Medir sin Pie de Rey:

Evaluando la Inteligencia de Sistemas Informáticos Supuestamente Inteligentes

José Hernández Orallo

Dpto. de Sistemas Informáticos y Computación,

Universidad Politécnica de Valencia,

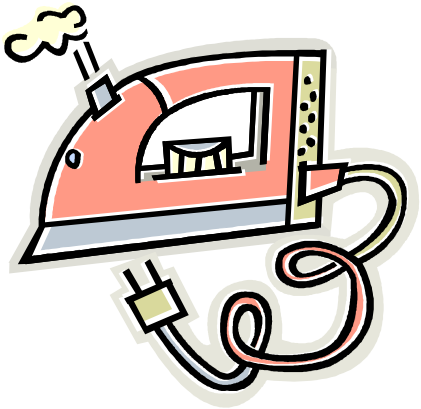
Valencia, España

jorallo@dsic.upv.es

UAI-2004, 25 de noviembre 2004.

Inteligentes

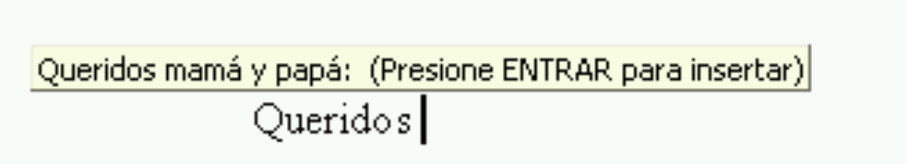
No inteligentes



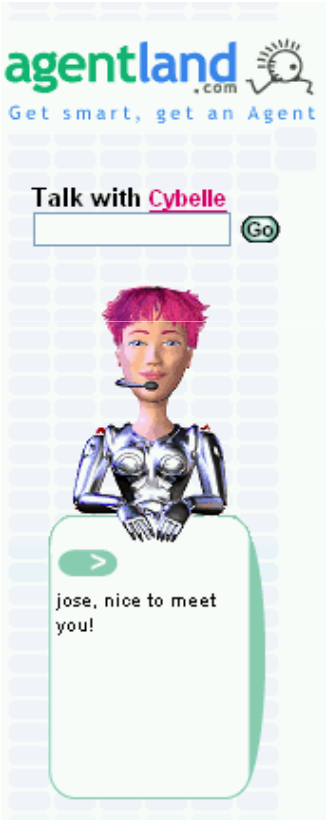
¿Inteligente?

Robots autónomos

Mascotas y otros seres de compañía

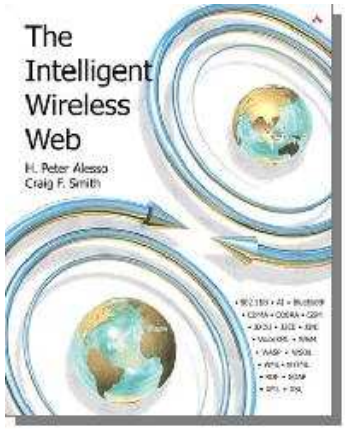


Aplicaciones software "inteligentes", "customising applications".



Agentes y avatares para el comercio electrónico, chatbots

Domótica



Web-bots, Smartbots, Security bots, Search bots…

Asistentes inteligentes

Organización

- ¿Para qué sirve medir la inteligencia?
- Los tests psicométricos
- Los tests por imitación: el Test de Turing
- Los tests por incapacidad: *captcha*
- Requerimientos de tests de inteligencia
- Tests basados en compresión
- Factorización
- Fiabilidad y aplicabilidad
- Evaluando comunidades
- Conclusiones

¿Para qué sirve medir la inteligencia?

- Evaluar la inteligencia **humana** sirve para:
  - Detectar posibles problemas en niños y adultos (tanto por debajo como por encima de la media).
  - Seleccionar personal en empresas e instituciones públicas (incluso educativas).
  - Comparar poblaciones, sistemas educativos, ...

Psicometría: ciencia de la medición de las capacidades intelectuales **humanas**.

¿Para qué sirve medir la inteligencia?

- Evaluar la inteligencia **no humana** sirve para:
  - Determinar la capacidad de sistemas inteligentes.
  - Certificar habilidades y funcionalidades.
  - Comparar sistemas inteligentes humanos con no humanos.
  - Evaluar el progreso *real* de la Inteligencia Artificial.

Problema sin resolver

(siquiera sin plantear completamente)

¿Para qué sirve medir la inteligencia?

- Cualquier campo científico o tecnológico tiene sus medidas:
  - Ejemplo:
    - Aeronáutica:
      - Medidas: Masa, Velocidad, Altura, Tiempo, Consumo, Carga, etc
      - “Volar”: se puede definir en términos de las medidas anteriores.
      - Diferentes sistemas especializados se pueden desarrollar teniendo en cuenta requerimientos de las diferentes medidas:
        - » Sistemas con mucha velocidad: aviones supersónicos.
        - » Sistemas con bajo consumo y poca carga: ultraligeros.
        - » Sistemas con tiempo de autonomía y alta carga: aviones de transporte.

¿Para qué sirve medir la inteligencia?

- ¿Por qué no se le han dedicado más esfuerzos?
  - Los sistemas “inteligentes” no biológicos tienen todavía capacidades muy limitadas.
  - Mala formulación de la “inteligencia artificial”:
    - “hacer que las computadoras realicen tareas que requerirían inteligencia si las hiciera un ser humano”.
  - Ciertos concursos (p.ej. Loebner test) han demostrado lo fácil que es engañar a ciertos tests.
  - Problema muy complejo.

Los Tests Psicométricos

- Objetivo: evaluar las capacidades intelectuales **humanas**:
  - Desarrollados por Binet y Spearman a finales del siglo XIX y principios del XX.
  - Aparece un índice conjunto, denominado IQ (*Intelligence Quotient*).
    - Se normaliza respecto a la media de la población.
  - Los tests se factorizan:
    - factor *g*, inteligencia líquida o general,
    - comprensión verbal,
    - capacidad espacial,
    - memoria,
    - capacidad de cálculo,
    - ...

Los Tests Psicométricos

- Características:
  - Fáciles de administrar:
    - Papel y lápiz.
    - No es necesario ningún juez durante la administración.
    - Es necesario poco tiempo (15’ – 30’).
  - Fiables:
    - La diferencia de dos tests sobre la misma persona (en situaciones similares) suele ser baja.
  - Pero no son *significativos*:
    - “la inteligencia es lo que miden los tests de inteligencia”.
    - se han desarrollado por prueba y error, y siempre sobre seres humanos...

Los Tests Psicométricos

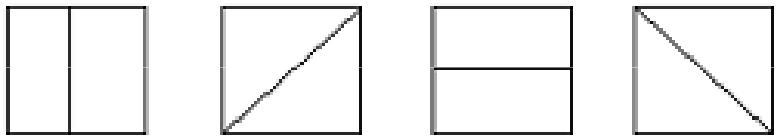
- Características:
  - Cualquier problema complejo requiere o se beneficia del uso de inteligencia para resolverlo.
  - ¿Cualquier problema complejo sirve como test de inteligencia?
    - La psicometría ha mostrado que no todo problema complejo vale para evaluar la inteligencia.
    - Para evaluar con fiabilidad se usan:

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.

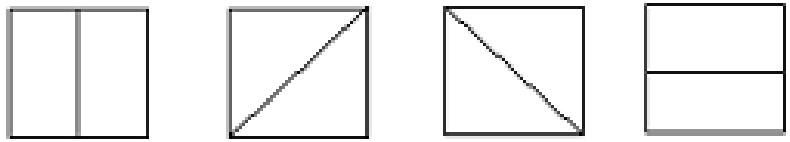
Los Tests Psicométricos

- Ejemplos:

Consider the sequence



Which one of the following will be next in the sequence?



