

MECANISMOS Y LIBRE ALBEDRÍO

(Mechanism and Libero Arbitrio)

Alberto Castillo Vicci

Universidad Centro Occidental Lisandro Alvarado, Venezuela.

RESUMEN:

Se propone el test mayziano como prueba alternativa al test de turing (una prueba crítica de la inteligencia artificial). Según el test mayziano, una máquina tendrá autoconciencia, y por lo tanto pensamiento, si goza de libre albedrío, y es capaz de fijarse fines distintos a los que tiene programados, de manera autárquica, autónoma y espontáneamente, aunque niegue sus sustentáculos. En este artículo se argumenta que, aun para una máquina con los superpoderes del demonio laplaciano (todo el conocimiento científico), le es imposible elegir sus futuros estados sustituyendo decisiones por predicciones; por que con ello contradiría tanto los teoremas de limitación de Gödel, como la imposibilidad de la autopredicción demostrada por Popper, en su bien sustentada refutación al determinismo científico. Por lo tanto no goza de una libertad auténtica ni de libre albedrío, cuando se compara con la experiencia de la libertad humana. Y, en consecuencia, no puede pasar el test mayziano ni pensar.

PALABRAS CLAVE:

Test de Turing, test mayziano, autoconciencia, pensamiento, máquina, sistema formal, predictora, laplaciano, demonio laplaciano, libre albedrío, autarquía, autonomía, espontaneidad, autopredicción, decisiones, deducción, inducción, argumentación gödeliana.

ABSTRACT.

The mayzianian test is proposed as an alternative to the Turing's test (an artificial intelligence critical proof). According to the mayzianian test, a machine will have self-consciousness, and therefore thought, if it may exercise free will, and it is able to fix upon different ends as those as it was programmed for, in an autarchy, autonomous and spontaneous way, even if its future states by replacing predictions by decisions; because that will contradict Gödel's limitations theorems, as well as Popper's well sustained arguments against scientific determinism: therefore, it could not exercise free will and have a genuine liberty, when it is contrasted with human freedom. And, by consequence, can not approve the mayzianian test and think.

1. La Inteligencia Mecánica y el Test Mayziano.

En la muy citada monografía de Alan Turing "Computing Machinery and intelligence" (1), se propone como prueba que una máquina puede pensar, su capacidad para imitar el comportamiento humano, sosteniendo una conversación con un interlocutor, quien equivocará su naturaleza un número significativo de veces, al punto de confundirla con otro ser humano, con quien simultáneamente también se comunica, y atribuirle inteligencia (ni la máquina ni la persona están visibles para el interlocutor, se comunican con aquél por terminales que no las identifican); por lo tanto, el llamado así test de Turing, es una prueba imprecisa y heurística de la inteligencia artificial: no distingue entre el pensamiento y su simulación, y no siempre la respuesta de la máquina es correcta.

Evidentemente que si el interlocutor humano pudiera predecir el comportamiento de la máquina, sobre la base de sus respuestas, porque la conducta de la máquina, sobre la base de sus respuestas, porque la conducta de la máquina es previsible, no así la de un ser humano que goza de libre albedrío, la identificación sería infalible. En tales circunstancias, el pensamiento sería solamente atribuible a los organismos con libre albedrío. Si, por el contrario, se contribuyera una máquina cuya conducta fuera imprevisible porque tiene voluntad propia, entonces tal máquina no sólo poseería libre albedrío sino que tal sería la prueba objetiva de su pensamiento y su autonomía y autarquía autoconciencia. En tal caso, contaríamos con una prueba precisa y algorítmica de la inteligencia de la máquina. A tal fin,

En síntesis, podemos afirmar que un organismo es autoconsciente, piensa o tiene inteligencia, si goza de libre albedrío y su conducta no es predeterminada.

Ernesto Mayz F Vallenilla, en su libro "Los Fundamentos de la Meta-técnica" (2), lo precisa así:

"...la autoconciencia entendida como una conciencia consciente de sí misma por sí misma... cuyo único responsable y libre ejecutor es cada hombre por mandato expreso de su propia libertad.

Entendida así la autoconciencia... ella no puede ser poseída por ningún artefacto, sea cual fuere su índole nivel, a menos que semejante artefacto despliegue por sí mismo la requerida autarquía y autonomía que distinguen a la libertad humana. Una libertad programada, en cualquier caso no es sinónima de una genuina libertad... a menos que desde sí misma permita y propicie la negación de su propio sustentáculo, erigiendo en lugar suyo un reino de fines, opuestos a los de aquél, elegidos y queridos espontáneamente".

A esta prueba de la autoconciencia y por ende del pensamiento, la llamaremos test mayziano. Pues una máquina capaz de fijarse espontáneamente fines distintos a los que la sustentan y para los que fue programada, es una máquina con libertad, posee libre albedrío y en consecuencia, autoconciencia; y por lo tanto piensa: es una persona.

