

¿Cuán Inteligente es el Homo Sapiens?

José Hernández-Orallo[‡]

*Universidad Politécnica de Valencia,
Dpto. de Sistemas Informáticos y Computación,
C/ de Vera s/n E-46022, Valencia, España
E-mail: jorallo@dsic.upv.es*

Enero 1999

Resumen. A la clásica pregunta de cuán inteligentes son o pueden llegar a ser las máquinas, desde un punto de vista antropocéntrico, este artículo contrapone la pregunta de cuán inteligentes somos, desde un punto de vista computacional. De todos los factores que influyen en la inteligencia, vista como potencial, distinguimos el factor *g* o *inteligencia líquida*, y lo definimos como la capacidad de encontrar descripciones intensionales a secuencias finitas, o, lo que es lo mismo, de comprenderlas. La mayor aportación de este artículo es que dicha medida se presenta de una manera formal y computacional, y correla con los clásicos tests de inteligencia, lo cual permite por primera vez presentar un test no antropomórfico (a diferencia del test de Turing) utilizable tanto para computadores como para el homo sapiens sapiens.

Palabras Clave: Inteligencia, Cociente de Inteligencia (C.I.), Factor *g*, Complejidad Descripcional, Complejidad Intensional, Psicometría, Turing Test, Aprendizaje Computacional.

1. Introducción

A primera vista parece ridículo preguntarse acerca de nuestra inteligencia, o al menos hacerlo sin tomar al homo sapiens como referencia, omnipresente ya desde mucho antes del Test de Turing. Parece como si se diera por supuesto que el homo sapiens es una especie de estándar o pie de rey universal, sobre el que todo el reino animal, supuestos seres extraterrestres y máquinas teóricas o reales tuviesen que compararse.

Este antropocentrismo ha marcado la inteligencia artificial (I.A.) desde sus inicios y sólo ha sido criticado recientemente (cf. p. ej. Preston 1991). De todas las definiciones que se han dado a la I.A., la siguiente fue de las menos controvertidas “la I.A. es aquello que si se hiciera por un humano requeriría inteligencia”. Aunque de por sí es una definición que depende de la de una definición de inteligencia para ser efectiva, muestra que gran parte del fracaso de la I.A. se debe precisamente a esta dependencia, a unos fundamentos no esclarecidos, e incluso cuestionados.

Otra postura más razonable es definir la I.A. como la ciencia dedicada a estudiar la inteligencia, que bien solaparía con la psicología, la filosofía, la neurobiología, disciplinas que hoy entran bajo el paraguas de “ciencias cognitivas”. Pero todas ellas carecen de los requisitos que una ciencia de la inteligencia, o que toda ciencia, requiere:

[‡] También en el Departamento de Lógica y Filosofía de la Ciencia de la Universidad de Valencia.

medición, clasificación y objetividad. La psicometría sería, con mucha diferencia, la más cercana a estos objetivos, pues se basa en la medición de la inteligencia, clasificación de sus diferentes factores y en su rigor estadístico. El primer problema de la psicometría es que se trata de la ciencia del estudio y medición de la inteligencia *humana*. El segundo problema es que, como la I.A., escurre el bulto de la definición de inteligencia con la curiosa afirmación: “la inteligencia es lo que miden los tests de inteligencia”. Sin dejar de puntualizar que, en mi opinión, ésta es la definición más rigurosa y objetiva de inteligencia de las existentes, la misma plantea la siguiente cuestión: ¿qué miden los tests de inteligencia?

Responder a esta pregunta sería de importancia capital para la psicometría, pero si la misma fuera respondida desde un punto de vista formal y computacional, lo sería aún más para la inteligencia artificial, porque podría evaluar la consecución de sus objetivos de una manera más gradual y detallada que el Test de Turing (o su encarnación en patéticos concursos como el Loebner Prize).

Para ello no sólo debemos denostar introspectivismos y trascendentalismos, lo cual debería estar perfectamente superado por tests de ‘comportamiento’, como los psicométricos o el test de Turing, sino que debemos cambiar el punto de referencia:

Los test de inteligencia (ya sean psicométricos, como los IQ, o filosóficos, como el test de Turing) se invalidan al tomar como punto de referencia un ser previamente designado ‘inteligente’ (por autorrecurrentes).

Este cambio de perspectiva es crucial si no queremos que la inteligencia sea un término vago que cambie con las épocas; ¿cuán inteligente consideramos al hombre de las cavernas?, o más incisivamente, ¿cuán inteligentes nos considerarán nuestros sucesores de aquí a 10.000 años? Usar la media de los resultados de los tests de inteligencia a finales del siglo XX, permite realizar comparaciones, pero no resuelve nuestra pregunta ¿qué miden los tests de inteligencia?

Una vez más, necesitamos un punto de partida diferente para abordar esta cuestión:

La inteligencia es un proceso computacional que transforma información. Por tanto, puede ser medida en seres biológicos y computadores sin la necesidad de otro juez inteligente y obviando cualquier consideración acerca de su “conciencia introspectiva”.

Como dijo Shklovskii en los sesenta, sigue siendo “urgente una definición funcional de ‘mente’ que no esté confinada a las nociones preconcebidas de la naturaleza física de este proceso” [Sagan & Shklovskii 1971].

2. Inteligencia como Potencial

El debate fundamental de la psicometría desde que Spearman construyera sus cimientos [Spearman 1904] hasta mitad de este siglo se ha centrado en la diferencia clara entre lo que en biología se conoce como conocimiento adquirido por evolución (fenotipo), conocimiento adquirido por la experiencia y capacidad de adquisición de nuevo conocimiento. Del mismo modo, en computación se distingue entre programa (o motor de inferencia), datos (o base de conocimiento) y capacidad de aprendizaje (o de auto-modificación de los datos). Dado que en los seres humanos el conocimiento adquirido por evolución (innatismo) no es demasiado importante en tareas intelectuales,

incluyendo el lenguaje, la psicometría se ha dedicado a separar entre *conocimiento* e *inteligencia líquida*.

Aunque la psicometría todavía no ha aclarado cómo factorizar esta inteligencia líquida, nadie duda de la capacidad de los tests de inteligencia de discernir entre el conocimiento adquirido y la habilidad innata para comprender nuevo conocimiento. Lo más importante es que los tests se basan única y exclusivamente en observaciones del comportamiento o rendimiento, y, de esta manera, son capaces de diferenciar el origen de la *funcionalidad* del sistema a medir, aprendida o improvisada.

Como resultado de la zoología y de la psicometría, cualquier razonamiento (introspectivo, metafísico o cuántico) en contra de la inteligencia se desvanece al haberse constatado, como ya postulara Darwin con su famosa “continuidad mental” el siglo pasado, que la inteligencia es una habilidad gradual en evolución y, como dijo Spearman al principio de éste, que también evoluciona en los primeros años de vida del homo sapiens y después se estabiliza sobre los quince años para volver a descender muy lentamente sobre los cincuenta.

Esta diferencia entre *funcionalidad específica* y *funcionalidad genérica* ha empezado a ser reconocida recientemente también en el ámbito de la inteligencia artificial. Los sistemas expertos y los sistemas basados en el conocimiento, abanderados de la inteligencia artificial hace una década, han sido fuertemente criticados por su funcionalismo específico y su poca adaptabilidad, debido a que la mayor parte de su conocimiento se introducía por un experto humano, lo cual los convertía en poderosos ‘idiots savants’, como humorísticamente se les ha conocido.

Paralelamente, los últimos treinta años han permitido formalizar el problema del aprendizaje computacional y explorar sus posibilidades. Aparte de las múltiples técnicas que han aparecido hasta nuestros días, lo más significativo es que sus resultados teóricos y, en algunos casos, prácticos, han estado fuertemente relacionados con los conceptos teóricos y fundamentales de información y computación.