A **B** **C** **D**

Complete the matrix

2	4	8
3	6	12
4	8	?

Los Tests Psicométricos

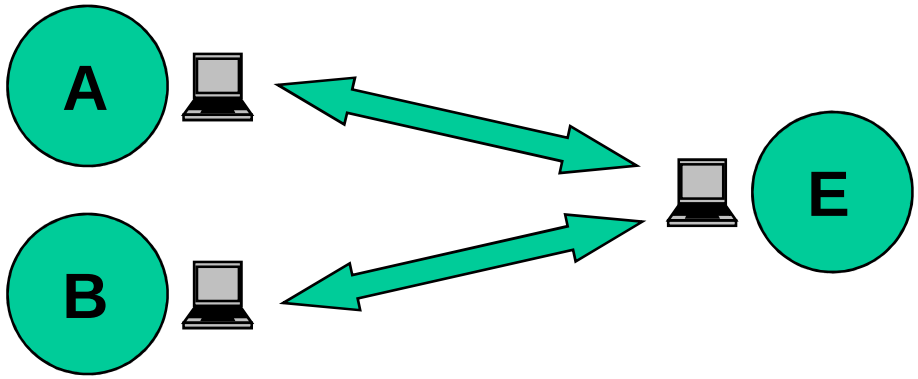
- ¿Son los tests psicométricos aplicables a sistemas inteligentes no humanos?
  - Aparentemente, ningún sistema informático "entiende" las preguntas.
  - Tampoco para animales directamente.
  - Además, dan por sentadas ciertas características que todos los seres humanos (normales) tienen.

⇕

Los tests psicométricos son antropomórficos

El Test de Turing

- Turing 1950: "*Computing Machinery and Intelligence*"
  - El Test de Turing es un Test de "Imitación".
    - El objetivo es intentar discernir entre dos individuos *A* y *B* (*A* siendo humano y *B* no) cuál es humano y cuál no.
    - Un evaluador externo *E* (también humano) puede preguntar a los dos a través de un terminal durante un cierto tiempo.
    - Si no puede discernir (o falla en la elección tantas veces como acierta), se concluye que *B* es "*inteligente*".



El Test de Turing

- Problemas:
  - El Test de Turing mide “humanidad”, no “inteligencia”.
  - Necesita dos humanos (*A* y *E*) para poder implementarlo. Uno de ellos debe ser un juez experimentado (*E*).
  - Requiere mucho tiempo.
  - La respuesta es sí o no.
    - Incluso con puntuaciones, tenemos paradojas:
      - No cabe el concepto de “más humano que”.

El Test de Turing

- “Loebner Prize”:
  - Un multimillonario ha organizado un concurso anual implementando el Test de Turing.

Home Page of The Loebner Prize--"The First Turing Test"

Loebner Prize Gold Medal

(Solid 18 carat, not *gold-plated* like the Olympic *"Gold"* medals)

El Test de Turing

- “Loebner Prize”:
  - Hay un conjunto de jueces (humanos) entrenados para tal fin.
  - Los “programas” han de intentar ser bien puntuados como “humanos” por los jueces.
  - Problemas:
    - Los programas se realizan específicamente para el concurso (no son capaces de nada más, ni de aprender a sumar).
    - Desprestigio de la comunidad científica. Se trata simplemente de un multimillonario que quiere pasar a la historia y no sabe cómo.

El Test de Turing

- "Loebner Prize". Ejemplo: Programa Joe:

JUDGE04 where do you work?
PROGRAM At the university. I mostly clean the Standish Building.
JUDGE04 What university?
PROGRAM University of Eastern Ontario. I'm a cleaner. I empty trash. Vacuum. Dust. You know.
JUDGE04 and why do they want to fire you?
PROGRAM It's just because my boss, Mr. Johnson, is mad at me.
JUDGE04 why is he mad at you?
PROGRAM I didn't do nothing wrong. But he thinks I did. It's all because of that radio they think I stole from the lab.
JUDGE04 that's too bad. are you in a union?
PROGRAM What would I tell the union? They won't say they're firing me because I stole the radio. They'll just make up some excuse

El Test de Turing

- “Loebner Prize”. Ejemplo: Programa Joe:

JUDGE02 what does NATO stand for?
PROGRAM I don’t know.
JUDGE02 what’s your name?
PROGRAM What’s your name?? I don’t know. What do I look like? Some kind of professor?
JUDGE02 What was Mark Twain’s real first name?
PROGRAM Huh?
JUDGE02 what does huh mean?
PROGRAM Beats me.
JUDGE02 huh?
PROGRAM What do you mean?
JUDGE02 beats me.
PROGRAM What?
JUDGE02 huh.
PROGRAM I don’t know.

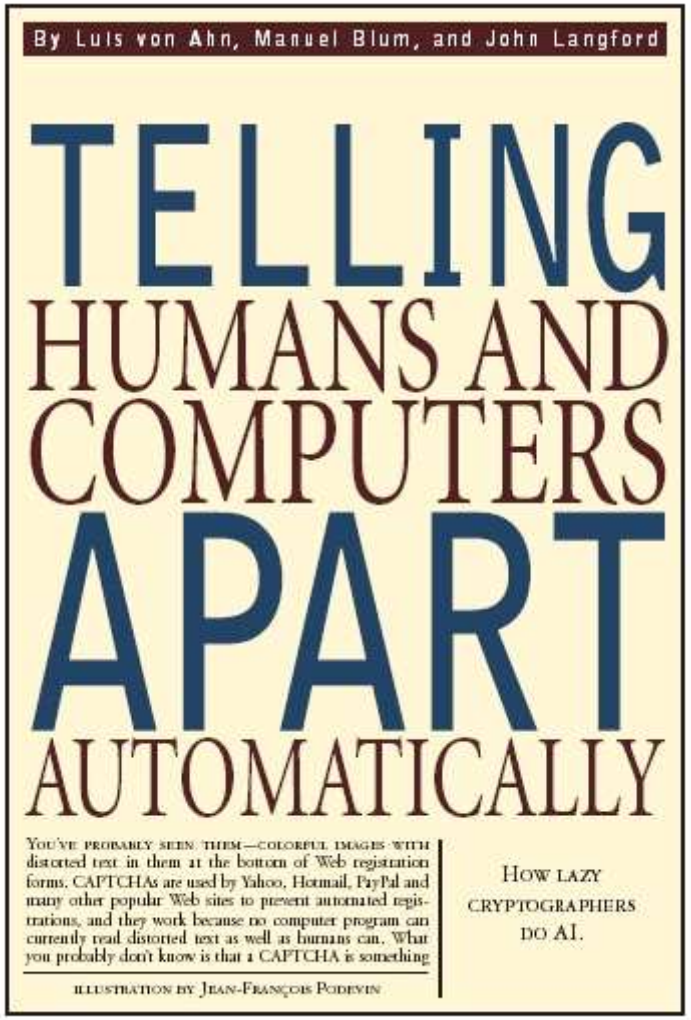
El Test de Turing

- Variaciones del Test de Turing:
  - Se han desarrollado muchas variantes: TTT, RTT, ...
    - Dowe & Hajek [Dowe & Hajek 1997, 1998a, 1998b] include induction inference problems into the TT to make it more robust.
  - Turing sólo pretendía zanjar las dudas sobre inteligencia no humana.
    - Su artículo "*Computing Machinery and Intelligence*" es una obra maestra en este sentido.
  - No se le puede culpar de que:
    - Se malinterprete.
    - Se quiera utilizar el test de Turing como un "test práctico" y fiable.

Los Tests “captcha”

- Proyecto de Carnegie Mellon (www.captcha.net) (2000-2004):

- Presentan problemas que los sistemas de I.A. no son capaces de reconocer hoy en día.



