

Inteligencia Artificial, la vía negativa y el estatus computacional de la persona

*Artificial Intelligence, the Negative Way and the Computational
Status of Persons*

JOSÉ MARTÍN CASTRO-MANZANO
UPAEP, México
josemartin.castro@upaep.mx

RESUMEN

Una de las preguntas más interesantes que la inteligencia artificial y la filosofía de la computación han hecho nacer es la pregunta por nuestro estatus computacional como personas humanas. En este trabajo discutimos dicho estatus mostrando la argumentación por vía negativa de Bringsjord para justificar que somos entidades computacionales *sui generis*. Para discutir el estatus computacional revisamos el concepto de “inteligencia artificial” y su relación con la Tesis de Church-Turing. Posteriormente comentamos la discusión Lucas vs Bringsjord para concluir por qué se dice que somos hipermáquinas.

Palabras clave: inteligencia artificial, persona, Tesis de Church-Turing, máquina de Turing, hipermáquina.

ABSTRACT

One of the most interesting questions that artificial intelligence and philosophy of computing have generated is the question about the computational status of human persons. In this paper we discuss the computational status of human persons by showing Bringsjord’s negative way argumentation in order to justify that we are *sui generis* computational entities. As a means to discuss the computational status we review the concept of “artificial intelligence” and its relation with the Church-Turing Thesis. Then we comment on the Lucas *vs* Bringsjord discussion for the purpose of concluding why it is said that we are hypermachines.

Keywords: artificial intelligence, person, Church-Turing Thesis, Turing machine, hypermachine.

Recepción del original: 17/01/15
Aceptación definitiva: 08/06/15

*Para venir a lo que no eres
has de ir por donde no eres*
San Juan de la Cruz

1. Introducción

La pregunta por el estatus computacional de las personas humanas es la versión formal de la pregunta informal: *¿Somos máquinas?* Esta pregunta es importante de suyo, pero también lo es por sus consecuencias en metafísica, antropología y ética, pues el estatus computacional de las personas humanas se relaciona con la naturaleza de las personas humanas, su lugar en el mundo y sus relaciones mutuas en el contexto tecnológico contemporáneo.

Dada esta importancia, en este trabajo discutimos el estatus computacional de las personas humanas, mostrando la argumentación por vía negativa de Selmer Bringsjord para justificar que somos entidades computacionales *sui generis*.

Para discutir el estatus computacional revisamos brevemente el concepto de “inteligencia artificial” (IA) y su relación con la Tesis de Church-Turing (Sección 2). Posteriormente comentamos la discusión Lucas *vs* Bringsjord para exponer en qué sentido se dice que somos hiper máquinas (Sección 3). Por último, terminamos con un resumen de la discusión y una exposición de nuestra postura (Sección 4).¹

2. Inteligencia artificial

El objetivo del siguiente recorrido (2.1) no es definir qué es la IA (2.2), sino explicar el origen de la Tesis de Church-Turing y su relación con nuestro estatus computacional como personas.

2.1. Detalles de la historia de la IA

Como desiderátum, la IA se puede rastrear en varios momentos que, por razones de espacio, únicamente mencionamos, pero que proporcionan una idea general de lo que la IA pretende: en la *Biblioteca* de Apolodoro,² la máquina de

¹ El autor agradece a los árbitros anónimos sus amables comentarios, sus precisas correcciones y sus útiles sugerencias.

² APOLODORO, *Biblioteca*, Madrid: Gredos, 2002, I, 1.9.23.

Antikythera,³ la silogística de Aristóteles,⁴ el Popol-Vuh,⁵ el legendario medieval alrededor del Golem,⁶ el homúnculo renacentista de Paracelso y de Fausto,⁷ el *Ars Magna* de Llull,⁸ la *roue pascaline*,⁹ el razonamiento según Hobbes,¹⁰ el *calculus ratiocinator* y la *mathesis universalis* de Leibniz,¹¹ el álgebra de Boole,¹² los diagramas de Venn,¹³ el piano lógico de Jevons,¹⁴ *Frankenstein* de Shelley,¹⁵ la máquina diferencial de Babbage,¹⁶ *Begriffsschrift* de Frege,¹⁷ R.U.R. de Capek,¹⁸ el cálculo λ de Church,¹⁹ las máquinas de Post,²⁰ los autómatas de Turing,²¹ la representación neuronal de McCulloch y Pitts²² y el teorista lógico de Simon, Newell y Shaw.²³

³ PRICE, D. DE S., "Gears from the Greeks. The Antikythera Mechanism: A Calendar Computer from ca. 80 B. C." en *Transactions of the American Philosophical Society*, New Series 64 (7): 1-70, 1974.

