se puede establecer identidad de esas condiciones con las de p y q , respectivamente, solamente si también podemos formular las condiciones de verdad de p y q en términos de entidades conectadas con sus partes (usualmente consideradas como entidades a las que algunas partes simbólicas de p y q hacen referencia).

El párrafo anterior trata de dar una idea de por qué TC^1 sólo puede aplicarse a lenguajes referenciales, en los que se hace referencia a objetos o entidades, que pueden ser el punto de conexión con interpretaciones de las fórmulas lógicas. El párrafo muestra también en qué forma la “eficacia” de TC^1 está demostrada para lenguajes que sean traducibles a las fórmulas de TC^1 en las interpretaciones definidas para ellas. La metateoría muestra que en todas esas interpretaciones se cumplen ciertas propiedades o relaciones de las fórmulas (por ejemplo, que si B es una consecuencia lógica de A en *toda* interpretación se cumple que si A es verdadera en ella, B también lo será). Si logramos encontrar enunciados que tengan las mismas condiciones de verdad que ciertas fórmulas en alguna interpretación, podremos inferir propiedades y relaciones entre tales enunciados de la manera ilustrada más arriba.

Volviendo al problema de Benacerraf: si no postulamos objetos aritméticos (como en “What Numbers Could not Be”, donde Benacerraf niega la existencia de los números), no podemos in-

interpretar de la misma manera referencial corriente los enunciados de la aritmética ordinaria. Podemos, de todos modos, seguir simbolizándolos de la manera lógica usual.⁶ Pero, si no desarrollamos una semántica existosa diferente de las usuales, no tendremos *ninguna* explicación de por qué las técnicas lógicas pueden reconstruir adecuadamente las inferencias aceptadas por los matemáticos, que conducen a teoremas cuya utilidad se hace sentir en la física. Mientras no se conozca una semántica alternativa tal, podemos preguntarnos: ¿es por pura casualidad que la teoría lógica diseñada para lenguajes referenciales de cierto tipo se aplica tan bien en las teorías matemáticas usuales? Adoptar el segundo cuerno del dilema de Benacerraf obliga, pues, a revisar nuestra explicación de la eficacia de la lógica en ciertos dominios y plantea algunos enigmas de interés.

2. Sobre la completitud de la lógica

Ciertos resultados metalógicos sobre infinitud y completitud son bien conocidos por los lógicos matemáticos pero han pasado más desapercibidos para el público filosófico general. Los traeré aquí a colación porque, como veremos, arrojan conclusiones un poco sorprendentes si se combinan con el segundo cuerno del dilema de Benacerraf y algunas informaciones adicionales.

Se pueden citar resultados más fuertes, pero bastará para nuestros propósitos mencionar que si *no hay* conjuntos infinitos, TC^1 es incompleta. La razón es que hay ciertas fórmulas que pueden verificarse sólo si algunas de sus letras predicativas se interpretan mediante conjuntos infinitos; si no los hay, tales fórmulas serán insatisfactibles y sus negaciones serán verdaderas para cualquier interpretación. Como no se trata de teo-

⁶ Podemos seguir las costumbres habituales de simbolización de manera mecánica sin hacer los supuestos ontológicos correspondientes: colocaremos constantes individuales en lugar de numerales, etc.

remas (en cuyo caso la lógica estaría reñida con la teoría de conjuntos actual), tales negaciones harían incompleta la lógica, bajo el supuesto asumido. Daré a continuación un ejemplo.⁷

Mediante una fórmula de TC^1 puedo expresar la idea de que la extensión de un predicado F es una parte propia de la extensión de un predicado G :

$$(1) (x)(Fx \supset Gx) \& (\exists x)(Gx \& \sim Fx).$$

También puedo afirmar que una relación R tiene la extensión de F como dominio y la extensión de G como rango, o codominio:

$$(2) (x)[(\exists y)xRy \leftrightarrow Fx] \& (y)[(\exists x)xRy \leftrightarrow Gy].$$

Finalmente, puedo indicar que R es biunívoca:

$$(3) (x)(y)(z)[(xRy \& xRz) \supset y = z] \& (x)(y)(z)[(xRy \cdot zRy) \supset x = z]$$

La conjunción de (1), (2), (3), sólo puede ser verdadera en una interpretación en la cual un conjunto (el asignado a G) puede ponerse en correspondencia biunívoca con una parte propia (la extensión de F). Pero se puede demostrar que un conjunto finito no puede ponerse en correspondencia biunívoca con una parte propia.⁸ Luego, si no hay conjuntos infinitos la conjunción de (1), (2) y (3), será falsa en toda interpretación. Su negación será verdadera en toda interpretación. Será, pues, una verdad lógica. Pero no es teorema de TC^1 (de lo contrario, TC^1 contradiría la teoría de conjuntos matemática y no podría ser utilizada como lógica subyacente de la misma). Luego, si no hay conjuntos infinitos, TC^1 es incompleta.