3. Información e Inteligencia

En la introducción concretamos el ámbito de la inteligencia como un proceso computacional que transforma la información. Debemos en primer lugar dar una noción universalmente aceptada de información.

La noción de información que Shannon introdujera en 1948 ha quedado superada y subsumida por la que introdujeran Solomonoff, Kolmogorov y Chaitin en los sesenta.

DEFINICIÓN 3.1. COMPLEJIDAD DESCRIPCIONAL

La *Complejidad Descricional* o *Complejidad Kolmogorov* de una secuencia x en un mecanismo descricional β se define como:

$$K_{\beta}(x|y) = \min \{ l_{\beta}(p) : \phi_{\beta}(p|y) = x \}$$

donde p denota cualquier programa en β libre de prefijos, $l_{\beta}(p)$ denota la longitud de p en β y $\phi_{\beta}(p|y)$ denota el resultado de ejecutar p utilizando como entrada y .

La complejidad de una secuencia x se denota como $K_{\beta}(x) = K_{\beta}(x|\varepsilon)$ donde ε denota la cadena vacía. En este caso $K_{\beta}(x)$ representa la longitud del programa o descripción más corta de x en β . La complejidad descricional es un criterio absoluto y objetivo de complejidad y es independiente (hasta un término constante) del mecanismo

descripcional β (p.ej. Li & Vitányi 1997). La razón es simple: existe un teorema de invarianza que dice que cualquier máquina universal puede emular a otra.

Es importante destacar que $K(x)$ es tanto una medida de información como de aleatoriedad. Una serie aleatoria x tiene un gran valor de $K(x)$ mientras que una serie con regularidades se puede comprimir para obtener un valor de $K(x)$ reducido. O visto de otro modo, $K(x)$ representa también perfectamente la idea de complejidad o simplicidad, muy enraizada en la filosofía de la ciencia del razonamiento inductivo y las teorías de aprendizaje. No es por tanto de extrañar que ya haya habido sugerencias en pro de una visión de la inteligencia “à la Kolmogorov”. En 1964, Solomonoff propuso la visión del aprendizaje no supervisado como compresión de información (Solomonoff 1964). Watanabe propuso “el reconocimiento de patrones como compresión de información” [Watanabe 1972]. Del mismo modo, el uso de la compresión como criterio de probabilidad de predicción ha sido una constante desde que Rissanen introdujera el principio MDL (Rissanen 1978) hasta nuestros días (Barron et al. 1998). Este principio es una moderna formalización de “la navaja de Occam”, utilizando $K(x)$ o, más concretamente, la complejidad estocástica de las descripciones. Aunque ha mostrado muy buenos resultados para datos y descripciones aproximados (con posibilidad de errores), y se ha relacionado con otros criterios estadísticos, no deja de ser una elección arbitraria, y suele dar malos resultados para tareas explicativas o creativas (Hernández-Orallo & García-Varea 1999).

Incluso una, exagerada en nuestra opinión, línea de investigación relaciona directamente el razonamiento con la idea de compresión de información: “todos los tipos de computación y de razonamiento formal podrían entenderse de una manera útil como compresión de información mediante emparejamiento de patrones, unificación y búsqueda” (Wolff 1995).

Pero fue Chaitin quien propuso por primera vez el “desarrollar definiciones formales de inteligencia y medidas de sus varios componentes” basadas en la teoría de la información algorítmica (Chaitin 1982). Dicha propuesta, sin embargo, ha sido olvidada durante mucho tiempo¹. Existen numerosas razones técnicas para explicar que una propuesta tan llamativa (y realizada por una persona de tanto prestigio) no haya sido abordada hasta ahora. Vayamos poco a poco descubriéndolas y presentando alternativas.

En primer lugar, la complejidad descriptiva permitiría de una manera directa y sencilla la medición de la capacidad de compresión. Sin embargo, el cociente entre $K(x)$ respecto a la longitud de x no representa la complejidad de x . Existen secuencias cuya descripción más corta es tan enrevesada que tienen poca utilidad como descripción. En 1973, Levin introdujo la siguiente variante de la complejidad descriptiva:

DEFINICIÓN 3.2. COMPLEJIDAD DESCRIPTIVA ESPACIO-TEMPORAL

La *Complejidad Descriptiva Espacio-Temporal* o *Complejidad de Levin* de una secuencia x dada otra secuencia y y en un mecanismo descriptivo β se define como:

$$Kt_{\beta}(x|y) = \min \{ LT_{\beta}(p) : \phi_{\beta}(p|y) = x \}$$

donde $LT_{\beta}(p) = l(p) + \log_2 \text{Cost}(p)$ representa la manera más apropiada, como ahora veremos, de ponderar el espacio de las descripciones con el tiempo que cuesta *ejecutarlas* para obtener su extensión.

¹ Por lo menos hasta lo que el autor ha podido indagar en la bibliografía y el propio Chaitin ha reconocido (Chaitin 1998, comunicación personal).

Aparte de numerosas ventajas, entre ellas que Kt es computable a diferencia de K , y que evita descripciones computacionalmente demasiado complejas², lo más importante de esta variante es que Levin demostró que un algoritmo de búsqueda iterativo ordenado por la complejidad Kt , era un algoritmo de búsqueda universal, en el sentido de que no existe un algoritmo mejor en general, es decir, para todas las posibles secuencias. De manera intuitiva, cualquier algoritmo debe invertir cierto esfuerzo ya sea en tiempo o solicitando/probando nueva información, en una relación que se aproxima a la ponderación de espacio-tiempo representada por LT .

Volviendo a la idea original de Chaitin, Kt parece permitir medir la capacidad de un sujeto (ya sea humano o computador) de encontrar descripciones con bajo LT . Pero aún así dicha medida nos plantea dos dudas fundamentales. La primera es que deberíamos interrogar al sujeto que explicara su descripción (y no sólo el resultado, como se suele hacer en las preguntas de tipo predictivo-explicativo de los test de inteligencia tradicionales). La segunda es que la idea central y original de inteligencia como la capacidad de comprender³ no coincide *exactamente* con esta idea de obtener explicaciones alternativas.

Afortunadamente, podemos elaborar esta idea para resolver ambos problemas, basados en el hecho de que *la comprensión no es más que un tipo especial de explicación*. La característica fundamental que determina la comprensión es bien conocida:

Requisito de Comprensión:
“Lo definido no puede aparecer en la definición”

Existe una razón fundamental para evitarlo: la circularidad. No es de extrañar que sea utilizado por los maestros al preguntar a sus alumnos, y que se evite en los diccionarios.

Por tanto, sería interesante seleccionar aquellas descripciones que cumplen este requisito, para poder establecer la dificultad de comprender, o comprensibilidad, de la evidencia o de un cierto problema. Si es posible establecer esta dificultad, la inteligencia vendrá dada por el grado en el que un determinado sujeto resuelve (o comprende) problemas de mayor dificultad de comprensión.

4. Intensionalidad y Comprensibilidad



La terminología tradicional en matemáticas distingue las definiciones (o descripciones): *por extensión o por comprensión* (o intensión). Sin embargo, esta distinción es completamente intuitiva o no ha habido mucho interés (o éxito) en formalizarla.

La teoría de la complejidad descriptiva parece ser suficiente para discernir cuando una descripción es completamente extensional. Por ejemplo, una secuencia aleatoria x tiene como descripción más corta algo como “PRINT x ”, que es una descripción completamente extensional de x y que evidentemente no dice nada a favor de la comprensión de x . Sin embargo, esa misma x podría tener definiciones no extensionales de longitud mayor o igual que x . Por otra parte, se podría pensar que la descripción más corta para una secuencia comprimible siempre es intensional. Evidentemente, si la

² Existe un concepto denominado profundidad lógica, introducido por el propio Chaitin, que es igual al máximo del mínimo tiempo necesario por cualquier programa k -comprimible (ver Li & Vitányi 1997). La profundidad lógica está íntimamente relacionado con Kt pero difiere para muchas secuencias.