⁴ ARISTÓTELES, "Primeros Analíticos", en *Tratados de Lógica*, Madrid: Gredos, 1998.

⁵ NEGRETE, J., "El Popol-Vuh. Un experimento *gedanke*" en NEGRETE, J., GONZÁLEZ, E., *De la filosofía a la inteligencia artificial*, México: Noriega Editores, 1992.

⁶ SCHOLEM, G., "The Idea of the Golem", en *On the Kabbalah and Its Symbolism*, trad. Manheim, R., New York: Schocken, 1965.

⁷ CAMPBELL, M. B., "Artificial Men: Alchemy, Transubstantiation, and the Homunculus", en *Republics of Letters: A Journal for the Study of Knowledge, Politics, and the Arts* 1, núm. 2, 2010.

⁸ *Selected Works of Ramon Llull (1232-1316)*, ed. y trad. Bonner, A., vol. 2, Princeton, N.J.: Princeton University Press, 1985, pp. 317-646.

⁹ MARGUIN, J., *Histoire des instruments et machines à calculer, trois siècles de mécanique pensante 1642-1942*, Hermann, 1994.

¹⁰ HOBBS, TH., *Leviathan: Or the Matter, Forme, and Power of a Common-Wealth Ecclesiasticall and Civill*, ed. Shapiro, I., Yale University Press, 2010.

¹¹ LEIBNIZ, G.W., "Lengua universal, característica y lógica". *Obras filosóficas y científicas*, 5. Comares, Granada, 2013. COUTURAT, L., *La logique de Leibniz d'après de documents inédits*, Hildesheim, 1962.

¹² BOOLE, G., *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, Oxford: Basil Blackwell, 1951. BOOLE, G., *An Investigation of The Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*, Dover, 1958.

¹³ VENN, J., "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings", en *Dublin Philosophical Magazine and Journal of Science* 9 (59): 1-18, 1880.

¹⁴ BARRETT, L., CONNELL, M., "Jevons and the Logic 'Piano' ", en *The Rutherford Journal*, vol. 1, Issue 1, 2006.

¹⁵ SHELLEY, M., *Frankenstein; or, The Modern Prometheus*, New York: Pearson Longman, 2007.

¹⁶ LINDGREN, M., *Glory and Failure: The Difference Engines of Johann Müller, Charles Babbage, and Georg and Edvard Scheutz*, trad. McKay, C.G., Cambridge, Massachusetts: MIT Press, 1990.

¹⁷ FREGE, G., *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle, 1879. FREGE, G., "Concept Script", trad. BAUER-ENGELBERG, S., en VAN HEIJENOORT, J., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, 1967.

¹⁸ CAPEK, K., *R.U.R.*, trad. SELVER, P., PLAYFAIR, N., Dover, 2001.

¹⁹ CHURCH, A., "A set of postulates for the foundation of logic" en *Annals of Mathematics*, Series 2, 33:346-366, 1932. CHURCH, A., "An unsolvable problem of elementary number theory" en *American Journal of Mathematics*, Volume 58, No. 2., 1936, pp. 345-363. CHURCH, A., "A Formulation of the Simple Theory of Types" en *Journal of Symbolic Logic* 5: 1940.

²⁰ POST, E. L., "Finite Combinatory Processes. Formulation I" en *Journal of Symbolic Logic*, 1, pp. 103-105, 1936.

²¹ TURING, A.M., "Intelligent Machinery" reimpresso en EVANS, C.R, ROBERTSON, A.D.J., *Cybernetics: Key Papers*, Baltimore: University Park Press, 1968. p. 31. La primera versión de las máquinas de Turing se encuentra en TURING, A.M. "On computable numbers, with an application to the Entscheidungsproblem" en *Proceedings, London Mathematical Society*, 2, 42 pp. 230-265, y no. 43, pp. 544-546, 1936.

²² MCCULLOCH, W.S., PITTS, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity" en *Bulletin of Mathematical Biophysics*, Vol 5, pp 115-133, 1943.