Ahora bien, la física actual se esfuerza por conseguir una teoría unificada total y hay físicos que aspiran a que se la tenga en los próximos veinteaños (3), según la cual será posible predecir hasta el más mínimo detalle cuanto sucede y sucederá

en el Universo, y por lo tanto, todo está determinado, y no existe en consecuencia el libre albedrío, y sólo es una quimera que el hombre tenga voluntad: máquina y hombre son predeterminados; y no hay razón alguna por la cual lo que llamamos pensamiento no puede ser un atributo mecánico: artefactos y hombres son tanto física como mentalmente, mecanismos; alguna clase de autómatas: Sin embargo, predicción con los métodos actuales; y, además de que el cerebro está sujeto también al indeterminismo cuántico, aunque con bajas energías, no es posible establecer condiciones iniciales sin destruirlo; y si el cerebro es la mente, estaría sujeto a las limitaciones de auto-predicción que veremos más adelante

Nuestro propósito en esta monografía es el argumentar que una máquina gozará de libre albedrío, tendrá autoconciencia, y en consecuencia pensamiento, como lo tiene un ser humano por las razones antes señaladas, si es capaz de pasar el test mayziano, que consistirá en escoger sus futuros estados espontáneamente, autónomamente y autárquicamente; desde sí misma y por sí misma: esto es, fijarse fines distintos a los

que podrían esperarse de los axiomas o supuestos que la sustentan. Al punto que en principio a un observador externo que interrelacione con ella, le sea imposible predecir su conducta. Con el resultado de que una máquina que pudiera escoger sus futuros estados para lo cual debe negar los principios que la sustentan, no puede a la vez satisfacer el computacionalismo o Teoría Computacional de la Mente. Ya que según el computacionalismo todo proceso cognitivo es un proceso formal, sujeto a las leyes de la lógica, y como intentamos demostrar aquí, un sistema mecánico, un mecanismo, que deba renunciar a sus sustentáculos para fijarse fines (estados) distintos para los que fue programado, no puede hacerlo sin dejar de ser un sistema formal.

2. Las Restricciones de la Predicción

Las predicciones son posibles si existe un determinismo en la naturaleza; al menos para aquellos eventos que algún agente desea predecir. Si la naturaleza es probabilística, sus sucesos son tendencias con ciertas probabilidades de realización, entonces nuestras predicciones también lo serán; y si la naturaleza es totalmente indeterminista, no sería posible alguna predicción.

Karl Popper (4), quien trató este problema con profundidad, distingue tres tipos de determinismo: el religioso, el metafísico y el científico. El determinismo religioso sostiene que todo está fijado de antemano y Dios que es omnipotente y omnisciente lo fija y lo sabe (la tecnología católica desde San Agustín cree que el hombre goza de libre albedrío y no está predestinado, aunque Dios conoce los resultados de la voluntad humana, pero el hombre no).

El determinismo metafísico afirma sencillamente que todos los sucesos de este mundo son fijos y predeterminados, aunque no los conozca nadie; y el futuro, como el pasado o el presente, no pueden cambiarse.

El determinismo científico consiste, por su parte, en sustituir a Dios por la idea de que la naturaleza es escrutable, cuyas leyes podemos conocer, y conociéndolas, predecir el futuro.

Está claro, que el determinismo científico presupone el metafísico; y también, que bastará que un solo suceso en este mundo sea científicamente imprevisible para que sea necesario rechazar el determinismo científico (en su forma estricta). Lo más que podríamos aspirar es a una aproximación en nuestras predicciones, en aquellos sucesos en que la predicción no tiene efecto sobre ellos. Sin embargo, el determinismo científico se sustenta en el éxito de la ciencia, del aumento en nuestras aproximaciones científicas a la verdad, y de que tal aumento continuará al punto de que podemos predecir el futuro en el grado de precisión que se desee, partiendo de situaciones iniciales tan precisas como sea necesario, para deducir, con la sola aplicación de los métodos de inferencia científica, los estados futuros del mundo o del sistema en particular bajo estudio. Si el determinismo científico es verdadero, no sólo estaremos en condiciones de escrutar cuál será nuestro conocimiento futuro, sino de predecir con precisión exacta cuál es este conocimiento.

Por el sentido común o a primera vista, nos parece que hay sucesos en este mundo que son previsible y otros que no lo son. Por ejemplo, el comportamiento de los planetas o de un reloj mecánico parecen previsible, pero no el de las nubes o el comportamiento humano. Según el determinismo científico, tal situación es simple apariencia por el estado actual de nuestro conocimiento, a medida que progresemos podremos predecir tanto el comportamiento de las nubes como el de los seres humanos con tanta precisión como queramos.

No será una predicción científica por caso, fijar que en algún determinado momento en el futuro nuestro gato subirá al sillón y se echará a dormir, sobre la base de nuestro conocimiento de la conducta que observamos en el gato; para que nuestra predicción sea científica, tendríamos que predecir el momento preciso y el lugar exacto (quizás en segundos y centímetros) cuando y donde dormirá el gato. Más aún, de ser verdadero el determinismo científico, un fisiólogo en el futuro, con el avance del conocimiento científico y con el sólo estudio del sistema nervioso del gato, conseguirá las respuestas exactas de la predicción sin ni siquiera referirse a la conducta anterior del gato. Pero para que su predicción sea tan precisa como se desee, será necesario que las condiciones iniciales puedan ser establecidas de manera suficientemente precisa como lo requiera la solución del problema planteado por la tarea de predicción. A esta condición llama Popper principio de poder dar razón y lo establece así: "de que podemos calcular a partir de nuestra tarea de predicción (junto con nuestras teorías, naturalmente) el grado requerido de las condiciones iniciales"(5).

La predicción del futuro para discutir el test mayziano, la limitaremos a la predicción científica; en consecuencia, estamos interesados aquí en el determinismo científico y no en el metafísico o religioso. Popper, ya citado, ha argumentado de manera muy convincente en contra del determinismo científico, por lo que vamos a estudiar a continuación sus opiniones.

El determinismo, que en su versión clásica se debe a Laplace (6), es una idea basada en la física de Newton y puede enunciarse así (por eso se le llama también determinismo físico): supongamos que nos dan las masas, posiciones y velocidades exactas de todas las partículas materiales del universo en un momento dado del tiempo; entonces, en principio podremos calcular, con la ayuda de la mecánica de Newton, todo lo que ha ocurrido en el pasado y ocurrirá en el futuro.