¿Qué relación tiene esto con el dilema de Benacerraf? El segundo cuerno de su dilema constituía el rechazo de las en-

⁷ Mosterín da ejemplos interesantes en "La frontera entre lógica y teoría de conjuntos"; el mfo es más sencillo.

⁸ Si se acepta el axioma de elección también puede demostrarse que *todo* conjunto infinito puede ponerse en correspondencia biunívoca con una parte propia.

tidades abstractas. Una posición tal suele estar asociada con un vago naturalismo que suele considerar de gran relevancia para la filosofía los datos que suministran las ciencias naturales más avanzadas (obviamente aun cuando no se adopte una actitud anti-entidades-abstractas podemos asumir una actitud de respeto e interés por los datos y mejores hipótesis de tales disciplinas). Ahora bien, nada en el conocimiento natural reciente sugiere la infinitud de las entidades materiales. La física parece decididamente inclinada a hipótesis finitistas sobre el universo. Si, como muchos filósofos, usamos en nuestras especulaciones las teorías que los científicos consideran más confiables por el momento, tendremos que inclinarnos por suponer que no hay infinitas entidades materiales.⁹ Si también rechazamos las entidades abstractas como en el cuerno del dilema que estamos analizando, parece que la hipótesis que debemos considerar más probable nos conduce a que las extensiones de los predicados no pueden ser infinitas. En ese caso, el rechazo de las entidades abstractas, combinado con ciertos datos e hipótesis científicas, conduce a la consecuencia algo sorprendente de que TC^1 es incompleta, y en cierto modo una teoría abierta.

3. *La descripción y explicación de “hechos” lógicos*

Puede defenderse, aunque con menos certeza que en los casos anteriores, que el rechazo de las entidades abstractas *por el tipo de razones ofrecidas* en “Mathematical Truth” conduce a una consecuencia aun más extraña: se trata de una limitación anti-intuitiva en la posibilidad de explicar y aun describir “hechos lógicos” bastante elementales. Antes de pasar a los ejemplos, haré notar algo acerca del rechazo sugerido por el

⁹ Hay, aparentemente, una sola manera de aceptar que el universo es finito y sin embargo hay infinitas entidades materiales: adoptar algún postulado de divisibilidad infinita (en algún sentido que no implique necesariamente partición física) de los cuerpos, o las partículas, o alguna entidad física apropiada. Es muy difícil evaluar la plausibilidad de este postulado.

segundo cuerno del dilema de Benacerraf. Las entidades abstractas han sido rechazadas por muchas razones distintas. En el caso de Benacerraf, las razones aducidas son epistémicas. Los objetos que se conocen deben estar en alguna suerte de conexión causal, directa o no, con el sujeto cognoscente. Entonces no hay entidades abstractas conocidas por el hombre, porque las entidades de ese tipo no están en el espacio-tiempo y son “causalmente inertes”. El argumento no conduce a la recusación de entidades abstractas en general; más bien lleva el rechazo de entidades abstractas que puedan ser conocidas por nosotros. Pero para muchos efectos el último rechazo importa tanto como el primero (si se trata de introducir entidades abstractas para fundamentar el conocimiento de alguna disciplina en el conocimiento de tales entidades, por ejemplo). Además podría sostenerse que la idea de entidades abstractas que no podamos conocer de ninguna manera carece de sentido, o quizás de utilidad teórica. De cualquier manera, quería señalar aquí que las mismas razones usadas por Benacerraf para hacer dudar de las entidades abstractas, pueden usarse para poner en tela de juicio entidades posibles no actualizadas, ya sean del tipo de los hechos, o de los objetos posibles. En efecto, parece que tampoco tenemos interacción causal alguna con posibilidades que no se han actualizado.

Supongamos que las dudas ontológicas mencionadas en el párrafo anterior se formulan de la manera más fuerte, y ponen en cuestión no sólo la existencia *del conocimiento* de entidades abstractas y posibles sino también *la existencia misma* de tales entidades. Querría señalar aquí algunas consecuencias que se verán mejor discutiendo un ejemplo.