²³ NEWELL, A., SIMON, H.A., *The Logic Theory Machine. A Complex Information Processing System*, The Rand Corporation, 1956.

Todos estos proyectos e ideales de exteriorización y mecanización del razonamiento recibieron un denominador común cuando la IA nació como disciplina formal en la conferencia de Dartmouth de 1956,²⁴ y generaron, por sus supuestos y sus implementaciones, dos programas de investigación: uno más atento a los *abstracta*²⁵ (en la tradición de John McCarthy) y otro más dedicado a los *concreta* (en la línea de Rodney Brooks).²⁶ El primero culminó en la Hipótesis del Sistema Físico de Símbolos (HSFS), y el segundo, en la Hipótesis de la Incorporación. De estas dos hipótesis, la primera ha sido más influyente por su generalidad y porque se conecta directamente con la cuestión de nuestro estatus computacional.

La HSFS afirma, en palabras de sus desarrolladores, Newell y Simon, que:

A physical symbol system has the necessary and sufficient means for general intelligent action. By necessary, we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By sufficient we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence.²⁷

Lo relevante de esta hipótesis es que el concepto de “sistema físico de símbolos” no está restringido a sistemas artificiales, pues, en esencia, un sistema físico de símbolos es cualquier entidad que se puede traducir a una máquina de Turing.²⁸ Así, en *A Basis for a Mathematical Theory of Computation*, John McCarthy, el padre de la IA, nos dice que:

We know that any intellectual process that can be carried out mechanically can be performed by a general purpose digital computer.²⁹

Él introduce una relación de traducibilidad entre los procesos inteligentes que llevamos a cabo las personas humanas y las máquinas de Turing haciendo uso de la Tesis de Church-Turing,³⁰ la cual, en palabras del propio Turing, en su tesis doctoral, *Systems of Logic Based on Ordinals*, establece que:

²⁴ MCCARTHY, J., MINSKY, M., ROCHESTER, N., SHANNON, C., *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955. Cf. CREVIER, D., *AI: The Tumultuous Search for Artificial Intelligence*, New York: BasicBooks, pp. 44–46, 1993.

²⁵ GENESERETH, M.R., NILSSON, N. J., *Logical foundations of artificial intelligence*, Palo Alto, CA.: Morgan Kaufman, 1988.

²⁶ BROOKS, R.A., “Elephants Don’t Play Chess” en MAES, P., *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, MIT Press, 1994, pp. 3-15. En este trabajo, que puede leerse como un manifiesto de la *nouvelle AI* frente a la IA tradicional, Brooks considera que la HSFS es errónea y sugiere, en su lugar, una IA basada en lo que llama la *physical grounding hypothesis*.

²⁷ NEWELL, A., SIMON, H.A., “Computer science as empirical inquiry: symbols and search” en *Communications of the ACM* 19 (3): 113–126, 1976.

²⁸ Una máquina de Turing M se define como una 7-tupla $M = \langle Q, \Gamma, b, \Sigma, \delta, q_0, F \rangle$ donde Q es un conjunto finito no-vacío de estados, Γ es un conjunto finito no-vacío de símbolos de alfabeto, $b \in \Gamma$ denota el símbolo vacío, $\Sigma \subseteq \Gamma - \{b\}$ es el conjunto de entradas, $q_0 \in Q$ es el estado inicial, $F \subseteq Q$ es el conjunto de estados de paro y δ es una función tal que $\delta: (Q - F) \times \Gamma \rightarrow Q \times \Gamma \times \{L, D\}$, donde L y D son movimientos a la izquierda o a la derecha, respectivamente. Cf. HOPCROFT, J., ULLMAN, J., *Introduction to Automata Theory, Languages, and Computation*, Reading, Mass.: Addison-Wesley, 1979, p. 148.

²⁹ MCCARTHY, J., “A Basis for a Mathematical Theory of Computation” en BRAFFORT, P., HIRSHBERG, D., *Computer Programming and Formal Systems*, North-Holland, 1963, p. 1.

³⁰ KLEENE, S., *Mathematical Logic*, New York: Wiley, 1967, p. 237.