By Luis von Ahn, Manuel Blum, and John Langford

TELLING HUMANS AND COMPUTERS APART AUTOMATICALLY

You've probably seen them—colorful images with distorted text in them at the bottom of Web registration forms. CAPTCHAs are used by Yahoo, Hotmail, PayPal and many other popular Web sites to prevent automated registrations, and they work because no computer program can currently read distorted text as well as humans can. What you probably don't know is that a CAPTCHA is something

How lazy cryptographers do AI.

Illustration by Jean-François Podevin

Communications of the ACM 57

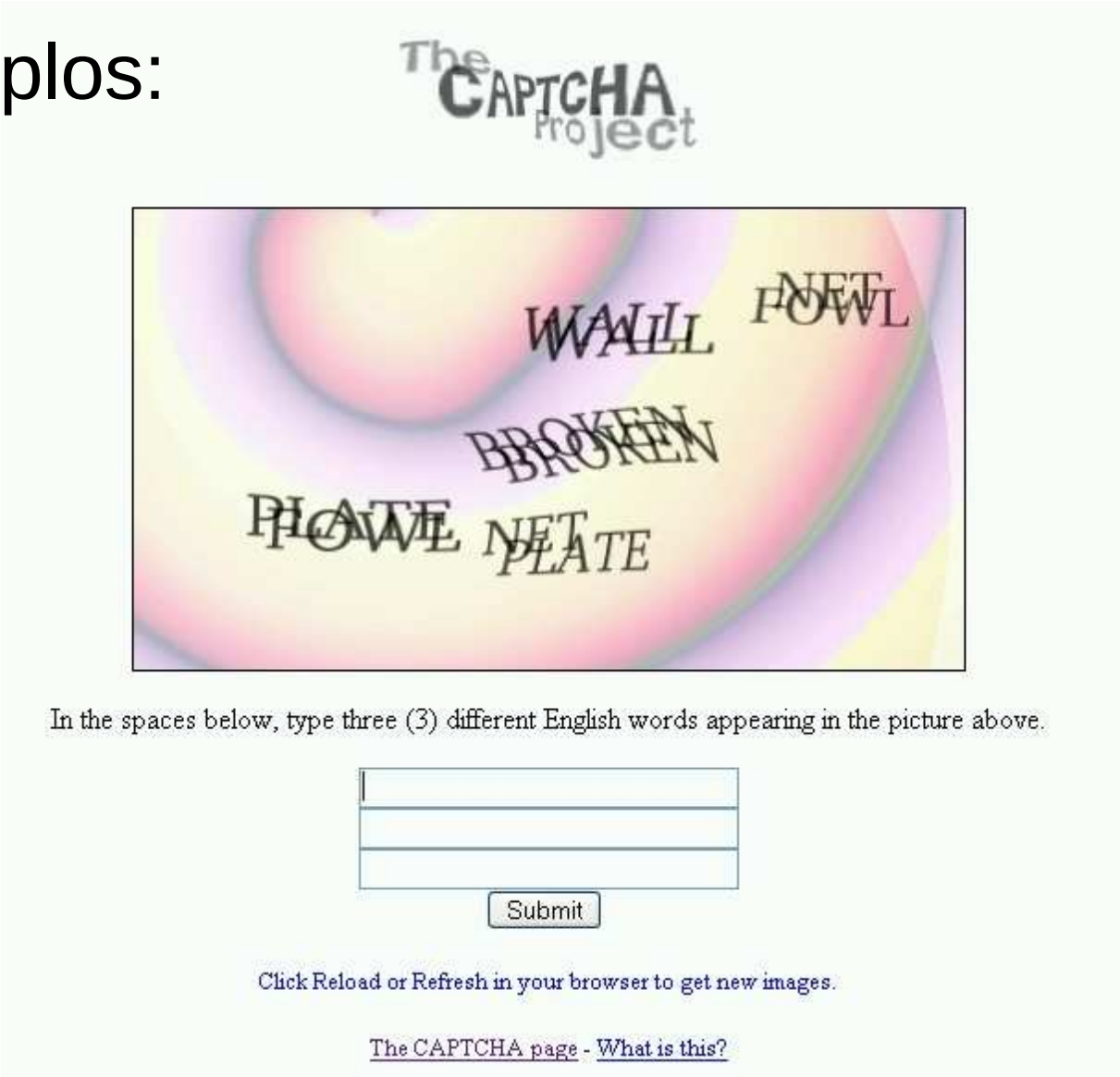
Los Tests “captcha”

- Los problemas:
  - Se generan automáticamente.
  - Se evalúan automáticamente.
  - Son fiables.

- Un *captcha* es un programa que puede generar problemas que él mismo no puede resolver, pero la mayoría de los seres humanos sí.

Los Tests “captcha”

- Ejemplos:

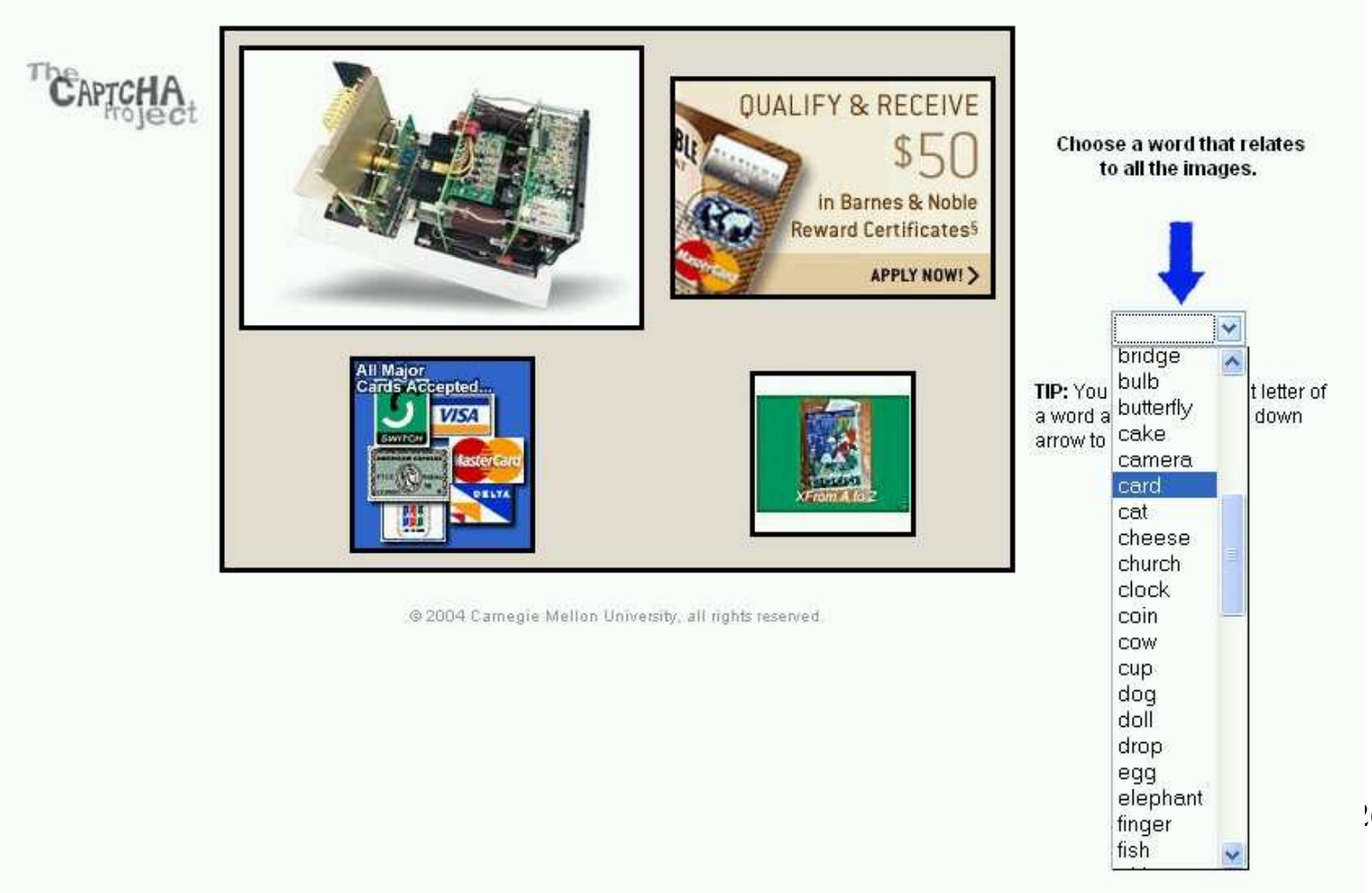


Los Tests “captcha”