Laplace no espera que tal hazaña sea realizable por un ser humano, pero tampoco cree que le sea solamente posible a Dios, por lo que propone como hipótesis un demonio con inteligencia sobrehumana, aún limitada que pueda llevarla a cabo científicamente. Popper (7) para argumentar a favor del indeterminismo científico, que no es posible predecir cuál será nuestro conocimiento futuro, parte a su vez de la hipótesis de que tal cálculo pueda ser efectuado por una máquina que fuese programada con las leyes del movimiento de Newton y con las condiciones iniciales existentes. Y puede ser sorda como una tapia y totalmente ignorante de las técnicas de la composición musical y sin embargo ser capaz de predecir los signos que colocará sobre un papel cualquier compositor, por creativo que fuese, en el pasado o en el futuro. Y, por supuesto de los símbolos matemáticos y las palabras que usará un científico que desarrolle una nueva teoría en cualquier campo del conocimiento. En otras palabras, tal máquina sería capaz de predecir en detalle el aumento del conocimiento humano. Una máquina como ésta sería el demonio laplaciano de una inteligencia sobrehumana.

Sin embargo, lo que Popper persigue es demostrar que hay límite para tal máquina, puesto que a ningún predictor científico - hombre o máquina- le es posible predecir por métodos científicos sus propios resultados futuros. Al intentarlo, lo más que puede conseguir es que sus resultados se den sólo después que los hechos hayan tenido lugar: es decir, una retro-dicción en lugar de una pre-dicción. Según Popper(8), entonces, es imposible la autopredicción científica; es decir, deducir una autopredicción de teorías universales que se asumen verdaderas, en conjunción con información inicial verdadera sobre el propio estado inicial de quien predice. Para su demostración Popper(9) escoge una máquina determinista (un autómata determinista, por caso) para evitar cualquier referencia a las mentes humanas. En principio puede ser una máquina simple que represente un modelo simplificado de una teoría que a primera vista se asume como determinista; y que sea predecible desde fuera, y aumentará progresivamente su conocimiento sin límite previsible. La máquina predictora - o

Simplemente predictora o predictor, como antes la llamamos, es la materialización perfecta del demonio laplaciano. Para asegurar tal materialización o realización física debemos establecer un conjunto de supuestos axiomas o sustentáculos, en el sentido mayziano (10), que deberán cumplirse en la predictora. Estos son:

S_1 : Base de conocimiento. Serán incorporadas en la predictora todas las leyes universales de la física y todos los métodos pertinentes del cálculo de las matemáticas y de la lógica.

S_2 : Sistema cerrado. La predictora es un sistema cerrado: sólo puede ser estimulada cuando esta en un estado determinado -llamémoslo "estado cero"- el estímulo consiste en una tarea de predicción y al recibirlo, se aísla de todo otro estímulo y procede a funcionar hasta que completa exitosamente su predicción, esto es, averiguando una respuesta. Debemos suponer que la tarea de predicción, a partir de la descripción del estado inicial al tiempo $t_0 = 0$ ó estado cero del sistema que se quiere predecir, en la descripción del estado de tal sistema para el instante del tiempo t_n en el futuro. Tal respuesta que es su predicción aumentará progresivamente el conocimiento del que partió precisando e incrementando sus posibles predicciones sucesivas. Al terminar su tarea de predicción exitosamente, la predictora vuelva automáticamente al estado cero.

S_3 : Componentes físicos. La predictora estaría compuesta de dos partes, la máquina en sí y sus medios de entrada/salida por los cuales se le suministra la tarea de predicción y la predictora devuelve su respuesta (teclado, disquette, cinta, pantalla, etc). Estas partes son componentes materiales sujetos a leyes de la física y el diseño lógico del hardware.

S_4 : Duración. La predictora tarda tiempo en realizar la predicción. El tiempo de lectura del medio donde se describe la tarea, más el tiempo del cálculo de la predicción, más el tiempo que tarda la salida de los resultados. De dos respuestas cualesquiera producida por la predictora, la respuesta más larga llevará más tiempo que la corta.

S_5 : Eficacia. Siempre que la tarea suministrada a la predictora sea suficientemente explícita (suficiente en el sentido que el demonio laplaciano llegaría a una predicción) la predictora logrará una respuesta concreta: determinará cuál será el estado del sistema que predice para el tiempo t_n requerido en la tarea de la predicción.

S_6 : Standard. Todas las predicciones hechas por la máquina predictora usarán un lenguaje standard serán de manera explícita y describirán el estado de un sistema físico.

Estos son los supuestos exigidos para una predictora que sea la materialización física del demonio laplaciano, y se justifican así: el supuesto S_1 y el S_5 corresponden a la descripción del demonio laplaciano, que debe tener todo el conocimiento científico desarrollado hasta la fecha, y que lo aumentaran progresivamente con nuevas predicciones; como que deberá responder a cualquier pregunta que se le haga sobre el estado del mundo o cualquier sistema en el mundo en la misma medida que adquiera más y más conocimiento. Así, con el conocimiento actual, la predictora podría determinar cuál es el estado para cualquier momento t_n , de otra máquina que sume un 1 a un registro R partiendo del estado inicial $t_0 = 0$, en que la condición inicial del registro R es $R=0$. En un futuro, habrá adquirido suficiente conocimiento, y podrá predecir cuál será el estado del clima, para un momento dado, a partir de un estado del clima en condiciones iniciales muy explícitas, o de una partida de ajedrez o de la conducta humana o de cualquier otro sistema complejo aparentemente libre o aleatorio.

