

PARADOJAS CLÁSICAS EN EL CONTEXTO DE LA LÓGICA NO MONÓTONA

Andreas Beck

Departamento de Lógica e Filosofía da Ciencia
Universidade de Santiago de Compostela

«When I came home, I expected a surprise and there was no surprise for me, so, of course, I was surprised».

L. Wittgenstein

«Decir 'método matemático' carece de sentido; debería decirse 'método natural', porque es éste último sólo el que se aplica a la Matemática».

G. C. Lichtenberg

Introducción

La cita anterior se encuentra en los *Sudelbücher* del físico y filósofo Georg Christoph Lichtenberg (1742-1799). No pocos científicos compartieron a lo largo de la historia esta apreciación relativa a la lógica matemática, lo cual condujo a la exigencia de aplicarla también a demostraciones fuera de las Matemáticas. Sin embargo, los intentos concretos de llevar a cabo esta exigencia revelaron la imposibilidad de la empresa. Prevalció la concepción de que, aunque la demostración matemática sea el medio ideal, se trata de un método demostrativo tan rígido que fuera de las Matemáticas sólo permite inferir conclusiones triviales, mientras que las afirmaciones interesantes no son demostrables.

Este reconocimiento es resultado de una reflexión teórica, porque pareciendo las conclusiones triviales menos interesantes, no han dado lugar a una investigación más profunda. Cuando se obtienen conclusiones no tan triviales, y se constata que es posible demostrar matemáticamente lo contrario (es decir, algo evidentemente falso), se habla de «paradoja» y suscita asombro.

Con el desarrollo de programas para el *razonamiento automático* en *inteligencia artificial* se modificó este punto de vista. Se probó que fuera de las Matemáticas la teoría de demostración matemática sólo conduce a resultados con sentido en casos excepcionales, lo que tiene su razón de ser especialmente en su *propiedad de monotonía*; y se desarrollaron como alternativas a la lógica matemática clásica lógicas no monótonas, que resultaron ser teorías de demostración adecuadas, particularmente en *bases de conoci-*

miento. Desde esta perspectiva, mucho de lo que desde el punto de vista de la lógica matemática clásica es considerado una paradoja, pierde ese carácter. La llamada «paradoja del verdugo» ofrece un bonito ejemplo de ello.

La paradoja del verdugo

K se sienta como acusado en la sala de juicios. En una sesión que dura todo el día X_0 , K es declarado culpable, y el Juez hace recaer sobre él la siguiente sentencia:

K será ahorcado exactamente al mediodía de uno de los cien días venideros X_1, \dots, X_{100} , y ello será para él una sorpresa, esto es, él no podrá saber la tarde anterior a su ejecución que será ahorcado al mediodía siguiente.

Esa misma tarde K reflexiona sobre su destino en la celda de su prisión. Comienza preguntándose si puede ser ahorcado el último día posible para la ejecución X_{100} , y se da cuenta de que eso es imposible: la tarde del día X_{99} él sabría que hasta entonces no había sido ahorcado, y que por lo tanto tendría que serlo el día X_{100} , lo cual ya no sería una sorpresa, y por lo tanto no es posible según la sentencia. En consecuencia, sólo son posibles para la ejecución los días X_1, \dots, X_{99} . Ahora bien: si el día X_{99} es el último día posible de ejecución, la tarde del día X_{98} K puede inferir con idéntica argumentación que tampoco va a ser ahorcado ese día, ya que en la tarde anterior al mismo ya lo sabría. Por lo tanto, restan como días posibles para la ejecución sólo los días X_1, \dots, X_{98} . De esta manera puede K eliminar sucesivamente el último día restante, y demostrar que la sentencia es contradictoria y no puede ser cumplida.

El resultado de la argumentación de K es evidentemente falso, de modo que hay que preguntarse dónde ha cometido un fallo. Hasta ahora se ha tratado ese problema filosóficamente. Quine, por ejemplo, argumentó que es imposible inferir analíticamente sobre el futuro, mientras que Shaw objetó por su parte que no era razonable suponer que K supiera que iba a ser ahorcado, y al mismo tiempo intentara demostrar que no iba a ser ahorcado. En este artículo no se va a discutir si estas objeciones son correctas o no, sino que se va a intentar poner de manifiesto que lo que a primera vista parece paradójico, en una reflexión más exacta no es más que un ejemplo típico que pone de manifiesto que en los sistemas basados en conocimiento hay que inferir con lógica no monótona.