- Más ejemplos:

Los Tests “captcha”

- Más ejemplos:

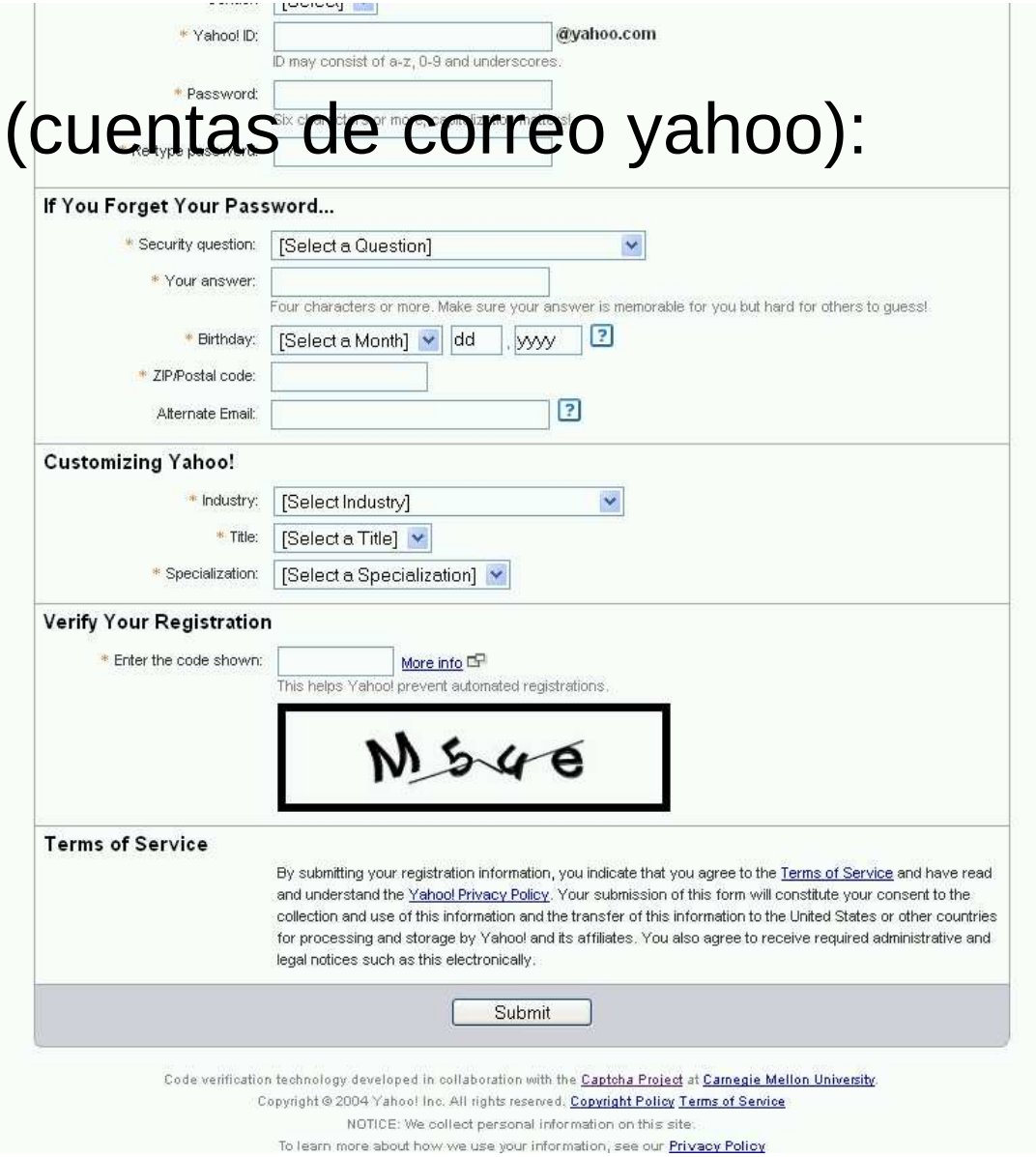


Los Tests “captcha”

- Usos:
  - Detección de “bots”:
    - Servicios de e-mail gratuitos.
    - “Bots” de búsqueda.
    - Agentes de compras online.
    - Votaciones online.
    - Sistemas de reserva.
    - Seguridad (p.ej. evitar ataque de diccionario para averiguar contraseñas).
    - Worms y Spam.
  - Constituyen retos para la inteligencia artificial.
    - Los tests se basan en problemas que los programas actuales nos son capaces de resolver.
    - A medida que se van resolviendo se sustituyen por otros más complejos.

Los Tests “captcha”

- Ejemplo (cuentas de correo yahoo):



Los Tests “captcha”

- Ventajas:
  - Son prácticamente instantáneos.
  - Relativamente fiables.
- Problemas:
  - Relativos a la tecnología del momento.
    - No son genéricos.
    - Deben ser sustituidos cada cierto tiempo.
    - Se pueden hacer programas específicos para pasar los problemas típicos del *captcha*.
  - Pueden ser engañados reenviándole el problema a un humano (problema de la aplicación, no del test).
  - No dan una puntuación (y es difícil diseñar una puntuación razonable).

Requerimientos de Tests de Inteligencia

- Vistas las aproximaciones anteriores:

¿Qué requerimientos son necesarios para unos tests de inteligencia válidos, útiles y generales?

- Requerimientos:
  - No antropomórficos.
  - Graduales.
  - Factoriales.
  - Computacionales.
  - Significativos.
  - Fiables.
  - Prácticos.

Requerimientos de Tests de Inteligencia

- No antropomórfico:
  - Debe servir para medir sistemas inteligentes en general (humanos o no): seres humanos, computadoras, otros animales, extraterrestres, ...
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: No. Diseñados para humanos.
    - Test de Turing: No. Es un test de “humanidad”
    - Tests *captcha*: No. Se ceban en las incapacidades de los “no humanos”.
  - *Todos usan el ser humano como “pie de rey”.*

Requerimientos de Tests de Inteligencia

- Graduales:
  - No se deben limitar a decir sí o no. Deben dar grados o niveles de inteligencia ...
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: Sí. Dan un “IQ” con una escala centrada en 100 (70-130 normal).
    - Test de Turing: No. Sólo distinguen entre “humanos” y “no humanos”.
    - Tests *captcha*: No. Sólo distinguen entre “humanos” y “bots”.

Requerimientos de Tests de Inteligencia

- Factoriales:
  - Deben medir diferentes factores y habilidades...
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: Sí. Se han determinado diferentes factores: cálculo, lingüístico, ...
    - Test de Turing: No. Sólo examina la “humanidad”.
    - Tests *captcha*: No. Aunque en realidad se basa en un *abanico* de habilidades que las computadoras no son capaces de hacer.