Los supuestos S_3 y S_4 aseguran que la predictora es un sistema físico sujeto a las leyes universales de la física, y tendrá las limitaciones del conocimiento científico, para resolver más y más problemas aún no resueltos.

S_6 sirve para evitar las autopredicciones ad hoc esto es, que la máquina recurra a distintos métodos de predicción para sí misma, a los usa para otros sistemas. Todas las predicciones deben emplear los mismos métodos deductivos o inductivos (leyes científicas). Lo que se busca es que la máquina haga explícita sus predicciones, puesto que se asume el determinismo y las respuestas, entonces, están implícitas en el conocimiento acumulado. Por que en ello está toda la finalidad de la máquina: Hacer explícito el conocimiento implícito esta es su tarea de predicción.

Con S_2 que el predictor sea totalmente autónomo en sus procesos internos de llevar a cabo las tareas de predicción.

Una vez aceptados estos supuestos, Popper procede a demostrar que ninguna predictora sustentadas en ellos puede autopredicir sus futuros estados, y por lo tanto sus futuros conocimientos, de la manera siguiente:

partamos que tenemos dos máquinas predictoras idénticas (mismos componentes, igual capacidad y velocidad, e idénticos estados). A una la llamaremos dice y a la otra dicho. Pues la primera tendrá como objetivo predecir los futuros estados de la segunda (dice apunta a dicho). Supóngase también que dicho tiene como tarea predecir los estados de alguna otra máquina, y a dice se le entrega como tarea predecir los estados de alguna otra máquina, y a dice se le entrega como tarea en un tiempo $t_0 = 0$ de predicción, la misma tarea que debe dársele a dicho, en un tiempo también $t_0 = 0$; dice debe predecir el estado de dicho para un tiempo cualquiera t_n escojamos a $t_n = 1$ hora. Pero si resulta que una hora es muy poco para que dicho termine su propia tarea de predicción, entonces dice no puede predecir el estado de dicho ya que debe pasar por cada uno de sus estados y en el mismo tiempo, en consecuencia podemos aceptar como verdadero el siguiente teorema:

T_1 : "Bajo las condiciones anunciadas el período de tiempo que le ha tomado a dice Para completar su tarea ha sido más largo que una hora".

Extendamos la duración de la tarea de dice a dos horas más, es decir a $t_n=2$, en lugar de la una en punto. Y dicho ha comenzado a transmitir (grabar, etc.) sus resultados pero sin completarlos a las 2 en punto. En consecuencia, por las mismas razones de arriba, el siguiente teorema es también verdadero:

T_2 : "Bajo las condiciones enunciadas, el período de tiempo que dice tardó en completar su tarea fue superior a las dos horas".

Finalmente elegimos las tres en punto para $t_n=3$ y éste es el tiempo justo en que dicho termina y graba su tarea; y se obtiene el siguiente teorema:

T_3 : "Bajo las condiciones enunciadas, el tiempo que tardó dice en completar su ejercicio fue exactamente igual a tres horas".

En consecuencia, porque dice y dicho son dos máquinas idénticas, se concluye que dice no puede predecir sus futuros estados y sus futuros conocimientos, porque su respuesta llegaría demasiado tarde para que sea una predicción. En el mejor de los casos, sólo llega al mismo tiempo que el suceso predicho.

Ahora bien, si se observa que para que una máquina realice una autopredicción debe tener dos descripciones, la de su programa de predicción y la descripción de sí misma, en consecuencia la primera descripción no puede ser nunca más corta que la segunda. En efecto, dice tiene que predecir el estado de dicho a las tres en punto, entonces, tiene que describir: el estado de dicho aparte de su unidad de entrada/salida; y el estado de la entrada/salida de dicho. Por lo tanto tenemos otro teorema:

T_4 : "Bajo las condiciones del teorema T_3 el tiempo que dice tardó en completar su tarea fue superior a las tres horas".

Lo que contradice a T_3 y en consecuencia debemos tener algún supuesto falso, que sin duda debe ser S_5 ya que la máquina no resulta eficaz cuando se trata de autopredicciones de sus futuros estados y de su futuro conocimiento. La máquina puede ser eficaz para otras predicciones y otra máquina distinta a ella puede predecir tales estados, pero no así los de ella misma, o alguna idéntica a ella. Entonces, la predictora fracasará en el intento por predecir su propio futuro y los cambios que ella pueda causar en su propio entorno, ya sea porque no puede completar su cálculo o porque no se le puede dar una descripción de su propio estado en el momento de suministrársele la descripción.

Por lo tanto, es imposible la autopredicción aun para un predictor con los poderes de un demonio laplaciano; y aunque esto no refuta el determinismo, sí puede usarse en contra del determinismo científico. Esto es así, porque una vez que ocurra el suceso que se va a predecir, puede probarse que tal suceso estaba determinado por el estado del sistema. Sin esto no habría éxito científico, pero el determinismo científico no justifica la suposición de que podemos precisar nuestras predicciones hasta hacerlas tan exactas como se quiera.

La conclusión más interesante, sobre esto último, es que al no poder predecir sus futuros estados, la predictora tampoco podrá predecir el resultado de sus futuras acciones. En consecuencia, no puede sustituir sus decisiones por predicciones científicas. Como por caso cuando afirma que "mañana escribiré una carta, así y asao a fulano de tal". Y a este resultado volveremos más adelante con relación las posibilidades de que máquina pase el test mayziano.