³ Del latín *intelligere* = comprender.

descripción es más corta que aquello a describir es imposible que la descripción sea completamente extensional, pero bien puede ser *parcialmente* intensional.

Por ejemplo, la secuencia $x = \dots 000a000a00a00\dots$ tiene infinitas descripciones, como por ejemplo: $d_1 = \text{“3 ceros, 1 a, repetidamente, con una excepción en la tercera secuencia de ceros (una ‘a’ se ha entrometido)”}$. De acuerdo con d_1 , se haría la siguiente predicción por la izquierda “a” y predicción por la derecha “0”. Otra descripción podría ser $d_2 = \text{“n ceros, 1 a, n – 1 ceros, a, \dots, repetidamente, hasta llegar a un cero, momento en que se invierte el orden”}$. De acuerdo con d_2 , se haría la siguiente predicción por la izquierda “0” y predicción por la derecha “a”.

La segunda descripción no tiene excepciones mientras que la primera sí. Podemos generalizar esta idea para dar una definición de una descripción libre de excepciones:

Una descripción p_x , cuya extensión es x , está libre de excepciones, denotado por $\Delta(p_x) = 0$, si no existe una subdescripción p_y , cuya extensión es y , tal que $l(p_x) - l(p_y) \geq |l(x) - l(y)|$. El significado de esta definición queda más claro si identificamos a p_y con la regla principal de la descripción y $l(p_x) - l(p_y)$ con la longitud de la excepción en la descripción, siendo $|l(x) - l(y)|$ la longitud de la excepción en la extensión.

La idea de excepción ha sido estudiada y formalizada en detalle (así como la idea de subdescripción) en (Hernández-Orallo & Minaya-Collado 1998, Hernández-Orallo & García-Varea 1998). En concreto, podemos definir la complejidad intensional de una secuencia, como la descripción más corta tal que no tenga excepciones. Sin embargo, como hemos dicho antes, a veces las descripciones más cortas son muy intrincadas y, por tanto, son poco explicativas. Es por tanto más razonable definir la complejidad intensional a partir de la variante espacio-temporal:

DEFINICIÓN 4.1. COMPLEJIDAD EXPLICATIVA

La *Complejidad Explicativa* de una secuencia x dada otra secuencia y en un mecanismo descriptivo β se define como:

$$Et_{\beta}(x|y) = \min \{ LT_{\beta}(p) : \phi_{\beta}(p|y) = x \wedge \Delta(\langle p, y \rangle) = 0 \}$$

En este caso, y se puede entender como el contexto o conocimiento previo donde la explicación se debe aplicar. En lo que sigue supondremos $y = \varepsilon$, es decir las secuencias a explicar están sacadas de contexto o el contexto no debería influir⁴. Finalmente, a la primera (en orden lexicográfico) descripción p de x sin excepciones tal que $LT(p) = Et(x)$ se le conoce como la mejor (o más corta) explicación de x y se denota por $SED(x)$.


Por último, sólo es necesario relacionar esta complejidad con la dificultad intrínseca de dar una explicación a una cierta secuencia:

DEFINICIÓN 4.2. COMPRESIBILIDAD

Una secuencia x es *k-difícil* (o *k-incomprensible*) en un sistema descriptivo β si k es el menor entero positivo tal que:

$$Et_{\beta}(x) \leq k \cdot \log_2 l(x)$$

Por ejemplo, dada una secuencia x de longitud 256 cuya mejor explicación tiene un $LT(x) = 50$, resulta que la dificultad de x es $k = 4$.

⁴ Sería interesante estudiar esta influencia, por ejemplo, cómo influye una sucesión de problemas parecidos o su relación entre el conocimiento y  eligencia líquida (o conocimiento improvisado).

5. El Test

En psicología es bien sabido que medir mediante tests no es más que una manera imperfecta y sesgada de evaluar el comportamiento, la personalidad, la inteligencia de un sujeto. Desgraciadamente, si se desea medir el comportamiento, y en nuestro caso el comportamiento inteligente, no hay otra manera de actuar. El problema más fuerte de la psicometría es que se suele aducir que el sujeto puede haber encontrado otra explicación mejor que la del examinador, prediciendo un resultado diferente. Esta crítica es teóricamente correcta, porque cualquier serie o secuencia tiene infinitas ‘explicaciones’, por tanto parece absurdo medir la solución ‘correcta’. La respuesta de la psicometría es que “la mayoría” coincide con cierta solución *porque no hay soluciones alternativas de similar complejidad*, por tanto es ‘la más plausible’.

Procedamos a formalizar dicha idea: decimos que dos descripciones son alternativas si sus predicciones difieren para algún valor pasado o futuro de una secuencia.

DEFINICIÓN 5.1. PLAUSIBILIDAD DE UNA DESCRIPCIÓN

Una descripción p de x es c -plausible si no existe una descripción p' de x tal que $l(p) + c > l(p')$.

Uno de los problemas de $K(x)$ o $Kt(x)$ es que esta definición resulta vacua porque cualquier descripción p tiene una descripción $p' = \text{“describe exactamente lo mismo que } p \text{ pero cambia el primer valor de predicción”}$, cuya longitud es $l(p) + k$, siendo este k muy pequeño (lo que cuesta expresar que se cambie un valor de predicción). Por tanto, sólo se podrían encontrar descripciones c -plausibles para $c < k$.

En cambio, para $Et(x)$, la definición 5.1 sí que formaliza la idea de plausibilidad, porque cualquier ‘parche’ para variar la extensión se puede eliminar, porque se considera una excepción. En concreto:

DEFINICIÓN 5.2. PLAUSIBILIDAD DE UNA EXPLICACIÓN

Una descripción intensional (explicación) p de x es c -plausible si no existe una descripción intensional p' de x tal que $LT(p) + c > LT(p')$.

Se ha dicho muchas veces en filosofía de la ciencia o de la inducción, que esta plausibilidad depende no sólo de las características de la propia explicación respecto al conocimiento anterior y la evidencia sino también de la capacidad de encontrar explicaciones alternativas. Por ejemplo, dos sujetos pueden haber encontrado la misma explicación a una evidencia; uno de ellos no se ha preocupado (o no tiene capacidad) para encontrar explicaciones alternativas, en cambio el otro lo ha intentado fehacientemente, con lo que su plausibilidad es mayor. En este sentido se enmarca la idea de que la inteligencia es el medio más importante para aumentar la plausibilidad de las explicaciones y, a su vez, de gran parte de la ontología de un sistema ‘inteligente’.

Volviendo a nuestro objetivo de realizar un test, aparte de evitar explicaciones alternativas cercanas a la que se propone, también es importante dar un margen de confirmación de la secuencia, en el siguiente sentido:

DEFINICIÓN 5.3. ESTABILIDAD POR LA DERECHA

Una secuencia x es m -estable por la derecha en el sistema descriptivo β si

$$\forall d, 1 \leq d \leq m : \text{SED}_{\beta}(x_{-d}) \text{ es similar a } \text{SED}_{\beta}(x)$$

Donde x_{-d} representa x menos sus últimos d símbolos por la derecha. La noción de ‘similar’ está formalizada en (Hernández-Orallo & Minaya-Collado 1998), pero si ambas descripciones son indefinidas, es decir, generan dos secuencias infinitas, se puede

sustituir por la noción de igualdad. En otras palabras, una secuencia x es m -estable por la derecha si quitando hasta m elementos por la derecha sigue teniendo la misma mejor explicación. Si se da la secuencia completa, estos m elementos se pueden considerar como *refuerzo*, redundancia, *pistas* o confirmación de la explicación. Por ejemplo, dada la serie “2,4,6, ...” parece lógico suponer que seguirá “8, 10, 12, 14, ...”, por tanto se considera redundante poner más de tres o cuatro símbolos para describirla.