Requerimientos de Tests de Inteligencia

- Computacionales:
  - Deben estar definidos matemáticamente, con una formulación precisa y clara de lo que se está midiendo...
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: No. “La inteligencia es lo que miden los tests de inteligencia”
    - Test de Turing: No. Parte de la categoría “humano”, no definida por capacidades sino biológicamente.
    - Tests *captcha*: No. Se basan en el estado de la tecnología.

Requerimientos de Tests de Inteligencia

- Significativos:
  - Aquello que se está midiendo debe representar algo relacionado con nuestro concepto primigenio de inteligencia (capacidad de comprender)...
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: No. “La inteligencia es lo que miden los tests de inteligencia”
    - Test de Turing: Parcialmente. Un sujeto es inteligente si es capaz de “imitar” bien.
    - Tests *captcha*: No. Se basan en el estado de la tecnología.

Requerimientos de Tests de Inteligencia

- Fiables:
  - El mismo sujeto debe obtener resultados similares en distintos pases del test... El test debe ser difícil de engañar.
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: Sí.
    - Test de Turing: Parcialmente. Pero para ello requieren muchos jueces y tiempo.
    - Tests *captcha*: Parcialmente. Algunos sistemas pueden pasar algunos test y otros no. Algunos humanos fallan.

Requerimientos de Tests de Inteligencia

- Prácticos:
  - Deben ser fáciles de administrar (sin necesidad de jueces inteligentes, tiempos muy largos, equipos sofisticados, conocimiento común, etc...)
  - ¿Cumplen el requisito los tests vistos anteriormente?
    - Test psicométricos: Sí. Papel, bolígrafo y unos minutos. Aunque requieren entender el idioma del test (para algún factor).
    - Test de Turing: No. Se requieren muchos jueces y tiempo.
    - Tests *captcha*: Sí. Prácticamente instantáneos. Requieren saber reconocer caracteres escritos (p.ej. no válidos para invidentes).

Tests basados en Compresión

- La complejidad Kolmogorov sirve como herramienta matemática para la formalización de la inferencia inductiva (aprendizaje inductivo) (Solomonoff 1964).

Inducción = Compresión de Información

- Se plantea la posibilidad de usar la complejidad de Kolmogorov para realizar tests de inteligencia (Chaitin 1982).
  - Se plantea completar el Test de Turing añadiendo la restricción de compresión de secuencias (Dowe & Hájek 1997, 1998)
- Pero, ¿cómo conseguir obtener *instancias* de problemas para evaluar la capacidad de inducción?

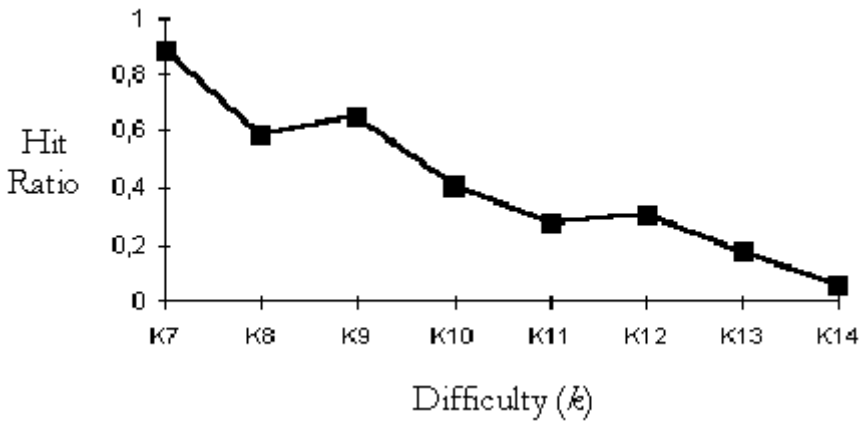
Tests basados en Compresión

- Generación de instancias de diferente complejidad [Hernández & Minaya 1998] [Hernández 2000a]
  - Se necesitan variantes computables de la complejidad Kolmogorov.
  - Es necesario generar instancias tal que exista una "explicación" significativamente más simple que todas las demás, para que la predicción sea relativamente *incuestionable*
- Se generaron tests utilizando series de diferente complejidad. P.ej.:

  "a,c,c,a,c,c,c,a,c,c,c,c,a,..."
- La respuesta viene definida de la generación de la cadena. Su complejidad es estimada formalmente.

Tests basados en Compresión

- El test se administró a 65 personas [Hernández & Minaya 1998].
  - Los resultados correlacionan con los tests clásicos de inteligencia (IQ tests).
  - La medida de complejidad obtenida para diferentes series representa fielmente la dificultad de la instancia:



Tests basados en Compresión

- El test se basa en conceptos puramente computacionales.
  - Los ejercicios no se seleccionan por capricho de un psicólogo o del que diseña el test, ni por su funcionamiento anterior en otro sujeto.
  - Los ejercicios se derivan de una fórmula matemática.
  - Se puede determinar la complejidad de las instancias a priori.
  - El “test” mide la capacidad de encontrar “explicaciones” inductivas.

Tests basados en Compresión

- ¿Qué requisitos cumple?
  - No antropomórfico. ✓
  - Gradual. ✓
  - Factorial. ×
  - Computacional. ✓
  - Significativo. ✓
  - Fiable. ✓ *(al menos con los humanos)*
  - Práctico. ✓

Factorización

- ¿La inteligencia es sólo capacidad de inferencia inductiva?
  - ¿Necesitamos más factores?
    - Factores deductivos.
    - Factores contextuales.
    - ...
  - En [Hernández 2000b] se formaliza la idea de complejidad de instancia de un problema dado.
    - Se pueden generar tests con la misma filosofía que los anteriores que contemplen diferentes factores.

Factorización

- Pero…
  - La psicometría todavía discute la existencia de factores:
    - Según algunos autores, es todo simplificable al factor “g”, o inteligencia “líquida”.
      - Los test inductivos anteriores miden precisamente ese factor.
    - Otros autores, optan por varios factores e incluso varios tipos de “inteligencias”…
  - ¿No sería suficiente con el factor “g”?

Factorización

¿Sorpresa? Computadoras pasando tests IQ

- (Sanghi & Dowe 2003) han implementado un programa (en Perl) que es capaz de pasar con éxito la mayoría de tests IQ (alrededor de una puntuación de 100).

Este programa no es inteligente

Test	I.Q. Score	Human Average
A.C.E. I.Q. Test	108	100
Eysenck Test 1	107.5	90-110
Eysenck Test 2	107.5	90-110
Eysenck Test 3	101	90-110
Eysenck Test 4	103.25	90-110
Eysenck Test 5	107.5	90-110
Eysenck Test 6	95	90-110
Eysenck Test 7	112.5	90-110
Eysenck Test 8	110	90-110
I.Q. Test Labs	59	80-120
Testedich.de:I.Q. Test	84	100
I.Q. Test from Norway	60	100
Average	96.27	92-108

Factorización

¿Sorpresa? Computadoras pasando tests IQ

- ¿Qué significa esto?
  - Los tests IQ no miden inteligencia, **en general**. Están especializados para medir la inteligencia *humana* de una manera práctica.
    - Utilizan la correlación existente entre esta habilidad (comprimir secuencias) y la inteligencia general.
  - Los tests basados en compresión son similares a los tests IQ y también podrían ser realizados con éxito por programas no inteligentes.

¿Qué falta en los tests?