Puede ocurrir que los matemáticos descubran una teoría de conjuntos K en el año 2001 y la desarrollen hasta el año 2012. Puede ocurrir también que ese año se derive de la teoría un teorema contradictorio quedando así establecida la inconsistencia de K . ¿Y qué ocurriría si en el año 2010 se abandonara totalmente el estudio de K , para no retomarse jamás en el espacio-

tiempo (en un ejemplo extremo porque la humanidad entera pereció en el 2010)? ¿Habría sido inconsistente K en ese caso? Observe el lector que aparentemente no se puede afirmar que K es inconsistente en la última situación mencionada sin apelar a alguna suerte de entidades posibles o abstractas: debemos suponer la existencia de una secuencia de fórmulas (una prueba de la contradicción) al modo de una entidad matemática abstracta, o, si se prefiere, postular la *posibilidad* de escribir una prueba tal. Para disminuir el uso de entidades posibles en la discusión del ejemplo podemos preguntarnos simplemente si existirán a lo largo del espacio-tiempo teorías que sean formuladas de manera concreta por algunos hombres (o sea que sean escritas, o publicadas, etc.), y que sean contradictorias sin que ello se descubra jamás. Pareciera que una situación tal puede darse en la realidad; sin embargo parece también que no podría describirse siquiera si no aceptamos más que entidades concretas.

No estoy seguro de las conclusiones sugeridas en esta sección, pero al menos quiero dejar planteado el problema de si el abandono total de entidades “causalmente inertes” podría limitar la capacidad de nuestras teorías para describir “hechos lógicos” que podrían darse en la realidad. Al hablar de hechos que “podrían darse” menciono aparentemente hechos posibles. Pero el problema que deseo dejar formulado es éste: si un hecho como el de la teoría K abandonada en el 2010 acaece, ¿puede describirse en un lenguaje que no admita ni entidades abstractas ni posibles no actualizadas? O, de una manera quizás más exacta: ¿puede darse sentido a la idea de *teoría contradictoria cuya inconsistencia jamás es descubierta* si se han hecho los “recortes ontológicos” antes aludidos?

4. *Observaciones finales*

Al exponer este trabajo en reuniones académicas se me ha sugerido varias veces que mi escrito da argumentos fuertes en

favor de la existencia de entidades abstractas (sobre todo las secciones 1 y 3). Desearía expresar aquí que ésta no es mi convicción. Lo que he querido señalar en las secciones anteriores es que si se adopta acerca de la matemática lo que sugiere uno de los cuernos del dilema de Benacerraf, se plantean varios problemas que obligan a revisar o reexaminar nuestras ideas de la lógica estándar. Pero las consecuencias que extraigo del cuerno del dilema elegido no parecen ofrecer con claridad una reducción al absurdo del supuesto analizado.¹⁰

BIBLIOGRAFÍA

- Acero, Juan José y Tomás Calvo Martínez, *Symposium Quine*, Universidad de Granada, 1987.
- Benacerraf, Paul, "Mathematical Truth", en Benacerraf y Putnam (eds.), *Philosophy of Mathematics*, 2a. ed., Cambridge University Press, 1983, pp. 403–420.
- Benacerraf, Paul, "What Numbers Could not Be", en Benacerraf y Putnam (eds.), *Philosophy of Mathematics*, 2a. ed., Cambridge University Press, 1983, pp. 272–294.
- Benacerraf, Paul y Hilary Putnam (eds.), *Philosophy of Mathematics*, 2a. ed., Cambridge University Press, 1983.
- Field, Hartry, "Tarski's Theory of Truth", en Platts (ed.), *Reference, Truth and Reality*, Routledge & Kegan Paul, 1980, pp. 83–110.
- Mendelson, Elliott, *Introduction to Mathematical Logic*, 3a. ed., 1987.
- Mosterín, Jesús, "La frontera entre lógica y teoría de conjuntos", en Acero y Calvo, *Symposium Quine*, Universidad de Granada, 1987, pp. 209–222.
- Orayen, Raúl, *Lógica, significado y ontología*, UNAM/Instituto de Investigaciones Filosóficas, México, 1989.
- Platts, Mark (ed.), *Reference, Truth and Reality*, Routledge & Kegan Paul, 1980.

¹⁰ Una contingencia que favoreció la concisión de este ensayo es que un intento que yo había hecho de dar una solución positiva a los problemas del dilema de Benacerraf resultó absolutamente defectuoso.

SUMMARY

In this paper I point out three consequences that apparently follow from adopting one horn of Benacerraf's dilemma (the one rejecting abstract entities as objects referred to by mathematical language). These consequences are:

1. The usual explanation of the successful application of logic to mathematics is lost.
2. Logic is incomplete.
3. Some "logical facts" cannot be described any more. In other words, giving up certain entities weakens the explanatory power of our theories to cope with some elementary logical facts.