Para facilitar el análisis vamos a simplificar la paradoja del verdugo y traducirla a un lenguaje formal. Vamos a partir de que la sentencia fue pronunciada un domingo por la tarde, y nos contentamos con el lunes, el martes y el miércoles como días posibles para la ejecución. La sentencia dice por lo tanto¹:

¹ Véase: «Recuperación de una paradoja», en R. Montague, *Ensayos de filosofía formal*, Madrid, 1977.

O (Al condenado K se le ahorca el lunes al mediodía pero no el martes ni el miércoles y K no sabe el domingo por la tarde que es verdadera la siguiente sentencia: «A K se le ahorca el lunes al mediodía»), o

(Al condenado K se le ahorca el martes al mediodía pero no el lunes ni el miércoles y K no sabe el lunes por la tarde que es verdadera la siguiente sentencia: «A K se le ahorca el martes al mediodía») o

(Al condenado K se le ahorca el miércoles al mediodía pero no el lunes ni el martes y K no sabe el martes por la tarde que es verdadera la siguiente sentencia: «A K se le ahorca el miércoles al mediodía»).

Sean L, M y X respectivamente las sentencias «A K se le ahorca el lunes al mediodía», «A K se le ahorca el martes al mediodía» y «A K se le ahorca el miércoles al mediodía». Sean d, l y m respectivamente las tardes de los días domingo, lunes y martes. Para cada tarde $p \in \{d, l, m\}$ sea $K_p(Q)$ la sentencia «K sabe Q en el tiempo p». Si el conocimiento Q es independiente del día, escribimos brevemente $K(Q)$.

Después de estas introducciones podemos formular el mandato y el hecho que K conoce ese mandato:

$$(A_0) \quad (L \wedge \neg M \wedge \neg X \wedge \neg K_d(L)) \vee (\neg L \wedge M \wedge \neg X \wedge \neg K_l(M)) \\ \vee (\neg L \wedge \neg M \wedge X \wedge \neg K_m(X)) \\ (A_1) \quad K_d(A_0)$$

En general tiene el operador K las siguientes propiedades:

K no puede saber algo que no es el caso. O sea, si K sabe Q entonces es el caso Q:

$$(A_2) \quad K(Q) \Rightarrow Q$$

Si K sabe Q y sabe que R se sigue de Q entonces sabe también R:

$$(A_3) \quad (K(Q \Rightarrow R) \wedge K(Q)) \Rightarrow K(R)$$

Si K sabe Q entonces sabe que sabe Q:

$$(A_4) \quad K(Q) \Rightarrow K(K(Q))$$

Si K puede saber algo independiente del día entonces lo sabe el domingo por la tarde:

$$(A_5) \quad K(Q) \Rightarrow K_d(Q)$$

Si K sabe algo un día, también lo sabe al día siguiente:

$$(A_6) \quad K_d(Q) \Rightarrow K_l(Q) \Rightarrow K_m(Q)$$

La argumentación de K sería en nuestro formalismo la siguiente:

(B0) Supongamos X:

Entonces se sigue de (A0): $\neg L \wedge \neg M \wedge \neg K_m(X)$.

Si K no ha sido ahorcado ni el lunes ni el martes al mediodía, sabe el martes por la tarde que no ha sido ahorcado ni el lunes ni el martes al mediodía. O sea:

$$(A_7) \quad (\neg L \wedge \neg M) \Rightarrow K_m(\neg L \wedge \neg M)$$

Además vale obviamente²:

² Para evitar el controvertido Axioma de la Necesidad epistemológica que proporcione la posibilidad de inferir $\vdash K(Q)$ de $\vdash Q$ (o sea K puede saber todo lo que es demostrable), vamos a considerar cada inferencia de ese tipo que K usa en su argumentación como axioma.

$$(A_8) \quad K(A_0) \Rightarrow K(L \vee M \vee X)$$

Y por tanto:

$$(A_9) \quad K(L \vee M \vee X) \wedge K_m(\neg L \wedge \neg M) \Rightarrow K_m(X)$$

lo que significa una contradicción y se sigue que X no es posible.