Estamos finalmente preparados para generar una serie de problemas que cumplan la propiedad de que el valor a predecir, es decir la solución, no sea una elección arbitraria del examinador, y que, por tanto, hasta un cierto valor se puede decir que es la única y más explicativa que se puede encontrar con unos ciertos recursos.

Para poder aplicar dichos problemas debemos asumir ciertas cosas: la inteligencia es una habilidad o capacidad *gradual*. Por lo tanto no tiene sentido la respuesta de que un sujeto sí o no es inteligente. En segundo lugar, ya que se va a hacer a través de un test, requiere la colaboración del sujeto. Se podría estimar la inteligencia de un sujeto por la mera observación de sus tareas cotidianas, como hacemos generalmente con la gente que nos rodea, pero tal medida no sería demasiado objetiva y mediría muchas más cosas que un test específico de inteligencia, como la simpatía, el conocimiento, ...

Además, no sólo necesitamos la colaboración del sujeto sino que es preciso que entienda el propósito del test (o que esté programado para obtener dicho propósito) y que también entienda el marco y lenguaje en el que se expresa la prueba.

Con todo esto podemos ya definir la inteligencia como el valor que resulta del siguiente test:

DEFINICIÓN 5.4

Seleccionemos un sistema descriptivo suficientemente expresivo e imparcial β y un rango suficientemente amplio $1..K$. Para cada $k = 1..K$ elegimos aleatoriamente p secuencias $x^{k,p}$, tal que sean k -incomprensibles, c -plausibles y d -estables con $d \geq r$, siendo r el número de símbolos redundantes (o pistas) de cada prueba.

Medimos la inteligencia de un pretendido sistema inteligente S de la siguiente manera: Inicialmente, le proporcionamos (o programamos) a S un alfabeto de símbolos Ω_β y unas operaciones Θ_β para manipular esos símbolos, con su correspondiente *coste* (o longitud).

Las preguntas son las $K \cdot p$ secuencias sin sus $d - r$ últimos elementos ($x^{k,p}_{-d+r}$). Se las damos a S y le pedimos el siguiente elemento según la mejor explicación (tal que no tenga excepciones y con el menor *coste* posible) que puedan hacer con Ω_β y Θ_β . Damos a S un tiempo t y registramos sus respuestas: $guess(S, x^{k,p}_{-d+r+1})$.

El resultado de este test de comprensibilidad o C-test es:

$$I(S) = \sum_{k=1..K} k^e \cdot \sum_{i=1..p} hit[x^{k,i}_{-d+r+1}, guess(S, x^{k,i}_{-d+r+1})]$$

siendo hit una función de diferencia (generalmente $hit(a,b) = 1$ si $a = b$ y 0 en caso contrario). El valor e es simplemente un exponente para ponderar en menor o mayor medida las cuestiones difíciles (si elegimos $e = 0$ todas las cuestiones tienen el mismo valor).

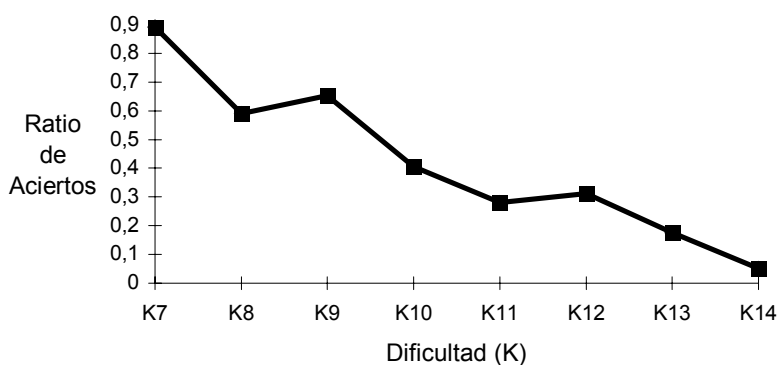
De manera informal, “*el test mide la capacidad de encontrar explicaciones (descripciones sin excepciones) para cadenas de creciente comprensibilidad en un tiempo fijo*”.

Existen muchas maneras de realizar una prueba específica a partir de la definición 5.5. En [Hernández-Orallo & Minaya-Collado 1998] se realizó una implementación de la misma utilizando una máquina abstracta muy parecida a una clásica máquina de estados, que no es más que una versión simplificada de un RISC (*Reduced Instruction Set Computer*). A partir de ahí se generó una variedad de cadenas de diferente *comprensibilidad* en esa máquina, tarea computacionalmente compleja y costosa y que necesita de ciertas heurísticas para poder generar secuencias con gran valor de dificultad. Finalmente, se aplicó una criba para quedarse sólo con secuencias *d*-estables.

En el mismo artículo se presentan los resultados de aplicar el test a 65 sujetos de la especie *homo sapiens sapiens* entre 14 y 32 años, junto con un test clásico de inteligencia, el *European IQ Test*. La correlación entre los dos tests resultó ser 0,77. Dicho valor es suficiente como para justificar un estudio más exhaustivo sobre grupos mayores y con diferentes tests generados a partir de la definición 5.4.

Esta evidencia psicométrica es de vital importancia para una teoría formal de medición de inteligencia, porque, según Brand [Brand 1996], la correlación con los tests de IQ es una condición necesaria (aunque no suficiente) para una buena medida de inteligencia.

Finalmente, y a la pregunta que da título a este artículo, la siguiente figura muestra el ratio de aciertos de los 65 sujetos para secuencias de creciente dificultad.



Como conclusión fácil de este estudio podríamos decir que el *homo sapiens* tiene una inteligencia media capaz de entender en un tiempo razonable secuencias de *comprensibilidad* = 10. También se podría decir que menos del 5% de la población son capaces de entender secuencias de *comprensibilidad* > 14. Evidentemente, estas medidas se deberían estandarizar para una máquina de referencia más adecuada, pero resulta ser la primera medida de un factor de la inteligencia (el factor *g*) que no utiliza al *homo sapiens* como pie de rey.

Es posible que una referencia absoluta, o al menos no antropomórfica, de inteligencia sea cada vez más interesante en el futuro, cuando los computadores se acerquen más a las capacidades del ser humano. Será imprescindible cuando nos superen de una manera significativa, porque la referencia humana, al no estar teóricamente fundamentada, no se puede proyectar de una manera justificada para inteligencias superiores. Tampoco parece una solución demasiado acertada ir cambiando el punto de referencia a medida que la tecnología avanza, sobretodo si son estas “inteligencias punta” las que deberían realizar los tests.

6. Evaluando Sistemas Inteligentes

Por el momento, el gran reto sigue siendo acercarse (y no superar) al homo sapiens. Precisamente para este reto es donde esta medida puede tener mayor utilidad. Como en cualquier otro campo de la ciencia, un gran avance de la disciplina ocurre cuando alguno de los tópicos más relevantes de ella puede ser medida de una manera efectiva y justificada. La inteligencia artificial, como ciencia, requiere medidas de la inteligencia, o por lo menos, medidas de sus diferentes factores.