[A] function is effectively calculable if its values can be found by some purely mechanical process. We may take this literally, understanding that by a purely mechanical process one which could be carried out by a machine. It is possible to give a mathematical description, in a certain normal form, of the structures of these machines.³¹

Y por tanto, cualquier proceso que pueda definirse de forma algorítmica puede traducirse a una máquina de Turing, o en términos de Church, a expresiones del cálculo λ que, a su vez, son equivalentes a un conjunto de funciones recursivas.

2.2. Definiciones operacionales de IA

Ahora, como ciencia, la IA se dedica al estudio, el diseño y la implementación de artefactos computacionales con cierta finalidad. Esta finalidad, siguiendo a Russell y Norvig, puede entenderse en cuatro sentidos (Tabla 1):

Finalidad	Pensar	Actuar
Racionalmente	S1 ³²	S2 ³³
Humanamente	S3 ³⁴	S4 ³⁵

Tabla 1. Sentidos de la IA según Russell y Norvig³⁶

Bajo la Tesis de Church-Turing y el paradigma de la HSFS, la IA es un caso de S1 (S2) que ha logrado llevar a buen término el proyecto de exteriorización y mecanización del razonamiento,³⁷ lo cual ha dado origen a la pregunta crucial del computacionalismo (¿Los sistemas de IA pueden pensar?), pero también ha abierto la cuestión simétrica sobre nuestro estatus computacional: si cualquier proceso que puede definirse de manera algorítmica (como el razonamiento) puede traducirse a una máquina de Turing, y viceversa, ¿somos (reducibles a) máquinas?

³¹TURING, A.M., *Systems of Logic Based on Ordinals*, tesis doctoral, Princeton University, 1939, p. 8.

³²HAUGELAND, J., *Artificial intelligence: the very idea*, Cambridge: MIT Press, 1985.

³³CHARNIAK, E., McDERMOTT, D., *Introduction to Artificial Intelligence*, Reading, Mass.: Addison-Wesley, 1985.

³⁴KURZWEIL, R., *The Age of Intelligent Machines*, Cambridge, MA: MIT Press, 1990.

³⁵SCHALKOFF, R.J., *Artificial Intelligence: An Engineering Approach*, Michigan: McGraw-Hill, 1990.

³⁶RUSSELL, S., NORVIG, P., *Artificial Intelligence: A Modern Approach*, Saddle River, NJ: Prentice Hall, 1994.

³⁷Con respecto a este proyecto nos parece interesante mencionar, como dice Carter, que: "Given our current conceptual understanding of the mind and technical understanding of the computational wetware of the brain which gives rise to it, by far the simplest way to create something which has the capacity for embodied experience and which is *ceteris paribus* guaranteed to develop a mind in the same sense that we have a mind is still the old-fashioned biological way—to create and raise a human being". CARTER, M. *Minds and computers. An introduction to the philosophy of artificial intelligence*. Edinburgh University Press, 2007, p. 206.

3. La vía negativa

La IA en el sentido S1 da lugar a la tesis de que somos máquinas (de Turing) y por ello, la IA ha sido criticada por diferentes argumentos que pretenden encontrar casos de S3 (S4) que no pueden traducirse (o reducirse) a casos de S1 (S2). Así, por ejemplo, el *dictum* de Ada Lovelace pretende socavar las nociones de autonomía y creatividad de la IA aludiendo a que éstas son propiedades *bona fide* de S3 (S4) que son imposibles en S1 (S2);³⁸ Searle critica la semántica y la intencionalidad de la IA argumentando que son imposibles en S1 (S2) pero que son naturales en S3 (S4);³⁹ Dreyfus apela a la falta de corporalidad en S1 (S2) para justificar que hay propiedades de S3 (S4) que son necesariamente corporales;⁴⁰ y Penrose y Lucas aprovechan las limitaciones intrínsecas de los sistemas formales para sostener que tales limitaciones son propias de S1 (S2), pero no de S3 (S4). En este trabajo ponemos atención a la discusión generada por estos últimos argumentos.

Así, a continuación exponemos el argumento de Lucas (3.1) y, después, la respuesta de Bringsjord (3.2) para, posteriormente, hacer un balance de la discusión y explicar en qué sentido se dice que nuestro estatus computacional es el de hiper máquinas (3.3).