Pero no sólo podemos concluir $\neg X$. El punto clave es que también podemos concluir que K sabe $\neg X$ porque K puede concluir $\neg X$, como acabamos de hacer independiente del día. Es decir:

$$(A_{10}) \quad K(A_0) \wedge K(B_0) \Rightarrow K(\neg X)$$

Ese resultado hace posible demostrar que tampoco se puede ahorcar a K el martes:

(B₁) Supongamos M :

$$\text{Entonces se sigue de } (A_0): \neg L \wedge \neg X \wedge \neg K_1(M)$$

Si K no ha sido ahorcado el lunes al mediodía, K sabe el lunes por la tarde que no ha sido ahorcado el lunes al mediodía. Es decir:

$$(A_{11}) \quad \neg L \Rightarrow K_1(\neg L)$$

Obviamente vale:

$$(A_{12}) \quad K(L \vee M \vee X) \wedge K(\neg X) \wedge K_1(\neg L) \Rightarrow K_1(M)$$

donde las premisas valen por (A_8) , (A_{10}) y (A_{11}) , y por tanto tenemos de nuevo una contradicción y se sigue que M no es posible.

De modo análogo podemos decir que el razonamiento (B_1) puede ser hecho por K independiente del día. Es decir:

$$(A_{13}) \quad K(A_0) \wedge K(B_1) \Rightarrow K(\neg M)$$

En total ya sabemos $\neg X$, $\neg M$, $K(\neg X)$ y $K(\neg M)$, con lo que se sigue directamente $\neg L$ porque:

(B₃) Supongamos L :

$$\text{Entonces se sigue de } (A_0): \neg M \wedge \neg X \wedge \neg K_d(L)$$

Vale:

$$(A_{14}) \quad K(L \vee M \vee X) \wedge K(\neg X) \wedge K(\neg M) \Rightarrow K(L)$$

y por siguiente $K_d(L)$ lo que es una contradicción.

Hemos demostrado que $\neg L$, $\neg M$ y $\neg X$ se sigue lógicamente del mandato (A_0) y del hecho de que K sabe (A_0) y es capaz de juzgar razonablemente. Como eso es contradictorio se sigue que el mandato no puede ser ejecutado.

¿El ferrocarril es paradójico?

Supongamos que K consigue huir de la prisión. En la misma tarde del domingo llega sin ser reconocido, por bajos y altos fondos, a la estación de Santiago de Compostela. Le han aconsejado ocultarse en Oporto. Saca un horario del bolsillo y lee a través de las gafas de sol que se ha puesto para camuflarse:

(H1) *Horario de trenes con salida de Santiago de Compostela.*

(Z1) 9.00 Santiago de Compostela-Vigo

(Z2) 12.00 Santiago de Compostela-Orense

- (Z3) 16.00 Santiago de Compostela–La Coruña
 (Z4) 20.00 Santiago de Compostela–Vigo
 (Z5) 23.00 Santiago de Compostela–La Coruña

K empieza a ponerse nervioso. ¿No había a las diez un tren directo a Oporto? ¿Habrá caído en una trampa? Mira a su alrededor. Su mirada recae sobre un tablón de anuncios colgado del techo con el siguiente texto:

(H2) *Horario de trenes con salida de Santiago de Compostela.*

- (Z1) 9.00 Santiago de Compostela–Vigo
 (Z2) 12.00 Santiago de Compostela–Orense
 (Z3) 16.00 Santiago de Compostela–La Coruña
 (Z4) 20.00 Santiago de Compostela–Vigo
 (Z6) 22.00 Santiago de Compostela–Oporto
 (Z5) 23.00 Santiago de Compostela–La Coruña

Evidentemente su horario está anticuado. La dirección de los ferrocarriles ha podido comprar otra máquina y ha añadido al horario un tren adicional para Oporto. K se dirige a la vía 3, donde ya esta esperando el tren. Vagón de fumadores, segunda clase. Se dispone a subir. Sin embargo, vacila. Algo no marcha bien.