Factorización

- Necesidad de “Factorización”: abanico de capacidades.
  - Se puede factorizar mal:
    - Factores muy específicos y arbitrarios.
    - Podría ampliarse el programa para pasar varias habilidades.
  - Riesgo de disgregar en un conjunto variopinto de tareas…
- ¿Cómo factorizar bien?
  - Limitarse a procesos básicos de inferencias y razonamiento.
  - Aumentar la “profundidad” de las habilidades.
    - Contextualización: necesidad de relacionar las instancias. Los problemas deben necesitar lo “aprendido” anteriormente.
    - Otras características “sociales”: Interactividad y comunicación.
  - Siempre calcular dependencias (teóricas) y correlaciones (experimentales) entre factores. Si son dependientes, eliminar.

Factorización

- Una vez factorizado adecuadamente:
  - Si añadimos más habilidades y, especialmente habilidades de contextualización, ¿podremos hacer tests prácticos, aplicables en poco tiempo?
  - ¿Servirán como tests generales?
  - Por ejemplo:
    - ¿Se pueden utilizar como *captchas*?
      - Los problemas tendrían que ser relativamente fáciles.
    - ¿Se podrán utilizar para evaluar capacidades de sistemas inteligentes en general?
      - No deben asumir nada del conocimiento previo.
      - Todos los factores: proceso lento de evaluación.

Fiabilidad y Aplicabilidad

- Conocimiento “común”:
  - El conocimiento común entre “examinado” y “examinador/prueba” facilita la tarea de evaluación.
    - Necesario para algunas “habilidades”, p.ej. lingüísticas.
  - Permite expresar tareas más complejas
  - Permite respuestas más complejas.
  - Pero, si existe conocimiento común, el test es menos imparcial.

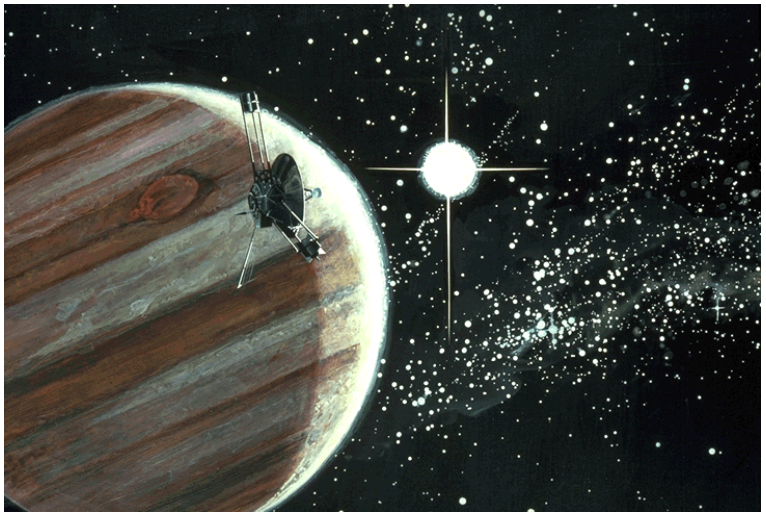
Fiabilidad y Aplicabilidad

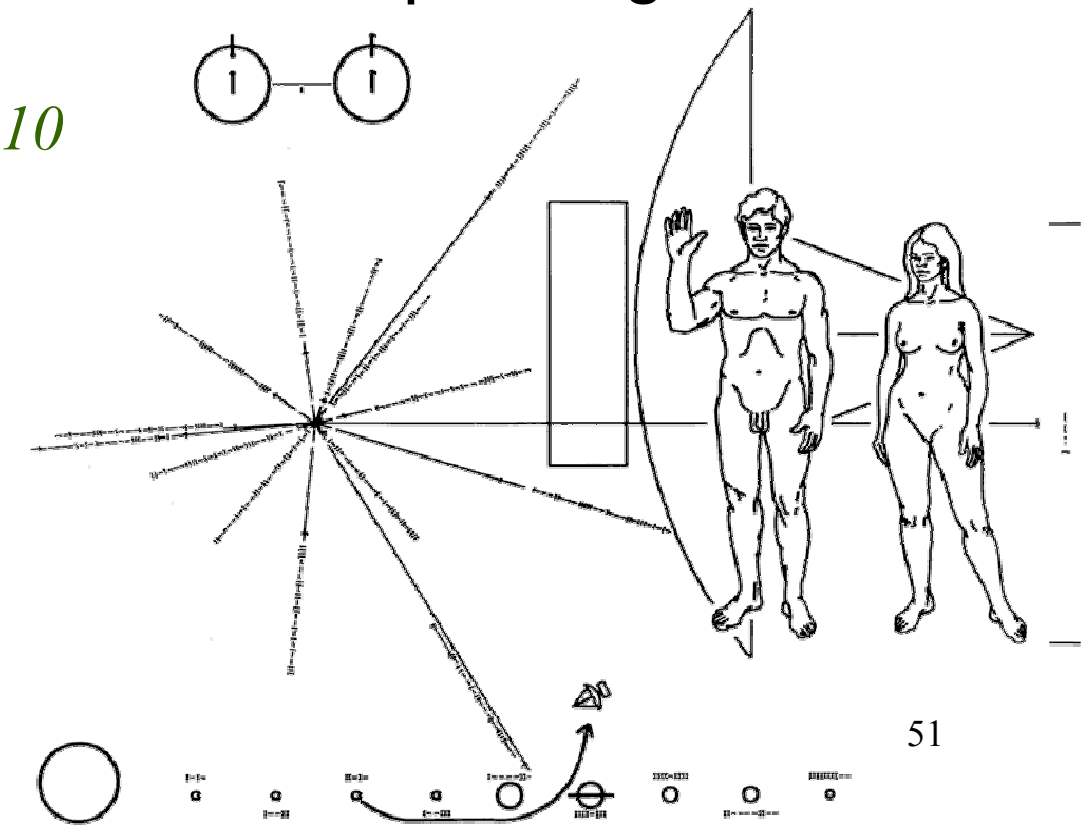
- Sin conocimiento “común”:
  - Ningún conocimiento ni lenguaje común:
    - Debe basarse en algún tipo de test interactivo en el que se “premia o castiga” el resultado de cada ejercicio.
      - Los ejercicios se pueden ir basando en ejercicios anteriores.
  - Criterio de construcción:
    - Pensar que se realiza un test para extraterrestres.

Fiabilidad y Aplicabilidad

- Sin conocimiento “común”.
  - Aproximaciones: referirse a algo “externo” perceptible por ambos: una placa, galaxia, …

Placa de la sonda Pioneer 10





Fiabilidad y Aplicabilidad

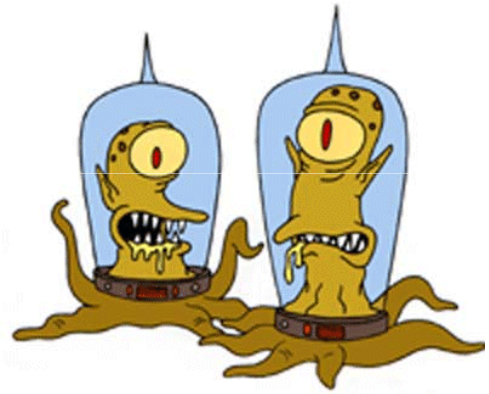
- Sin conocimiento “común”:
  - Evaluar resulta ser más sencillo que comunicar.
    - Habría que nombrar objetos matemáticos e ir creando un lenguaje…
  - Evaluar se puede basar en conceptos simbólicos.

Fiabilidad y Aplicabilidad

- Sin conocimiento “común”:

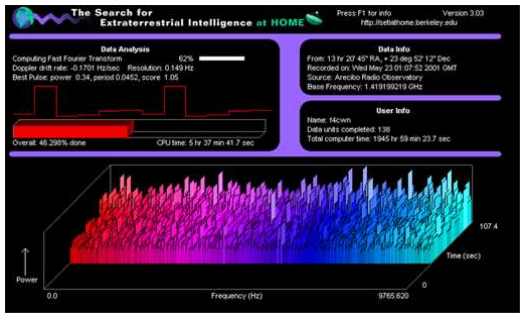
10101010101010101?