3.1. El argumento de Lucas

En *Minds, Machines and Gödel*, Lucas nos informa que:

Gödel's theorem seems to me to prove that Mechanism is false, that is, that minds cannot be explained as machines.⁴¹

Lucas interpreta la IA en el sentido S1, por lo que no sorprende que su respuesta a la pregunta por el estatus computacional de las personas humanas sea negativa. De manera más precisa, la razón por la cual Lucas sostiene que no podemos ser máquinas es porque cree encontrar un ejemplo de S3 (S4) que no puede reducirse a S1 (S2) apelando a una interpretación del Teorema VI de *Über formal unentscheidbare Sätze der Principia Mathematica* que afirma que:

Satz VI: Zu jeder -widerspruchsfreien rekursiven Klasse k von Formeln gibt es rekursive Klassenzeichen r , so daß weder v Gen r noch Neg (v Gen r) zu $Flg(k)$ gehört (wobei v die freie Variable aus r ist).⁴²

³⁸LOVELACE, A., "Translator's notes to L. F. Menabrea's 'Sketch of the analytical engine invented by Charles Babbage, Esq.'" en BOWDEN, B. V., *Faster than Thought: A Symposium on Digital Computing Machines*, New York: Pitman Publishing Co., 1953, 362-408.

³⁹SEARLE, J., *Mind, language and society*, New York: Basic Books, 1999.

⁴⁰DREYFUS, H.L., *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, MA.: MIT Press, 1992.

⁴¹LUCAS, J.R., "Minds, Machines and Gödel" en *Philosophy* 36: 112-127, 1961.

⁴²"Teorema VI: Para cada clase recursiva primitiva y $_$ -consistente de K FÓRMULAS hay un SIGNO DE CLASE r tal que ni v Gen r ni Neg(v Gen r) pertenecen a $Flg(K)$ ". GÖDEL, K., "Über formal unent-

A partir de una interpretación peculiar de las consecuencias de este teorema, Lucas cree encontrar un caso de S3 (S4) que no se puede reducir a S1 (S2); sin embargo, no parece claro cómo se puede llegar a semejante conclusión a partir de la lectura directa del teorema anterior. Por ello mostramos un par explicaciones, una informal y otra (más o menos) formal, para detallar el argumento de Lucas.

Para ofrecer una explicación informal, exponemos una versión de Rucker.⁴³ Supongamos que existe la MVU, la Máquina de la Verdad Universal, por sus siglas. La MVU es una máquina (de Turing) capaz de decidir, para cualquier proposición, si dicha proposición es verdadera o no. Sea G la proposición:

La MVU no puede afirmar que G es verdad.

Si la MVU afirma que G es verdad, entonces es inconsistente. Pero no puede ser inconsistente porque sólo puede afirmar verdades. Pero entonces debe ser el caso que la MVU no puede afirmar que G es verdad. Pero entonces la MVU es incompleta porque hay por lo menos una proposición verdadera que no puede afirmar. En consecuencia, si la MVU afirma que G es verdad, se contradice; pero si es imposible que se contradiga, entonces es incompleta. Por tanto, ya sea que la MVU afirme G o no afirme G , es inconsistente o incompleta. Y por tanto, la MVU no puede decidir algo que nosotros sí, a saber, que es verdad que la MVU no puede afirmar que G es verdad. Lucas lo resume así:

Gödel's theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true - i.e., the formula is unprovable-in-the-system-but which we can see to be true. It follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines.⁴⁴

Para desarrollar este argumento de manera semi-formal, consideremos la versión del argumento gödelinano de Penrose.⁴⁵ Sea $C(n)$ un cómputo que se aplica para cada natural n . Así, $C(n)$ será la acción de una máquina de Turing sobre n , que es la entrada a partir de la cual la máquina trabaja. La cuestión a resolver es si dicha máquina se detendrá o no, lo cual es una versión del *halting problem* o problema del paro.

scheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I" en *Monatshefte für Mathematik und Physik* 38: 173–98, 1931, p. 187. Cf. GÖDEL, K., *Obras completas*, trad. MOSTERÍN, J., Madrid: Alianza, 1989, p. 53.

⁴³RUCKER, R., *Infinity and the Mind: The Science and Philosophy of the Infinite*. Princeton, N.J.: Princeton University Press, 1982.

⁴⁴LUCAS, J.R. "Minds, Machines and Gödel", p.2.