Lógica monótona y lógica no monótona

Una de las propiedades características de la teoría de demostración matemática consiste en que todo lo que se puede inferir de un conjunto de axiomas Δ también se puede inferir de cada supraconjunto Ψ al que pertenece Δ . Esta propiedad, denominada *monotonía*, se expresa formalmente del siguiente modo:

(M1) $\Delta \vdash \varphi, \Delta \subseteq \Psi \rightarrow \Psi \vdash \varphi$

donde Δ, Ψ designan conjuntos de axiomas y φ designa un enunciado.

Para poder admitir inferencias para las que no vale (M1) es necesario un nuevo concepto, no monótono, de demostración: $\vdash\sim$. Generalmente vale para $\vdash\sim$:

(NM1) $\Delta \vdash\sim \varphi, \Delta \subseteq \Psi \not\vdash \Psi \vdash \varphi$

Como regla de inferencia de $\vdash\sim$ es conveniente aplicar reglas *por defecto*, es decir, reglas cuya forma es:

(A) $\alpha : \beta / \gamma$

donde α, β, γ designan enunciados. A través de (A) se sigue γ de un conjunto Δ , cuando α sigue de Δ y $\Delta \cup \{\beta\}$ es consistente.

En el caso del ferrocarril es evidente que K no ha concluido *monótonamente* de su horario que a las diez no hay ningún tren directo a Oporto, puesto que aplicó la regla *por defecto* conforme a la cual todos los trenes que no aparecen en su horario no salen de Santiago. Formalmente eso se expresaría como sigue:

Sea $\Delta :=$ (H1) y $\varphi :=$ (Z6 22.00 Santiago de Compostela-Oporto) entonces vale:

$$\Delta \Rightarrow \neg\varphi$$

donde la regla: $\neg\varphi / \neg\varphi$ es aplicada³.

Con $\Delta \subseteq \Psi$: = (H2) vale

$$\Psi \not\Rightarrow \neg\varphi$$

porque $\Psi \cup \{\neg\varphi\}$ es inconsistente y por lo tanto la regla: $\neg\varphi / \neg\varphi$ no es aplicable.

(M1) es equivalente⁴ a la propiedad que todo lo que es inferible de un conjunto de axiomas se puede añadir a este conjunto sin que el conjunto de los enunciados inferibles crezca. Formalmente esto se dice:

$$(M2) \quad \Delta \vdash \varphi \rightarrow \Delta \cup \{\varphi\} \mid - \mid \Delta$$

Esta propiedad tan práctica en las Matemáticas, conforme a la cual es posible añadir un enunciado inferido al conjunto de axiomas, no es válida en sistemas de demostración no monótonos.

Las inferencias utilizadas en la argumentación de K muchas veces no cumplen (M1). En (B₀) se infieren las conclusiones verdaderas $\neg X$, $K(\neg X)$ pero K aplica estos resultados como axiomas, lo que conduce al resultado $\neg M$, que en realidad no es inferible a partir de los axiomas anteriormente dados Δ ⁵, sino sólo a partir de $\Delta \cup \{\neg X\} \cup \{K(\neg X)\}$. De modo análogo K concluye en (B₂) con la ayuda de $K(\neg M)$ y $K(\neg X)$ a $K(L)$, y deduce a partir de ello la contradicción de la sentencia Δ . Esto es falso: en realidad lo único que K demuestra es que el conjunto establecido de axiomas Δ unido a los resultados por él mismo elaborados $K(\neg M)$, $K(\neg X)$ y $K(L)$ es contradictorio. Al no haber concluido monótonamente, esto no equivale a la contradicción de la sentencia como conjunto de axiomas establecidos.

Particularmente (A₁₀), (A₁₃) y (A₁₄) no cumplen la propiedad de monotónía y deben ser entendidos como $\vdash -$ inferencias. Ellas aplican las reglas por defecto

$$(A_{10}') \quad (K(A_0) \wedge K(B_0)): K(\neg X) / K(\neg X)$$

$$(A_{13}') \quad (K(A_0) \wedge K(B_1)): K(\neg M) / K(\neg M)$$

$$(A_{14}') \quad (K(L \vee M \vee X) \wedge K(\neg X) \wedge K(\neg M)): K(L) / K(L)$$

Si se corrige la formalización de K en el sentido expresado no obtenemos la contradicción. Más bien resulta que $K(L)$ no es consistente con la sentencia unida a $K(\neg X)$ y $K(\neg M)$. Por lo tanto, las premisas de la regla (A₁₄') no son satisfechas, y en consecuencia no es posible concluir $K(L)$.