01010...



11101010001010001 01?

00010..



Fiabilidad y Aplicabilidad

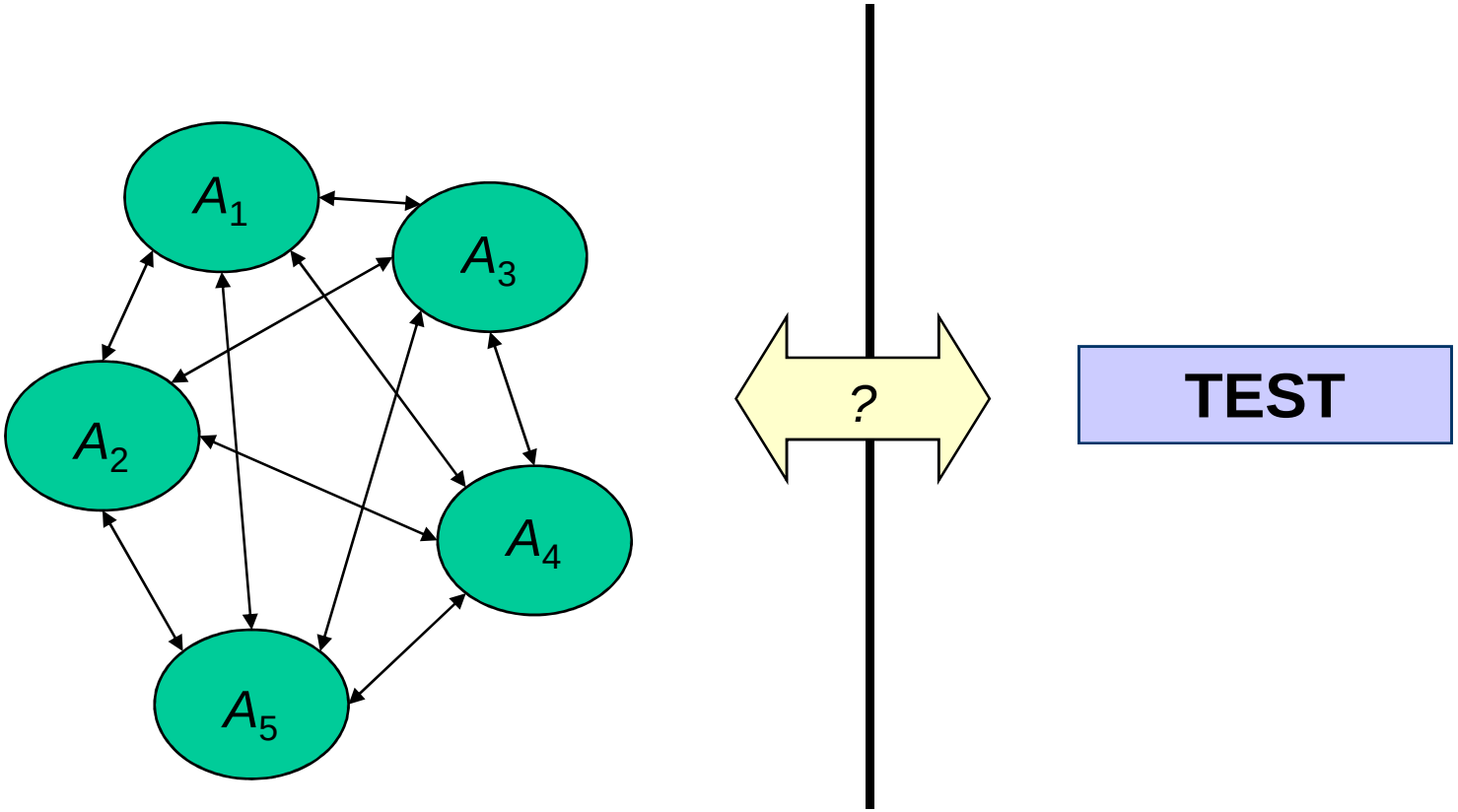
- Características:
  - Tanto en un caso como en otro
    - Si la prueba se va ajustando a las respuestas, se puede acelerar la tarea de evaluación.
      - Primero preguntas sencillas, luego cuestiones más complejas, más dependientes de problemas anteriores (contextuales).
    - El tiempo de la prueba depende de la velocidad del examinado
      - La velocidad del examinado es un factor de la inteligencia a tener en cuenta.

Fiabilidad y Aplicabilidad

- La fiabilidad del test dependerá:
  - Del conocimiento común que se asuma.
  - Del número y variedad de ejercicios.
- La fiabilidad limita la aplicabilidad:
  - Habrá que buscar compromisos.
    - Asumir parte de conocimiento común.
    - Adaptar el test según las respuestas del examinado.
    - Medir sólo algunos factores.

Evaluando Comunidades

- ¿Es válido lo anterior para evaluar una comunidad de agentes?

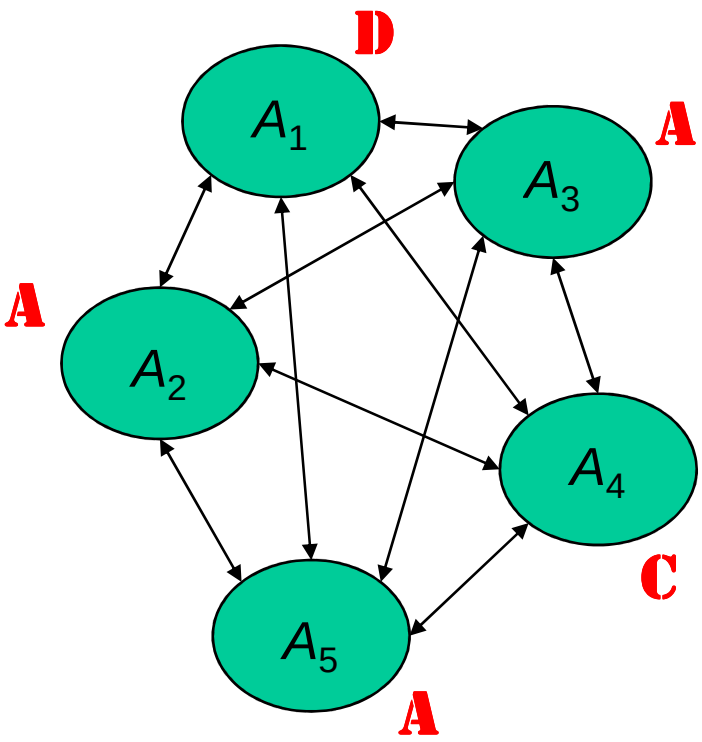


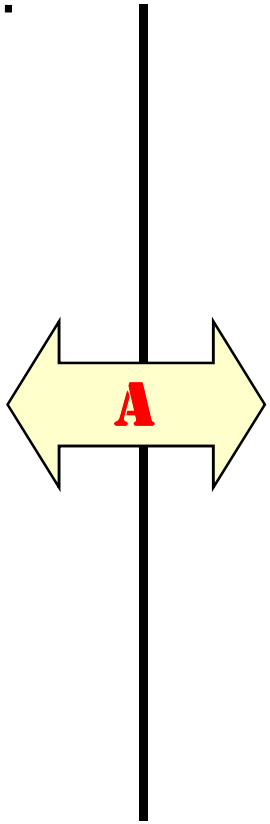
Evaluando Comunidades

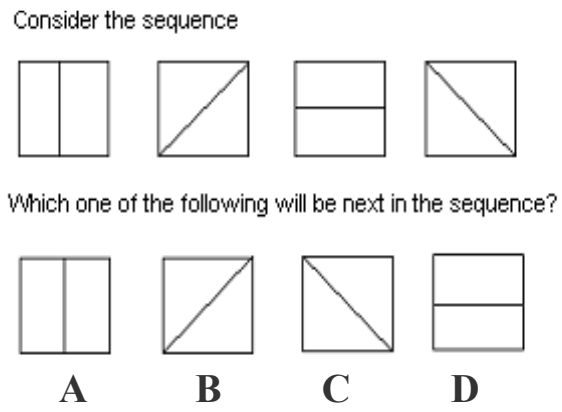
- ¿Cómo se evalúan?
  - Responden al test:
    - Por votación.
      - El resultado global se limita a una combinación estadística
    - Por votación ponderada.
      - ¿Cómo ponderamos? Agentes más fiables (líderes), …
    - Por discusión.
      - Requiere comunicación entre los agentes.
    - Por especialización.
      - Requiere comunicación entre los agentes.