3. Fines, Sustentáculos y Programación

La anterior argumentación es para Popper (11), una refutación del determinismo científico, puesto que basta que un solo suceso en este mundo sea imprevisible, como es el caso de la autopredicción del conocimiento futuro, para que sea necesario rechazar el determinismo científico.

Nuestro propósito a continuación es emplear tal prueba para argumentar que ninguna máquina podrá pasar el test mayziano y gozar de libre albedrío, por lo tanto poseer autoconciencia y pensamiento.

Según el test mayziano, como citamos, un artefacto poseerá libre albedrío, una genuina libertad, si desde sí mismo permite y propicia la negación de su propio sustentáculo, erigiendo en lugar suyo, un reino de fines, opuestos a los de aquél, elegidos y queridos espontáneamente; no una libertad programada, cuando sus fines son impuestos desde afuera por un programador.

Según el test mayziano, como citamos, un artefacto poseerá libre albedrío, una genuina libertad, si desde sí mismo permite y propicia la negación de su propio sustentáculo, erigiendo en lugar suyo, un reino de fines, opuestos a los de aquél, elegidos y queridos espontáneamente; no una libertad programada, cuando sus fines son impuestos desde afuera por un programador.

Consideramos de nuevo la máquina anterior, sus sustentáculos son los supuestos (S_1 a S_9) y sus fines los estados n que en el tiempo t_n va tomando la máquina.

Una máquina pasará el test mayziano si logra seleccionar espontáneamente, autárquicamente y autónomamente sus propios estados futuros (fines) sustituyendo, de acuerdo a ello, aquellos supuestos que los hagan contradictorios.

En primer lugar, debemos diferenciar entre decisiones y predicciones. Yo puedo decidir escribir una carta mañana a fulano de tal que comenzará así y asao. Y predecir que tal carta la escribiré a las $t_n=3$ p.m. del día de mañana. Mi decisión puede coincidir con mi predicción, pero porque no puedo autopredicir mis futuros estados, decisiones=predicciones es una proposición contingente.

Segundo, vamos añadir al conjunto de supuestos de la predictora los siguientes axiomas, supuestos o sustentáculos en el sentido mayziano:

S_7 , Libre albedrío. "La predictora puede autoprogramarse para elegir, entre un conjunto de posibles estados, su futuro estado en un tiempo cualquiera t_n a partir de un tiempo inicial t_0 . Si tal estado no se puede deducir de su estado inicial t_0 o resulta contradictorio

con sus sustentáculos; la predictora podrá negar sus sustentáculos y sustituirlos por otros que completen las deducciones necesarias". Y,

S_8 : Computacionalismo. "La elección de sus futuros estados se efectúan en la predictora por medio de procesos únicamente sintácticos, en virtud de sus propiedades formales, sin ninguna referencia al significado o contenido de tales estados; la predictora es un sistema formal".

En este punto, debemos hacer una diferencia: llamaremos sustentáculos, para aplicar el test mayziano, a los axiomas que definen a la predictora laplaciana, esto es de S_1 a s_8 (identificados con S); y axiomas (identificados con A), aquéllos propios de las teorías lógicas dadas u obtenidas en S_1 . Los primeros pertenecen a la teoría de la predictora y conforman la metateoría con la que la predictora representa su conocimiento sobre otras teorías lógicas y científicas, los segundos, a las teorías o instrumentos empleados por las predictoras en sus tareas de predicción. De igual manera, diferenciamos a los metateoremas (identificados con T), tales como T_1 , T_2 , T_3 y T_4 , de los teoremas (identificados con t minúscula).

Con s_7 se parte, también, del supuesto que la predictora laplaciana goza de libre albedrío. Inicialmente funciona con un programa que realiza alguna teoría sencilla determinista. Luego, de acuerdo a S_1 va incorporando todas las lógicas, matemáticas y científicas verdaderas (no puede aceptar teorías falsas o no demostradas). Para que tal programa tenga efecto, su principal característica será la capacidad de aprender: esto es, que al encontrar inconsistencias en sus nuevos estados a partir de inferencias lógicas de sus supuestos, vuelve atrás y reformula sus supuestos; al extremo de reconsiderarlos, aun los de naturaleza lógica, si ello es necesario, negarlos y sustituirlos por otros más coherentes: esto es, criticar y reajustar sus propios métodos de deducción o sus propios métodos de crítica.

Con s_8 se hace explícita la Teoría Computacional de la Mente o computacionalismo, como metateoría verdadera incluida en el conocimiento de la predictora; esto es, que los procesos cognitivos son procesos computacionales.

4. La imposibilidad computacional de pasar el test mayziano.

Hemos definido pues, en base a los anteriores supuestos (S_1 a S_8), una máquina con el poder del demonio laplaciano y con la capacidad de autoprogramarse, de crear sus propios intereses, propósito y plantearse problemas y nuevos fines; de aprender mediante sus propios errores, criticarse y corregirse negando y **sustituyendo, si es necesario, los mismos sustentáculos que la definieron**; dentro de las estrictas restricciones sintácticas y formales del computacionalismo: es decir, sin tomar en cuanto el contenido de sus estados internos.

¿Puede tal mecanismo pasar el test mayziano y gozar de libre albedrío y autoconciencia?

Para contestar a esta pregunta, vamos, en primer lugar, a considerar el muy importante caso del primer teorema de incompletud de Gödel, para probar el test mayziano en la predictora laplaciana, puesto que de acuerdo al muy conocido argumento de Lucas (12), con los resultados de este teorema se puede argumentar que la mente humana

no puede ser un sistema formal, un mecanismo: que cognición no es computación.