³ : $\neg\varphi / \neg\varphi$ debe leerse: Si $\neg\varphi$ es consistente con Δ , entonces vale $\neg\varphi$. En nuestro ejemplo esto se corresponde con la siguiente conclusión: Si φ es un trayecto desde Santiago de Compostela, que no pertenece al horario, entonces vale $\neg\varphi$.

⁴ Véase, Apéndice.

⁵ Δ sea el conjunto de axiomas que representa: la sentencia, el hecho que K conoce esa sentencia y es capaz de inferir lógicamente.

K muere y nos vamos a Oporto

Resumidamente, se mantiene que K puede excluir lícitamente el miércoles como uno de los posibles días para su ejecución. Pero como esta inferencia amplía el conjunto de enunciados demostrables a través de la sentencia, la inferencia no satisface (M1), y debe ser interpretada como no monótona. Por eso $\Delta \cup \{\neg X\}$ no es equivalente a Δ .

Si K aplicara correctamente la regla *por defecto*, no podría inferir el domingo por la tarde que puede ser ahorcado el mediodía del lunes, porque $\Delta \cup \{\neg X\} \cup \{K(\neg X)\} \cup \{\neg M\} \cup \{K(\neg M)\} \not\vdash \neg L^6$.

Por lo tanto, K debe contar con que es posible que sea ahorcado el lunes. Si transcurre el mediodía del lunes y K no fue ejecutado, entonces debe, a causa de $\Delta \cup \{\neg X\} \cup \{K(\neg X)\} \cup \{\neg L\} \cup \{K(\neg L)\} \not\vdash \neg M^7$

estar preparado para la posible ejecución el martes. Si no fue tampoco ahorcado el martes, entonces puede afirmar que la sentencia es contradictoria, a causa de la inconsistencia de

$\Delta \cup \{\neg L\} \cup \{K(\neg L)\} \cup \{\neg M\} \cup \{\neg K(M)\}$.

Si K opone esta objeción a su argumentación, reclamará la aplicación de reglas de demostración monótonas en lógica, en las cuales no se cumpla la propiedad de monotonía. Esto es, él no puede utilizar el tren a Oporto en su huida, porque a partir de su argumentación ese tren no debería existir.

Apéndice

Teorema: $(\Delta \vdash \phi, \Delta \subseteq \Psi \rightarrow \Psi \vdash \phi) \leftrightarrow (\Delta \vdash \phi \rightarrow \Delta \cup \{\phi\} \mid - \mid \Delta)$

Demostración:

‘ \rightarrow ’: ‘ \vdash ’: $\Delta \vdash \Delta, \Delta \subseteq \Delta \cup \{\phi\} \rightarrow \Delta \cup \{\phi\} \vdash \Delta$

‘ \vdash ’: $\Delta \vdash \Delta, \Delta \vdash \phi \rightarrow \Delta \vdash \Delta \cup \{\phi\}$

‘ \leftarrow ’: $\Delta \vdash \phi, \Delta \subseteq \Psi, \Delta \cup \{\phi\} \mid - \mid \Delta \rightarrow \Psi \vdash \phi$

Referencias

Genesereth, M./Nilsson, N., *Logical Foundations of Artificial Intelligence*, Los Altos, 1987.

González, A./Damkel, D., *The Engineering of Knowledge-Based Systems*, London, 1993.

Lichtenberg, G. C., *Sudelbücher*, München, 1965.

⁶ De manera análoga al caso $\Psi \not\vdash \neg\phi$ del ferrocarril, $\neg L$ no puede ser inferido a través de la correspondiente regla por defecto, en la medida en que $\neg L$ añadido a las premisas conduce a inconsistencia.

⁷ Como en 5). Se observa que el domingo por la tarde K todavía puede inferir $\neg M$, porque $\{\neg L, K(\neg L)\}$ aún no forma parte de las premisas.

- Montague, R., *Ensayos de filosofía formal*, Madrid, 1977.
- Quine, W., «On a So-Called Paradox», *Mind* 62, (1953), 65-67.
- RENFE Galicia, *Horario de trenes*, Santiago de Compostela, 1995.
- Shaw, R., «The Paradox of the Unexpected Examination», *Mind* 67, (1958), 382-384.