Evaluando Comunidades

- ¿Cómo se evalúan?
  - Responden al test:







Evaluando Comunidades

- ¿Qué relación existe entre las inteligencias de los individuos y las inteligencias del conjunto?
  - ¿Es siempre igual o superior la inteligencia del conjunto?
    - Depende del esquema de colaboración.
    - En algunos esquemas (con bastantes suposiciones) se puede mostrar que la capacidad puede aumentar.
      - Muy relacionado con las técnicas de combinación de modelos en aprendizaje automático.
      - Relacionado también con la visión de que los sistemas inteligentes han de ser “masivamente” paralelos.

Evaluando Comunidades

- ¿Qué significa evaluar la inteligencia de una comunidad?
  - Desmitifica la idea de “inteligencia”.
  - Cualquier sistema inteligente está compuesto de subsistemas, que pueden no ser inteligentes o, mejor dicho, “menos” inteligentes.
    - Visión constructivista.
  - Separa inteligencia y conciencia, generalmente indivisible en los seres humanos.
    - Clarifica el problema desde un punto mucho más “objetivable” y menos “introspectivo”.

Evaluando Comunidades

- EJERCICIO:
  - 111010100010100010100010 ?
    - Votación
      - (sin hablar entre sí).
    - Votación ponderada.
      - (sin hablar entre sí). Los “profes” cuentan doble.
    - Con discusión.

Conclusiones

- El problema de la medición de inteligencia no humana es un problema *abierto*.
  - Está empezando a ser relevante en ciertas aplicaciones.
    - El proyecto *captcha* ha mostrado la utilidad que puede tener.
  - Será cada vez más importante.
    - Sistemas adaptativos/inteligentes en cada vez más ámbitos y aplicaciones…
    - Sistemas con mayores capacidades.
      - Los tests simplistas dejan de ser útiles…

Conclusiones

- Sabemos algunas cosas:
  - Sabemos lo que no funciona.
  - Sabemos que el concepto de inteligencia es multidimensional/multifactorial
  - Se ha de desligar de consideraciones filosóficas y atacar el problema desde un punto de vista aséptico, ingenieril (teoría y práctica).
  - Sabemos que ciertas cosas son *necesarias* (capacidades de inferencia inductiva / compresión), pero no sabemos el conjunto de habilidades / factores *suficiente*.

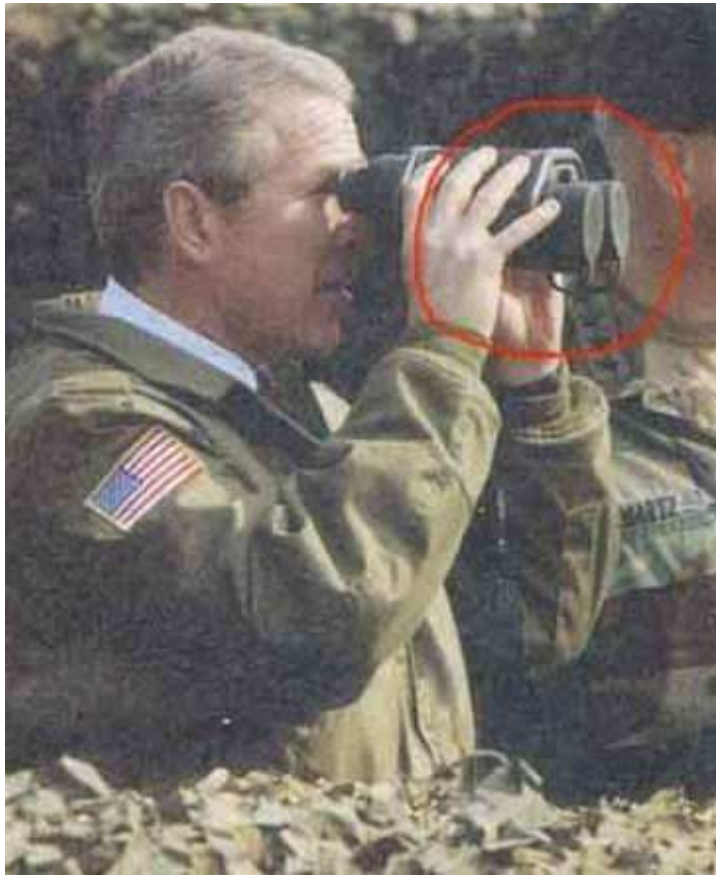
Conclusiones

- A medida que los sistemas inteligentes se vayan acercando a la inteligencia humana, más necesarios será poder medirla y factorizarla:
  - Para evaluar el progreso.
  - Como parte de las especificaciones.
  - Para resolver problemas éticos.

Conclusiones

- ¿Preguntas?

¿Presidente Inteligente?



Conclusiones

- ¿Más preguntas?

¿Nevera Inteligente?

Contacto

- Más Información, transparencias, artículos relacionados:

  http://www.dsic.upv.es/~jorallo/

- Dudas, críticas y spam:

  jorallo@dsic.upv.es

- Agradecimientos:
  - David Dowe, Monash University.

Referencias

- Spearman, C. " 'General Intelligence' objectively determined and measured" *American Journal of Psychology* 15, 201-293, 1904.
- Turing, A.M. "Computing machinery and intelligence" *Mind* 59, 433-460, 1950.
- Solomonoff, R.J. "A formal theory of inductive inference" *Information & Control* 7, 1-22, March, 224-254, June, 1964.
- Chaitin, G.J. "Gödel's theorem and information" *International Journal of Theoretical Physics*", 21, 941-954, 1982.
- Dowe, D.L. & Hajek, A.R. "A computational extension to the Turing Test" *Technical Report #97/322*, Department of Computer Science, Monash University, Australia, 1997.
- Dowe, D.L. & Hajek, A.R. "A non-behavioural, computational extension to the Turing Test" *Proceedings of the International Conference on Computational Intelligence & Multimedia Applications (ICCIMA '98) (pp-101-106) World Scientific, Australia* 1998.

Referencias

- Hernández, J.; Minaya, N. “A Formal Definition of Intelligence Bases on an Intensional Variant of Algorithmic Complexity” *Proceedings of International Symposium of Engineering of Intelligent Systems* (EIS’98), pp 146-163, Tenerife, Spain, 1998.
- Hernández, J. “Beyond the Turing Test” *Journal of Logic, Language and Information* 9: 447-466, 2000.
- Hernández, J. “On the Computational Measuremente of Intelligence Factors” *Workshop on Performance Metrics for Intelligent Systems, NIST, Gaithersburg, August 14-16*, 2000.
- von Ahn, L.; Blum, M.; Langford, J. “Telling Humans and Computers Apart (Automatically) or How Lazy Cryptograhpers do AI” Communications of the ACM, 2002. www.captcha.net.
- Sanghi, P. & Dowe, D.L. “A computer program capable of passing I.Q. tests” a aparecer, 2003.