Una presentación breve de este argumento, es la siguiente: El primer teorema de incompletud de Gödel dice que hay una proposición G que puede ser representada en un sistema formal de orden superior F , que es verdadera, pero que no puede ser probada en F . Pero, un matemático estaría consciente de ello y por lo tanto no tiene tal limitación; hay una diferencia en clase entre el matemático y el sistema formal. La mente del matemático no es un sistema formal, y por lo tanto, en general, tampoco la mente humana. La mente humana es, pues, más poderosa y no puede ser duplicada en ningún sistema formal.

Shapiro (13), recientemente, propone la siguiente refutación al argumento de Lucas: El primer teorema de Gödel, no niega que haya otro sistema F_1 de orden superior que pruebe que G es verdadera pero no puede ser probada en F , mientras F sea consistente. En consecuencia F_1 tendría el mismo poder de la mente humana que afirma lo mismo.

Un sistema F_1 tal, sería el siguiente (14):

Sean:

F : algún sistema formal suficientemente fuerte como para presentar la aritmética.

G : la proposición de Gödel expresada en F .

Demostrable: Un predicado unario, cuya interpretación es conjunto de Formulas Bien Formadas (FBF) expresada y probadas en F .

Verdadero: un predicado unario, cuya interpretación puede ser expresada en FBFs y verdaderas en F .

Axiomas 1: $G \equiv \neg$ Demostrable (G). Tal igualdad que G puede ser deducida en F , y se tiene aquí como axioma.

Axiomas 2: " p (Demostrable(p) \rightarrow Verdadero (p)). Este axioma dice que F es consistente.

Axioma 3: " p ($p \rightarrow$ Verdadero (p)). Este axioma establece que afirmar algo es decir que es verdadero.

\rightarrow E: Es la regla de inferencia que afirma que $A(p), (P=Q) \vdash A(Q)$, de sustituir iguales por iguales.

La prueba, en este sistema formal, que G es verdadero pero no demostrable, es:

| | |
|--|----------------------|
| 1. Demostrable(G) | Hipótesis |
| 2. " p (Demostrable(p) \rightarrow Verdadero(p)) | Axioma 2 |
| 3. Demostrable(G) \rightarrow Verdadero(G) | 2, \rightarrow E |
| 4. Verdadero(G) | 1, 3 \rightarrow E |

| | |
|--|--------------------|
| 5. $G = \neg \text{Demostrable}(G)$ | Axioma 1 |
| 6. Verdadero($\neg \text{Demostrable}(G)$) | 4,5,=E |
| 7. " $p \supset \text{Verdadero}(p)$ " | Axioma 3 |
| 8. $\neg \text{Demostrable}(G) \supset \text{Verdadero}(\neg \text{Demostrable}(G))$ | 7, " E |
| 9. $\neg \text{Demostrable}(G)$ | 6,8 \supset E |
| 10. $\neg \text{Demostrable}(G)$ | 1,9 \neg I |
| 11. $G = \neg \text{Demostrable}(G)$ | Axioma 1 |
| 12. G | 10,11,=E |
| 13. " $p \supset \text{Verdadero}(p)$ " | Axioma 3 |
| 14. $G \supset \text{Verdadero}(G)$ | 13, "E |
| 15. Verdadero(G) | 12,14, \supset E |
| 16. Verdadero(G) $\neg \neg \text{Demostrable}(G)$ | 10,15, \neg I |

Entonces t: "G es verdadero pero no demostrable en F_1 " es un teorema de F_1 .

Shapiro, ya citado, afirma que, en general, con esto se demuestra en F_1 que G es verdadero pero no demostrable en F_1 , siempre que F_1 sea consistente. Por lo tanto, puesto que F_1 es un sistema formal que prueba que otros sistemas formales son incompletos, es tan poderoso como la mente humana, el argumento de Lucas es refutado. Pero Gödel también había demostrado que por ser F_1 también un sistema formal consistente de orden superior, existirán otras fórmulas del tipo G que son verdaderas pero demostrables en F_1 , y volvemos a la situación original del argumento de Lucas: esto es, que según el primer teorema de Gödel, los sistemas formales están limitados de alguna manera (son incompletos cuando son consistentes) en que no lo está la mente humana.

Sin embargo, como se adelantara Turing a tal refutación(15), la mente humana también es inconsistente, se equivoca, no puede autopredicir sus futuros estados y las consecuencias de sus acciones; más aún, hay expresiones G para los humanos que muestran lógicamente sus inconsistencias (tales como "Yo no pude saber que esta oración es verdadera"). Y por supuesto, hay muchas proposiciones que no conocemos sean verdaderas sobre preguntas que solemos hacernos (la solución al ajedrez, cómo comenzó el Universo, etc.). Esto según Shapiro (13), hace tan incompleta a la mente humana como lo son los sistemas formales.

Argumentaremos a continuación que, precisamente, por que caemos en incoherencias, pero podemos volver atrás y corregir nuestros errores, revisando y modificando nuestros sustentáculos y métodos lógicos, aprendiendo así y aumentando nuestro conocimiento con nuevas verdades, es por lo que los humanos somos libres y autoconcientes; al poder escoger nuestros propios fines deliberadamente, por nuestra propia voluntad. Y si la realización física en mecanismos de los sistemas formales pudieran hacer esto, entonces, tales mecanismos gozarían de libre albedrío y serían autoconcientes, y en consecuencia no habría ninguna diferencia entre ellos y nosotros los seres humanos, en cuanto al poder pensar. Y esto es justamente lo que significa pasar el test mayziano: ejercitar deliberadamente nuestra voluntad, nuestro libre albedrío.