⁴⁵PENROSE, R., *Shadows of the Mind*, Oxford: Oxford, 1994, pp. 74–75.

Para encontrar la respuesta a dicha interrogante, es necesario otro procedimiento (el algoritmo A , que es también una máquina de Turing) que contiene todas las entradas posibles para demostrar que $C(n)$ no se detendrá, es decir, que:

Si A se detiene, $C(n)$ no se detiene.

Si se denomina $A(q,n)$ a la ejecución del algoritmo A sobre el cómputo $C_q(n)$, obtenemos que:

Si $A(q,n)$ se detiene, entonces $C_q(n)$ no se detiene.

Sobre esta proposición se aplica el principio de diagonalización y se toma $q=n$:

Si $A(n,n)$ se detiene, entonces $C_n(n)$ no se detiene.

Pero resulta que $A(n,n)=C_n(n)$, por el mismo proceso de diagonalización, de modo que:

Si $A(n,n)$ se detiene, entonces $A(n,n)$ no se detiene.

Este resultado es una contradicción y bien podría significar, en términos gödelianos, que existe una proposición verdadera que no se puede demostrar; o en términos cantorianos, que existe un número real que no se puede encontrar en la lista de todos los números reales.

Por tanto, concluye esta argumentación, como el algoritmo A es incapaz de determinar que $C_n(n)$ no se detiene cuando de hecho *nosotros sabemos* que no se detiene, resulta que los humanos, a diferencia de los sistemas inteligentes artificiales basados en la Tesis de Church-Turing, tienen algo que es no-computacional, algo en S3 (S4) que no se puede reducir a S1 (S2), pues de no ser así, seríamos incapaces de salir de la contradicción producida por el proceso de diagonalización. En consecuencia, las mentes humanas no pueden reducirse a máquinas, no son máquinas.

2.2. El argumento de Bringsjord

Bringsjord y Arkoudas, siguiendo la definición de hipercómputo de Copeland,⁴⁶ proponen un argumento brillante para justificar, no que seamos máquinas de Turing, sino que somos hipermáquinas.⁴⁷ El argumento es todavía más fuerte: si Lucas (como Penrose) está en lo correcto, las personas humanas no serían máquinas; serían hipermáquinas.

⁴⁶COPELAND, B.J., "Even Turing Machines can compute uncomputable functions" en CASTI, J., *Unconventional models of computation*, London: Springer, 1998.

⁴⁷BRINGSJORD, S., ARKOUNDAS, K., "The modal argument for hypercomputing minds" en *Theoretical Computer Science* 317 (2004) 167-190.

En el ahora clásico *Elements of the Theory of Computation* Lewis y Papadimitriou sugieren dos versiones del teorema de imposibilidad de las máquinas de Turing:

Theorem 6.1.5 a) There is no algorithm that can determine, given a Turing machine m and an input string w , whether m accepts w . b) For a certain fixed Turing machine m_0 , there is no algorithm that can determine, given an input string w , whether m_0 accepts w .⁴⁸

El argumento modal de Bringsjord comienza representando la versión b) del teorema anterior modalmente:

$$\forall m \exists w \neg \Diamond D m m_0 w$$

y supone, por reducción al absurdo, que somos máquinas de Turing, por la Tesis de Church-Turing:

$$\forall p \exists m p = m$$

de esto se sigue que:

$$\forall p \exists w \neg \Diamond D p m_0 w$$

sin embargo, dado que existen seres humanos:

$$\forall p \forall w \Diamond D p m_0 w$$

y por tanto:

$$\exists m \forall w \Diamond D m m_0 w$$

lo cual es contradictorio. Por tanto, la suposición de que somos máquinas debe rechazarse porque implica una contradicción; no obstante, continúan Bringsjord y Arkoudas, dado que sabemos que los humanos somos, entre otras cosas, seres que procesan información (y que de hecho hacemos las veces de máquinas de Turing⁴⁹), se sigue que debemos ser más fuertes que las máquinas de Turing, lo cual, por vía negativa,⁵⁰ nos deja una sola alternativa: que somos hipermáquinas.

⁴⁸ LEWIS, H.R., PAPADIMITROU, C.H., *Elements of the Theory of Computation*, Michigan: Prentice Hall, 1998, p. 278.

⁴⁹ "A man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal machine". TURING, A.M. "Intelligent Machinery", p. 9.