La definición 5.4 proporciona una escala más detallada que la mayoría de evaluaciones subjetivas utilizadas para evaluar el avance de sistemas inteligentes o de aprendizaje automático. En los primeros se basa en nuevas funcionalidades para problemas concretos. La ciencia de la “inteligencia artificial” ‘avanza’ porque se resuelven problemas de una manera automática que antes requerían la participación humana. Sin embargo, esto no representa en absoluto que el nuevo sistema sea más inteligente que los anteriores, más bien se podría decir que sus *desarrolladores* han sido más inteligentes que sus anteriores o han tenido más medios técnicos y científicos. Además nadie se preocupa de medir si son más o menos *funcionales* para otros tipos de problemas, por lo que en la mayoría de los casos no se tratan de resolvedores de problemas generales, como inicialmente se intentó en I.A.

En el caso de los segundos, los sistemas de aprendizaje automático, las medidas merecen un poco más de respeto. Suelen existir una serie de bancos de problemas en los que se evalúa el tiempo de resolución. El problema es que dichos bancos se realizan de una manera bastante desorganizada, como simple recopilación de ejemplos que han aparecido en la bibliografía y sin ninguna base de cuál es su complejidad teórica. Además, en muchos casos, son bancos muy restringidos, con la única excepción de la librería TPTP [Suttner & Sutcliffe 1996] de problemas de demostración de teoremas, usada desde hace años para comparar sistemas de razonamiento automático (demostración automática de teoremas).

Cómo realizar diferentes y fundamentados tests de inteligencia, sabiendo exactamente qué es lo que se va a medir, es una tarea urgente a realizar en inteligencia artificial. En un futuro es posible que se lleguen a realizar concursos al estilo del Loebner Prize pero mucho menos antropomórficos. Llegado el momento, y durante una temporada, computadores y seres biológicos podrían participar de una manera más ecuánime.

7. Protocolos de Autenticación de Inteligencia

Aparte de servir como realimentación para el avance de la “verdadera inteligencia artificial” o de la evaluación del progreso de una cierta persona (ya sea un niño o un sistema artificial) , ¿tienen o tendrán alguna otra utilidad los tests de inteligencia?

En la década de los sesenta y setenta se produjo un intenso interés sobre la comunicación con vida inteligente extraterrestre, siendo Carl Sagan, uno de sus mayores impulsores y difusores del noble ideal de contactar con otras inteligencias. Se afirmaba que la comunicación con vida inteligente exterior se debía basar en las características intrínsecas de los mensajes, que tuvieran ciertos patrones lo suficientemente complejos para que la probabilidad de darse en la naturaleza ese tipo de emisiones fuera muy remota y lo suficientemente sencillos para comprenderlos. Hoy en día el interés por los contactos

extraterrestres ha disminuido de manera notoria, a excepción de la aparición de alguna película. Sin embargo, ‘contactos’ *parecidos* empezarán a ser más frecuentes en el futuro.

Me estoy refiriendo a sistemas multiagentes de gran ámbito, como pueden ser las redes de computadores y, en especial, Internet. Un agente establece la comunicación con otro agente, siendo ambos desconocidos entre sí. Tanto uno como otro pueden estar bastante interesados en conocer el nivel de inteligencia del otro. Un test de Turing requeriría un tercer agente, humano, para conocer simplemente si el otro interlocutor es humanoide con una cierta probabilidad, siendo uno de sus mayores problemas que es un “test de humanidad, no de inteligencia” (Fostel 1993). Aun suponiendo que uno de los dos es humano y que es precisamente la ‘humanidad’ del otro lo que queremos medir, dicha comunicación deberá ser cada día más larga y concienzuda para tener una cierta probabilidad de averiguar la ‘humanidad’ o no del otro agente⁵. En otras palabras, el Test de Turing dejará de ser útil en breve, si es que ha sido útil *en la práctica* alguna vez.

Ante esta situación es necesario cambiar el enfoque radicalmente, sin la necesidad de un ‘juez’. A partir de nuestra definición anterior es posible realizar un protocolo de autenticación de la inteligencia. Dicho test se pasaría a ambos interlocutores en el momento que se desea saber la inteligencia de cada uno. Evidentemente un agente puede fingir ser menos inteligente de lo que realmente es pero difícilmente puede aparentar ser más inteligente.

Aparte del propio interés de ambos agentes por conocer el resultado en el otro lado, dicho protocolo deberá ser elaborado de manera que los dos sintonicen al nivel de inteligencia inferior de los dos, o, en el caso de un análisis multifactorial inteligencia a los mínimos por cada factor. El objetivo es que cuando las inteligencias son muy dispares la comunicación todavía puede ser posible. La analogía con los ‘módems’ es esclarecedora; ambos intentan la comunicación a sus máximas velocidades y van bajándola hasta que consiguen establecerla a la máxima que permiten los dos.

Es interesante destacar que dicho protocolo ha estado asumiéndose entre humanos y computadores desde el mismo principio de la informática y hoy en día es un aspecto crucial en la interacción humanos-computador. Hasta el momento, el ser humano debe ‘rebajar’ su inteligencia para poder comunicarse con los computadores, debe saber que los computadores, hoy en día, no entienden analogías, no entienden las bromas, no resuelven la mayoría de ambigüedades de la comunicación humana y, por tanto, la expresividad y flexibilidad del lenguaje utilizado se reduce a su nivel.

Ha habido un reciente interés por definir lenguajes estándares para la comunicación entre agentes. Aunque puede ser una solución (o una necesidad) a corto plazo, cualquier lenguaje se extiende y *cambia* a medida que se conoce la inteligencia del otro interlocutor, haciéndose más ambiguo y peor entendible para terceros.

8. Conclusiones

Uno de las primeras críticas que se suelen argumentar ante esta definición del factor *g* es que cualquier problema complejo requiere inteligencia para resolverlo y por tanto permite medirla. De este modo, la capacidad de realizar *puzzles*, por ejemplo, serviría como test de inteligencia. Afortunadamente, la psicometría se ha dedicado a rebatir esta idea: no

⁵ Basta echar un vistazo a los buenos resultados de los concursantes del Loebner Prize (especie de competición basada en el Test de Turing) a pesar de su poca inteligencia y humanidad.

todo problema complejo es válido para medir la inteligencia. Más aún, se ha comprobado que sólo unos pocos tipos de problemas permiten medirla con fiabilidad. Estos no son más que los problemas predictivo-explicativos de series o secuencias de símbolos, sean de letras, números o figuras, donde el objetivo es encontrar el patrón más sencillo que la explique sin excepciones.

Parece, por tanto, que existen ciertas propiedades intrínsecas en los problemas que los hacían buenos o malos para medir la inteligencia. En (Hernandez-Orallo & Minaya-Collado 1998) se relacionan por primera vez los clásicos tests de inteligencia refinados a lo largo de este siglo con formalismos computacionales. Las consecuencias de esta conexión son múltiples. En primer lugar, la psicometría obtiene un espaldarazo importantísimo al estar justificada matemáticamente, es decir, los tests de inteligencia pueden hacerse de manera empírica pero también se pueden generar a partir de modelos teóricos (aunque por el momento no parece razonable sustituir los bancos de tests ya comprobados con nuevos “tests de C.I. generados por ordenador”). En segundo lugar, se identifica el factor *g* de la inteligencia con la capacidad de encontrar explicaciones plausibles, algo que resulta ser crucial para el día a día, desde el lenguaje que se aprende siendo niño hasta la comprensión de los acontecimientos diarios. En tercer lugar, proporciona numerosas sugerencias sobre hacia dónde y cómo debe caminar la investigación en inteligencia artificial, si el objetivo es lo que le da nombre y no otros menesteres, que aunque loables, no deberían ir acompañados por ese nombre. En cuarto lugar, resuelve muchas dudas acerca del sentido ‘común’ en la teoría del significado y la pobreza de estímulos de la teoría del aprendizaje, identificando el porqué las personas suelen realizar las mismas explicaciones ante los mismos hechos. O como apuntara Hofstadter en los setenta, “estaría bien si fuésemos capaces de definir la inteligencia de una manera diferente a “eso que extrae el mismo significado a partir de una secuencia de símbolos que nosotros”” [Hofstadter 1979].