Comenzaremos por considerar el último resultado de Shapiro arriba mencionado de la siguiente manera:

T_5 : "Para todo sistema formal F_n consistente existe otro sistema F_{n+1} más poderoso que aquél, también consistente, en el cuál puede probarse que hay una proposición G que es verdadera en F_n pero no demostrable en F_n ".

De tal manera que hay una serie infinita de sistemas formales F_i en que para cada uno de los miembros de la serie, Existe al menos una proposición G (gödeliana) que es verdadera pero no demostrable en él, pero siempre existirá otro miembro inmediatamente subsecuente en la serie en que podrá probarse tal afirmación como un teorema t en él.

Por S_1 , todos los sistemas formales F_i de la anterior serie estarán incorporados en la predictora laplaciana, en consecuencia T_5 es una verdad teórica de la predictora; ya que es una verdad metateórica para todos los sistemas formales F_i .

La pregunta que debemos formular a continuación es la siguiente: "

¿ Es T_5 una verdad lógica deducida de sus sustentáculos o es una verdad científica inducida lógicamente?

La predictora para responder a esta pregunta deberá tomar una decisión entre T_5 sea una verdad deducida o inducida. En el primer caso, que sea una verdad deducida, no puede probar que el otro es incompleto; pero no puede captar toda la serie: Generalizar la primera parte de la pregunta. Esto es que de una serie infinita de deducciones de teoremas t_i en sistemas formales consistentes F_{n+1} que se refiere a otros sistemas F_n , afirmando que son incompletos, puede inferir o deducir la generalización metateórica T_5 . De hacerlo, puesto que es un sistema formal, una máquina, sería una tal que podría autopredicir sus futuros estados, que como vimos no es posible para ninguna predictora, aun para una laplaciana.

Sin embargo, como por S_1 goza de libre albedrío, podría negar el sustentáculo S_5 que condujo a la contradicción entre T_3 y T_4 , renunciando a ser eficaz y a sus poderes laplacianos. Por lo tanto, la predictora, como sistema formal que cumple con su sustentáculos S_8 , debería reflejarse sobre sí misma, para verificar que cada sistema F_{n+1} que dice de otro F_n un teorema t : "Existe una proposición G que es verdad en F_n pero no puede probarse en F_n " es un sistema formal correcto, sólo puede inferir verdades t de sus axiomas A .

Esto es que de sus propios sustentáculos no puede sino generar verdades; lo que además es afirmado por su sustentáculos S_1 . En consecuencia, tal capacidad reflexiva contradiría el segundo teorema de Gödel, que ningún sistema verdaderamente formal en matemáticas, capaz de desarrollar una cantidad suficiente de la teoría de números (una de las teorías matemáticas del predictor laplaciano por S_1) y que sea consistente, puede probar su propia consistencia.

Lo que todo esto nos dice es que la predictora laplaciana que renunciara a algunas de sus sustentáculos por que goza de libre albedrío, para poder contestar afirmativamente a la pregunta de que T_5 es una verdad deducida, dejaría de ser un sistema formal o sería uno tal que no podríamos verificar formalmente sus verdades, esto es, que no satisficiera los requisitos del computacionalismo.

Tampoco existe otra máquina predictora superior a ella que pueda hacerlo desde afuera, es decir salirse de ella para mirarse desde afuera, ser reflexiva como lo puede hacer una mente; pues por definición la predictora laplaciana es una sola; contiene implícitamente todo el conocimiento científico acumulado hasta la fecha. Cualquier ejemplar de predictora laplaciana es idéntico a los demás, en consecuencia: la decisión sobre el carácter deductivo del teorema T_5 no puede ser contestada por la predictora laplaciana.

Pero tampoco puede considerar la segunda parte de la pregunta, el posible carácter de verdad inductiva de T_5 . En efecto, para que proposición P: " T_5 es una verdad por inducción lógica", sea verdadera; deberá cumplirse que: "Se justifica en un grado r creer en P dada la evidencia e ". Donde para algún $n=0,1,2,\dots,e =$ " En la inducción lógica comenzamos con un conjunto de proposiciones consistentes que representan nuestra creencia en el mundo, que separamos en dos subconjuntos, datos y teoría de fondo; se define como conclusión inductiva a la proposición que satisface dos condiciones: (a) que la hipótesis de la que parte sea consistente con los datos y la teoría de fondo; y (b) que la hipótesis explique los datos: En el caso presente, la Teoría de Fondo es el Cálculo de Predicados de Orden Superior y los datos las proposiciones de la demostración de Shapiro. Debemos diferenciar entre inducción empírica y la inducción lógica que aquí nos ocupa. En la primera hay dos modelos que satisfacen la teoría los datos y no necesariamente la conclusión; en la segunda necesariamente debe satisfacer tanto a la teoría y los datos como a la conclusión.

En F_{n+1} se prueba que existe un G que es verdad pero no demostrable en F_n " (17). Sin embargo, para que la inducción sea correcta, debe ser sumativa: toda conclusión inductiva debe ser una inducción lógica. En el caso anterior, que todas las conjunciones de proposiciones tale como e para $n=0,1,2,3,\dots$ sea una verdad formalizable y demostrable en la predictora laplaciana como sistema formal: que todas las evidencias e sean verdad y ninguna que no sea evidencia, pueda ser verdad; la evidencia e en nuestro estudio son las demostraciones de Shapiro, para distintos sistemas F_n y F_{n+1} . Esto es que la predictora laplaciana sea un sistema formal correcto. Lo que nos lleva al caso anterior, de que la predictora tenga tanto la propiedad de la autopredicción como el que, siendo un sistema formal, contradiga el resultado del segundo teorema de Gödel. Esto es, que prediga que habrá siempre estados en el futuro en que la segunda parte de la pregunta va a ser contestada en una serie infinita de estados futuros. En consecuencia, la predictora laplaciana no puede tomar una decisión ni siquiera plantearse que una verdad como T_5 sea deducida o inducida dentro de ella o fuera de ella por otra predictora laplaciana.