⁵⁰ Cuando el Pseudo-Dionisio explica que algunos de los nombres divinos son negaciones y que tales negaciones son más propias de Dios que las afirmaciones, distingue a las negaciones de las privaciones. En la discusión Lucas vs Bringsjord ocurre un fenómeno similar. Cf. "La Théologie Mystique" en *Pseudo-Denys l'Aréopagite. Oeuvres complètes. Traduction, préface et note*, Paris: Aubier, 1980.

3.3. Balance de los argumentos

Aunque parece que la conclusión del argumento de Bringsjord y Arkoudas coincide con la respuesta negativa de Lucas y Penrose, es falso que se correspondan. Michael Arbib lo ha dejado claro:

Gödel's theorems limits a human as much as a machine; if a human closes his mind to new information, then there is much he cannot know. He can know much more if he continually modifies his ideas on the basis of comparing deductions from his own ideas with observations of the real world. So, any really adequate cybernetic theory of thought must contain a full understanding of growing inductive machines that continually change their internal structure as a result of interacting with the external environment.⁵¹

El punto de Bringsjord y Arkoudas, como el de Arbib, es que los argumentos gödelianos al estilo de Lucas y Penrose no sólo no funcionan porque no son casos legítimos de S3 (S4) irreducibles a S1 (S2), sino que, inadvertidamente, ofrecen una respuesta positiva al problema sobre el estatus computacional de las personas humanas como entidades computacionales *sui generis*.

Por tanto, en nuestra opinión, el argumento de Bringsjord y Arkoudas resulta ganador por dos razones: una de forma y otra de contenido. En términos de forma, porque cancela la conclusión del argumento gödeliano de Lucas (y Penrose) mostrando su invalidez,⁵² y en términos de contenido, porque provee una respuesta a la pregunta por nuestro estatus computacional por vía negativa, aceptando el movimiento argumentativo de Lucas: si, como quiere Lucas, no somos máquinas, no se sigue que *no somos* máquinas: se sigue que somos hipermáquinas o supramáquinas y, como tales, computamos funciones que una máquina de Turing no puede resolver, como el mismo problema del paro, pues podemos simular programas para un número infinito de pasos, y decidir si los programas se detienen. Ejemplos de estos sistemas los podemos encontrar en las Máquinas de Zeus,⁵³ las Máquinas Oráculo,⁵⁴ que el mismo Turing había vislumbrado, y en las Máquinas de Turing Inductivas⁵⁵ como a las que se refiere Arbib.

⁵¹ ARBIB, M., *Minds, machines and mathematic*, Springer, 1987.

⁵² BRINGSJORD, S., XIAO, H., "A refutation of Penrose's Gödelian case against artificial intelligence" en *J. Exp. Theoret. Artificial Intelligence* 12, 2000, 307-329.

⁵³ BOLOS, G.S., JEFFREY, R.C., *Computability and Logic*, Cambridge: Cambridge University Press, 1989.

⁵⁴ VAN MELKEBEEK, D., *Randomness and Completeness in Computational Complexity*, LNCS, 1950, Springer, 2000, p. 43.

⁵⁵ BURGIN, M., "Inductive Turing Machines", en *Notices of the Academy of Sciences of the USSR*, v. 270, No. 6, 1983, pp. 1289-1293.

Conclusión

Una de las preguntas más interesantes que la IA y la filosofía de la computación⁵⁶ han hecho nacer es la pregunta por nuestro estatus computacional como personas humanas: ¿Somos (reducibles a) máquinas? Después de la revisión del concepto de IA bajo la Tesis de Church-Turing y la HFSH tendríamos que responder, siguiendo la solución de Bringsjord y Arkoudas, que *no lo somos*; pero tampoco podríamos afirmar que *no lo somos*, pues en esta solución el sentido de las negaciones no es la privación. Por tanto, entraríamos en el contexto de una tradición apofática en la que nos quedaría una alternativa por vía negativa: sostener que somos *hipermáquinas* o *supramáquinas*.

Bajo esta argumentación, los versos de San Juan de la Cruz con los que hemos iniciado este trabajo resuenan con nuevos matices y dan luz sobre nuestro estatus computacional.

⁵⁶FLORIDI, L., *Philosophy and computing. An introduction*, New York: Routledge, 1999.