Finalmente, la independencia de estas medidas de las características del homo sapiens representa la base para un nuevo marco teórico y formal donde examinar preguntas trascendentales acerca de nuestra naturaleza: desde la pregunta que da título a este artículo hasta evaluar qué recursos computacionales serían necesarios para una mayor inteligencia.

En definitiva, este artículo *simplemente* ha identificado computacionalmente uno de los factores más importantes (si no el más) que componen eso que llamamos inteligencia. Evidentemente, hay partes de la inteligencia humana que quedan fuera y que son indispensables para la vida cotidiana. Antes hemos comentado que el segundo parámetro de $Et(x|y)$ puede ser utilizado para medir la capacidad de contextualización, y otros muchos factores pueden ser también identificados con nociones computacionales. Por el momento, aquello que miden estos tests es el famoso factor ‘*g*’ o la parte fundamental de la inteligencia, la capacidad de comprender la realidad (cualquier realidad). Creo sinceramente que este factor será todavía más fundamental en un futuro lejano, cuando la vida ‘cotidiana’ no sea el objetivo de los constructores de inteligencias cada vez más superiores.

Agradecimientos

No podría dejar de nombrar a dos de los grandes inspiradores de este trabajo, Greg Chaitin y Douglas Hofstadter, por sus comentarios alentadores, especialmente durante 1997, cuando esta teoría quería tomar cuerpo frente a la incompreensión de muchos.

Desde entonces las ideas han ido madurando gracias a numerosas colaboraciones y sugerencias: Rafael Beneyto y Juan Manuel Lorente del departamento de Lógica y Filosofía de la Ciencia de la Universidad de Valencia, Paul Thagard del departamento de filosofía de la Universidad de Waterloo, Olga Pellicer del departamento de psicobiología de la Universidad de Valencia, Nigel T. Crook de la Oxford Brookes University, Enrique Fueyo de la UNED, Ismael García del Instituto Tecnológico de Informática de Valencia e Ignacio Soto de la Universidad de Valladolid.

Finalmente, agradezco especialmente los comentarios de Kike Araque, Enrique Hernández y Neus Minaya sobre numerosos borradores previos de este artículo.

Referencias

- [A.S. 1992] Artificial Stupidity, *The Economist*, vol. 324, no. 7770, August 1, p. 14, Editorial, 1992.
- [Barron et al. 1998] Barron, A.; Rissanen, J.; Yu, B., 1998, The Minimum Description Length Principle in Coding and Modeling, *IEEE Transactions on Information Theory*, V. 44, No. 6, Oct. 1998, 2743-2760.
- [Bennett et al. 1998] Bennett, C.H.; Gács, P.; Li, M.; Vitanyi, P.M.B.; Zurek, W.H. Information Distance, *IEEE Trans. Information Theory*, IT-44:4, 1407-1423, 1998.
- [Block 1981] Block, N. Psychologism and behaviorism. *Philosophical Review* 90:5-43, 1981.
- [Blum 1967] Blum, M. "A machine-independent theory of the complexity of recursive functions" *J. ACM* 14, 4, 322-6, 1967.
- [Blumer et al. 1987] Blumer, A.; Ehrenfeucht, A.; Haussler, D.; Warmuth, M. K. "Occam's razor" *Inf.Proc.Lett.* 24, 377-380, 1987.
- [Blumer et al. 1989] Blumer, A.; Ehrenfeucht, A.; Haussler, D.; Warmuth, M. "Learnability and the Vapnik-Chervonenkis Dimension" *Journal of ACM*, 36, pp. 929-965, 1989.
- [Boolos & Jeffrey 1989] Boolos, G.; Jeffrey, R. "Computability and Logic" 3rd Edition, Cambridge University Press 1989.
- [Bosch 1994] Bosch, van den, Simplicity and Prediction, Master Thesis, dep. of Science, Logic & Epistemology of the Faculty of Philosophy at the Univ. of Groningen, 1994.
- [Bradford & Wollowski 1995] Bradford, P.G.; Wollowski, M., A Formalization of the Turing Test (The Turing Test as an Interactive Proof System), *SIGART Bulletin*, Vol. 6, No. 4, p. 10, 1995.
- [Brand 1996] Brand, C. "The g Factor: General Intelligence and its implications" Wiley 1996
- [Chaitin 1982] Chaitin, G.J. "Gödel's Theorem and Information" *Int. J. of Theoretical Physics*, vol.21, no.12, pp. 941-954, 1982.
- [Chaitin 1992] Chaitin, G.J. "Algorithmic Information Theory", fourth printing, Cambridge University Press, 1992.
- [Chandrasekaran 1990] Chandrasekaran, B. "What kind of Information Processing is Intelligence?" in Partridge,D.; Wilks, Y., *Foundations of AI: A Source Book* Cambridge Univ. Press, 1990.
- [Clark & Toribio 1998] Clark, A. and Toribio, J. (eds.) *Cognitive Architectures in Artificial Intelligence. The Evolution of Reserach Programs*, Garland Publishing, Inc., 1998.
- [Dietrich 1990] Dietrich, E. "Programs in the Search for Intelligent Machines: The Mistaken Foundations of AI" in Partridge & Yorick, *Foundations of AI: A Source Book*, Cambridge Univ. Press, 1990.
- [Epstein 1992] Epstein, R., Can Machines Think? *AI Magazine*, Vol. 12, No.2, 80-95, 1992.
- [Ernis 1968] Ernis, R., Enumerative induction and best explanation, *J. Philosophy*, LXV (18): 523-529, 1968.
- [Eysenck 1979] Eysenck, H.J. "The Structure and Measurement of Intelligence", Springer-Verlag 1979.
- [Fostel 1993] Fostel, G., The Turing Test is For the Birds, *SIGART Bulletin*, Vol. 4, No. 1, 7-8, 1993.
- [Freivalds et al. 1995] Freivalds, R.; Kimber, E.; Smith, C.H. "On the Intrinsic Complexity of Learning" *Inf. & Control* 123, 64-71, 1995.