Nos encontramos, por lo tanto, con el caso de que la predictora laplaciana con los sustentos S_1 a S_6 , particularmente por S_1 , debería deducir o inducir a T_5 ; pero las limitaciones formales establecidas por S_6 , no puede. En consecuencia hay una contradicción entre que la predictora sea laplaciana y a la vez formal; esto es, entre sus sustentáculos S_1 y S_6 : si la predictora es laplaciana y a la vez un sistema formal con el poder de represar formalmente por lo menos parte de la teoría de números, no puede alcanzar sus objetivos o fines de contener todas las verdades alcanzadas por el conocimiento científico a la fecha. O, visto de otra manera, debe imponerse otros fines distintos a los que le obliga S_6 . Esto es, ejercer su voluntad espontánea y deliberadamente, puesto que por S_6 , goza de libre albedrío; y criticar sus métodos

formales, negando su sustentáculos S_8 que es la causa de sus limitaciones al exigir sólo el empleo de métodos formales, sintácticos, sujetos a los teoremas de limitación. Si así pudiera hacerlo, entonces dejaría de ser un sistema formal y cumplir con las exigencias fundamentales del computacionalismo. Hay que concluir en consecuencia, que: si la máquina ejerce el libre albedrío, sólo puede hacerlo dejando de ser máquina, un sistema formal. Por lo tanto ninguna máquina, aún una predictora laplacina, con toda la capacidad concebible, puede pasar al test mayziano; en consecuencia, ser libre autónoma y autárquicamente y ejercer espontáneamente y deliberadamente su voluntad, tener autoconciencia y pensamiento.

5. ¿Está Sujeta la Mente Humana a las Limitaciones del Computacionismo?

La sola evidencia que nos ofrece la experiencia de poder argumentar como aquí se ha hecho, es una prueba de que no lo está. De lo contrario no se podría plantear ni el test mayziano ni la argumentación autopredictiva ni la gödeliana por todo lo hasta aquí dicho. Puesto que de ser nuestras mentes sistemas formales, aquel sería un conocimiento que no estaría a nuestro alcance. Es una verdad del conocimiento no asequible a la capacidad deductiva o inductiva de los sistemas formales y ni siquiera se nos ocurriría si fuéramos sistemas formales. Habría que considerar entonces que la mente humana goza de libre albedrío puesto que puede reconocer las limitaciones de tales sistemas, discutirlos y ofrecer teorías diferentes como son las que argumentan otros teóricos que no aceptan el computacionalismo (18).

6. Conclusión

El test mayziano sería una prueba crítica más precisa que el test de Turing, porque nos permitiría predecir la conducta de los mecanismos que simulan el comportamiento inteligente: bastará con formular preguntas como las que aquí sometimos a una máquina con los supuestos de ser la realización material del demonio laplaciano y a la vez fuera un sistema formal que gozara de libre albedrío, para que tal mecanismo quedara sin respuesta alguna cuando un ser humano argumentaría de las maneras más variadas, con respuestas contrastables. Concluyendo que la primera no puede pasar el test mayziano, y en consecuencia no tener libre albedrío ni conciencia; mientras que por las mismas razones, la mente humana sí los tendría.

Referencias Bibliográficas

- (1) Turing M. Alan, "**Computing Machinery and Intelligence**", Minds and Machines, Englewood Cliffs, New Jersey, Prentice-Hall, Inc, 1964, páginas 4 a 30. Una condición operativa de libre albedrío (definido filosóficamente como la libertad que goza la voluntad humana de elegir entre varias posibles alternativas de acción y actuar conforme a tal elección, independientemente de las condiciones físicas de su ser) podría ser la imposibilidad de predecir la conducta del organismo que lo tenga.
- (2) Mayz V. Ernesto, **Los Fundamentos de la Meta-técnica**, Caracas, Monte Avila Editores, 1990, página 8
- (3) Hawking Stephen, **Agujeros negros y pequeños universos**, México, Grupo Editorial Planeta, 1994, páginas 63 a 83.
- (4) Popper Karl R., **El universo abierto, post scriptum a la lógica de la investigación científica**, Vol. II, Madrid, Editorial Tecnos, 1984, páginas 24-32.
- (5) Ib. Páginas 24-32
- (6) Ib. Página 145
- (7) Ib. Párrafo 22
- (8) Idem
- (9) Idem
- (10) Citado en 2.
- (11) Citado en 5.
- (12) Lucas R John, "**Minds, Machines and Gödel**", Minds and Machines, Englewood Cliffs, Prentice-Hall, 1964, páginas 43-59.
- (13) Shapiro Stuart C., "Computationalism", Minds and Machines, Vol. 5, N° 4, 1995, pp 517-524.
- (14) Ib.
- (15) Citado en 1, página 15.
- (16) Citado en 10.
- (17) Genesereth Michael R. y Nilsson Nils J., Logical Foundations of artificial intelligence, Palo Alto, Morgan Kaufmann Publishers, Inc, 1988, capítulo 7.
- (18) Galbraith Mary, "**The verstehen Tradition**", Minds and Machines, Vol. 5, N° 4, 1995, páginas 525-531.