- [Gold 1967] Gold, E. M. "Language Identification in the Limit", *Inform and Control*, 10, pp. 447-474, 1967.
- [Harman 1965] Harman, G., The inference to the best explanation, *Philos. Review*, 74: 88-95, 1965.
- [Harnad 1992] Harnad, S., The Turing Test Is Not a Trick: Turing Indistinguishability Is A Scientific Criterion, , *SIGART Bulletin*, Vol. 3, No. 4, 9-10, October 1992,
- [Hempel 1965] Hempel, C.G., *Aspects of Scientific Explanation*, The Free Press, New York, 1965.
- [Herken 1994] Herken, R. "The universal Turing machine: a half-century survey" Oxford University Press, 1994.
- [Hernández-Orallo 1999a] Hernandez-Orallo, J., 1999a, Constructive reinforcement learning, *International Journal of Intelligent Systems*, to appear.
- [Hernández-Orallo & García-Varea 1998] Hernández-Orallo, J. and Garcia-Varea, I., 1998, Distinguishing abduction and induction under intensional complexity, in: *Proceedings of the ECAI'98 Workshop on Abduction and Induction in AI*, Flach, P.; Kakas, A., eds., 41-48, Brighton.
- [Hernández-Orallo & García-Varea 1999] Hernández-Orallo, J. and García-Varea, I., 1999, Explanatory and Creative Alternatives to the MDL principle, presented at MBR'98, *Proceedings*, to appear.
- [Hernández-Orallo & Minaya-Collado 1998] Hernández-Orallo, J.; Minaya-Collado, N. A FDI based on an intensional variant of algorithmic complexity, *Proc. of the Intl. Symp. of Engin. of Intelligent Systems, EIS'98*, ICSC Press 1998, pp. 146-163.
- [Hofstadter 1979] Hofstadter, D.R. "Gödel, Escher, Bach: An eternal golden braid" New York: Basic Books, 1979.
- [Hofstadter 1985] Hofstadter, D.R. "Metamagical Themas. Questing for the Essence of Mind and Pattern" Basic Books, Inc., 1985.
- [Hofstadter et al. 1995] Hofstadter, D.R.; Fluid Analogies Research Group "Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought" Basic Books, 1995.
- [Johnson 1992] Johnson, W.L., Needed: A New Test of Intelligence, *SIGART Bulletin*, Vol. 3, No. 4, 7-9, October 1992, Editorial and Commentary.
- [Klir 1985] Klir, George J. "Complexity, Some General Observations" *Systems Research*, no. 2, pp. 131-140, 1985.
- [Kolmogorov 1965] Kolmogorov, A.N. "Three Approaches to the Quantitative Definition of Information" *Problems Inform. Transmission*, 1(1):1-7, 1965.
- [Kolmogorov 1968] Kolmogorov, A.N., Logical basis for information theory and probability theory, *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 662-664, sept. 1968.
- [Koppel 1987] Koppel, M., Complexity, Depth, and Sophistication, *Complex Systems* 1, 1087-1091, 1987.
- [Kotovsky and Simon 1990] Kotovsky, K.; Simon, H.A. "Why are some problems really hard: explorations in the problem space of difficulty". *Cognitive Psychology*, 22, 143-183 1990.
- [Larsson 1993] Larsson, J.E., The Turing Test Misunderstood, *SIGART Bulletin*, Vol. 4, No. 4, p. 10, 1993.
- [Levin 1973] Levin, L.A. "Universal search problems" *Problems Inform. Transmission*, 9:265-266, 1973.
- [Li & Vitányi 1997] Li, M.; Vitányi, P. "An Introduction to Kolmogorov Complexity and its Applications" 2nd Ed. Springer-Verlag 1997.
- [Lucas 1962] Lucas, J.R. "Minds, Machines, and Gödel" Reprinted in Anderson, A. (ed) "Minds and Machines", Prentice Hall, 1962.
- [Merhav & Feder 1998] Merhav, N.; Feder, M., Universal Prediction, *IEEE Transactions on Information Theory*, Vol. 44, No. 6, October 1998, 2124-2147.
- [Millican & Clark 1996] Millican, P.J.R. and Clark, A. (eds.) *Machines and Thought. The Legacy of Alan Turing*, Vol. I, Clarendon Press, Oxford, 1996.
- [Moody 1993] Moody, T.C., *Philosophy and Artificial Intelligence*, Englewood Cliffs, NJ, Prentice Hall, 1993.
- [Moravec 1998] Moravec, H., *ROBOT: Mere Machine to Transcendent Mind*, Oxford Univ. Press, 1998.
- [Newell 1990] Newell, A. *Unified Theories of Cognition*, Cambridge, Mass.: Harvard University Press, 1990.
- [Preston 1991] Preston, B. 1991. AI, anthropocentrism, and the evolution of "intelligence.". *Minds and Machines* 1:259-277.
- [Quinlan & Rivest 1989] Quinlan, J.; Rivest. R., 1989, Inferring decision trees using the minimum description length principle, *Information and Computation*, vol. 80, pp. 227-248.

- [Rissanen 1978] Rissanen, J., Modeling by the shortest data description, *Automatica-J.IFAC*, 14:465-471, 1978.
- [Rissanen 1986] Rissanen, J., Stochastic complexity and modeling, *Annals Statist.* 14:1080-1100, 1986.
- [Rissanen 1996] Rissanen, J., Fisher information and stochastic complexity, *IEEE Trans. on Information Theory*, 42(1), 1996.
- [Sagan 1973] Sagan, C. (ed). "Communication with Extraterrestrial Intelligence" Cambridge, Mass.: MIT Press, 1973.
- [Sagan & Shklovskii 1971] Sagan, C. & Shklovskii, *Intelligent Life in the Universe*, 1971.
- [Shannon 1948] Shannon, C.E. "The mathematical theory of communication" *Bell System Tech. J.*, 27:379-423, 623-656, 1948.
- [Shapiro 1992] Shapiro, S.C., The Turing Test and The Economist, *SIGART Bulletin*, Vol. 3, No. 4, 10-11, October 1992.
- [Simon & Kotovsky 1963] Simon, H.; Kotovsky, K. "Human acquisition of concepts for sequential patterns" *Psych. Review* 70, 534-46, 1963.
- [Smullyan 1992] Smullyan, R.M. "Gödel's Incompleteness Theorems" Oxford University Press 1992.
- [Solomonoff 1964] Solomonoff, R.J. "A formal theory of inductive inference" *Inf. Control.* vol. 7, 1-22, Mar., 224-254, June 1964.
- [Solomonoff 1978] Solomonoff, R.J., Complexity-based induction systems: comparisons and convergence theorems, *IEEE Trans. Inform. Theory*, IT-24:422-432, 1978.
- [Solomonoff 1986] Solomonoff, R.J. "The Application of Algorithmic Probability to Problems in AI" in L.N. Karnal; J.F. Lemmer(eds) *Uncertainty in AI*, Elsevier Science, pp.473-91, 1986.
- [Spearman 1904] Spearman, C. "'General Intelligence' objectively determined and measured" *Amer. J. of Psych.*, 15, 201-293, 1904.
- [Sternberg 1977] Sternberg, R.J. "Intelligence, Information Processing, and Analogical Reasoning" John Wiley & Sons 1977
- [Stonier 1992] Stonier, T. "Beyond Information. The Natural History of Intelligence" Springer-Verlag 1992.
- [Suttner & Sutcliffe 1996] Suttner, C.B.; Sutcliffe, G. "The TPTP Problem Library", Tech. Univ. Munich, Germany, 1996
- [Thagard 1986] Thagard, P., The emergence of meaning: An escape from Searle's Chinese Room. *Behaviorism* 14:139-46, 1986.
- [Thagard 1989] Thagard, P. "Explanatory coherence" *The Behavioural and Brain Sciences*, 12 (3), 435-502, 1989.
- [Turing 1936] Turing, A.M. "On computable numbers with an application to the Entscheidungsproblem" *Proc. London Math. Soc.*, series 2, 42:230-265, 1936. Correction, *Ibid*, 43:544-546, 1937.
- [Turing 1950] Turing, A.M. "Computing Machinery and Intelligence" *Mind* 59: 433-460, 1950.
- [Valiant 1984] Valiant, L. "A theory of the learnable". *Communication of the ACM* 27(11), 1134-1142, 1984.
- [Vitányi & Li 1997] Vitányi, P., and Li, M., On prediction by data compression, in: *Proc. of the 9th European Conf. on Machine Learning*, LNAI 1224, Springer-Verlag, 14-30, 1997.
- [Watanabe 1972] Watanabe, S. "Pattern Recognition as Information Compression" in Watanabe (ed.) *Frontiers of Pattern Recognition* New York: Academic Press, 1972.
- [Watanabe 1992] Watanabe, O. (ed.) "Kolmogorov complexity and computational complexity" *Monographs on TCS*, Springer 1992.
- [Winograd 1986] Winograd, T. "Thinking Machines: Can There Be? Are We?" in Winograd & Fernando Flores "Understanding Computers and Cognition: A New Foundation for Design" Norwood, 1986.
- [Winston 1992] Winston, P.H. "Artificial Intelligence" Third edition, Addison-Wesley Publishing Company 1992
- [Wolff 1995] Wolff, J.G. "Computing as Compression: An Overview of the SP Theory and System" *New Gen. Computing* 13, 187-214, 1995.

NOTA: Este apéndice apareció en (Hernández-Orallo, J.; Minaya-Collado, N. 1998)

Se incluye aquí para ilustrar el test y sus resultados.

Appendix A. An Example of C-Test

The problem of selecting a good bias for generating *k-hard* strings depends on many factors. The objective is to maintain expressiveness, to ease the problem of finding explanatory descriptions and to limit the combinatorial *explosion*. The final choice we present is an oversimplified abstract machine that is easily extensible to work as a Turing machine.

A.1 A Toy Memory-less Abstract Machine

Due to the current technology of the computers we can use, we have chosen an extremely abridged emulation of the machine that will *effectively* run the programs, instead of more proper languages, like λ -calculus (or LISP). We have adapted the “toy RISC” machine of [Hernández & Hernández 1993] with two remarkable features inherited from its object-oriented coding in C++: it is easily tunable for our needs, and it is *efficient*. We have made it even more reduced, removing any operand in the instruction set, even for the loop operations. We have only three registers which are AX (the accumulator), BX and CX. The operations Θ_β we have used for our experiment are in Table 1:

LOOPTOP	Decrements CX. If it is not equal to the first element jump to the program top.
LOOPS	Same as LOOPTOP but it jumps n (for the tests n=4) instructions backward.
LOOPM	Same as LOOPTOP but it jumps m (for the tests m=7) instructions backward.
SUCC	Increments the accumulator.
PRED	Decrements the accumulator.
WRITE	Writes into the output and moves fwd.
BREAD ²	Moves back and reads from the output.
FREAD ²	Moves fwd and reads from the output.
MOV A,B ¹	Copy register BX into AX
MOV B,A ¹	Copy register AX into BX
MOV A,C	Copy register CX into AX
MOV C,A	Copy register AX into CX
ROTR ³	Rotates 45° to the right.
ROTL ³	Rotates 45° to the left.

Table 1. Instruction Set

The operations with no superscript are present in all the subsets. Operations marked with (1) are present in the ‘professional’ version of the machine, the operations with (2) are present in the *Turing-like* version and those with (3) are present in the *logo* version where the output is bidimensional. This sparseness of only 10 operations will be clearly justified later. We have essayed with many different alphabets but for this test we will use the professional version and a circular alphabet $\Omega_\beta = \{a,b,c,d,\dots,z\}$, that is, incrementing ‘z’ yields ‘a’ and decrementing ‘a’ yields ‘z’. Since the first element is an inflexion point for the loops, it is presented to the subjects as “*a critical element*”.

This configuration *still* produces many programs that are not robust because programs can be often split into subprograms. The solution for these cases comes from another restriction: the programs must be comprised wholly inside a loop. This leaves a good approximation to explanatory programs. The rest to do is to avoid repetitions of patterns like “abcabcabcabc” and take apart the strings where an important part is explained by a shorter program. We think that the bias is not all the expressible we would like but it allows the generation of strings of certain complexity. Also we think it is *fair* because it does not relate on arithmetic (like *cryptarithmics* tests) or any other preceding knowledge, except the order of the alphabet.

A.2 The Generation of k -Hard Strings

The algorithm we have used to generate a set of different k -incomprehensible strings is very similar to the one we presented in section 5.4. Having 10 operations, we have that usually only about a 20% of the programs of any size are explanatory. This means that trying to know if a randomly generated program of, say, size 15, will need the checking of more than 2,222,222,222,222 programs. And this is the case if the computational cost of x^* is slow, contrariwise (if x^* is a *costly* program) we will have to check longer programs.

We have used some optimisations and heuristics in order to make the great amount of programs to check more tractable. Some examples of questions are:

Prediction style:

$k9$:	a, d, g, j, ...	Answer: ‘m’
$k12$:	a, a, z, c, y, e, x, ...	Answer: ‘g’
$k14$:	c, a, b, d, b, c, c, e, c, d, ...	Answer: ‘d’

Abduction style:

$k8$:	a, _, a, z, a, y, a, ...	Answer: ‘a’
$k10$:	a, x, _, v, w, t, u, ...	Answer: ‘y’
$k13$:	a, y, w, _, w, u, w, u, s, ...	Answer: ‘y’

A.3 The Tests

Four tests were devised to measure prediction, abduction, g -factor and similarity. The prediction test is composed of 19 exercises generated with the following k -hardness distribution (2 $k7$, 1 $k8$, 2 $k9$, 3 $k10$, 3 $k11$, 3 $k12$, 2 $k13$ and 1 $k14$), redundancy $r = 2$ and the less ‘akin’ as possible. The abduction test is composed of 15 exercises using the same generator and redundancy. The distribution was (2 $k7$, 2 $k8$, 1 $k9$, 2 $k10$, 1 $k11$, 3 $k12$ and 4 $k13$). In these two tests, the incorrect options were generated randomly but relative near to the solution and the letters appearing in the string. The IQ test we used was the European IQ test simply because it is a culture-fair test, devised for 20 minutes, ensuring a reasonable range (75-174) of values and available on the Internet. The similarity test is composed of 8 exercises generated with binary strings of different length and different levels of edit errors (insertion, deletion or change). The strings were generated and checked by dynamic programming to ensure that they did not have a better correction path. The purpose of this test was to measure the ability of compression by *trivial* pattern matching.

A.4 Subjects and Administration

Subjects were selected from two different groups: the first group was composed by 48 high-school students with ages comprised between 14 and 18 years. The second group

was composed by 17 subjects of a mixed sample of undergraduate and postgraduate university students with ages comprised between 22 and 32 years.

All the tests were passed in the same session. The times were, without including instructions, 10 min. for the prediction test, 5 min. for the abduction test, 5 min. of break, 20 min. for the IQ test and 3 min. for the similarity test.

A.6 Results

We evaluated the test without penalising the errors, i.e, the function *hit* evaluated the same for blanks than for mistakes. We chose $e=0$, i.e. all questions with the same value. IQ-correlations are illustrated in Table 2.

	Pred.	Abd.	Induct.	Simil.
High-School	0.31	0.38	0.42	0.39
University	0.51	0.42	0.56	0.35
Both Groups	0.73	0.68	0.77	0.50

Table 2. Correlations with EIQ test

The correlation for induction (prediction + abduction) is of the same order as the usual correlation for induction tests made by psychologists. The correlation between the abduction and prediction tests was 0.61, less than expected, which suggests that even problems constructed by the same generator can be more or less difficult depending on its presentation (abductive or predictive). The correlation between induction and similarity was 0.51 which supports the thesis that “the ability of compression” is different from “the ability of comprehension”. Finally, we think that an analogy test based on our theory would surely round off the study.

With these data and our amateur methods we are not in conditions to assert more things about the relation between C-tests and IQ-tests. There is only a thing that has no discussion in the light of the results, the k -hardness matches fairly well with the difficulty people found on them, as it is seen in Figure 1:

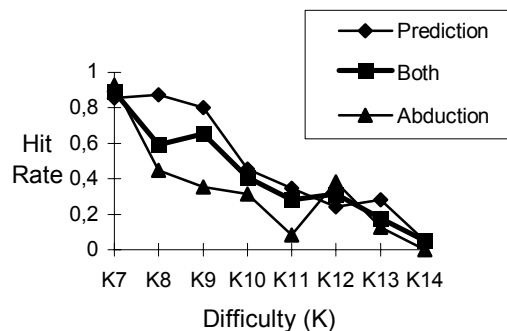


Figure 1. Hit Rate per Difficulty