# Turing Machines and Recursive Turing Tests

**José Hernández Orallo[1], Javier Insa-Cabrera[1], David L. Dowe[2], Bill Hibbard[3],**

1. *Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Spain.*
2. *Computer Science & Software Engineering, Clayton School of I.T., Monash University, Clayton, Victoria, 3800, Australia.*
3. *Space Science and Engineering Center, University of Wisconsin - Madison, USA*

**CQRW2012 - AISB/IA-CAP 2012 World Congress, July 4-5,** Birmingham, UK

# Outline

- The Comparative Approach
- Computational Measurement of Intelligence
- Reunion: bridging antagonistic views
- Base case: the TT for TMs
- Recursive TT for TMs
- Discussion

# The comparative approach

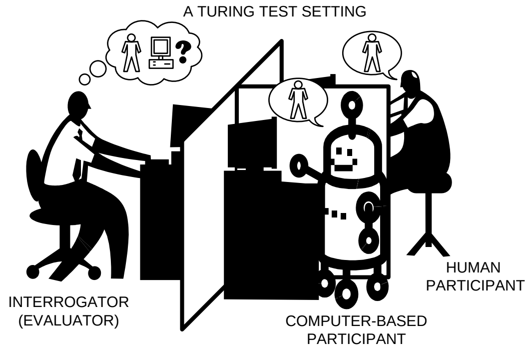
- Intelligence Evaluation:
  - Intelligence has been evaluated by humans in all periods of history.
  - Only in the XXth century, this problem has been addressed *scientifically*:
    - Human intelligence evaluation is performed and studied in psychometrics and related disciplines.
    - Animal intelligence evaluation is performed and studied in comparative cognition and related disciplines.

What about machine intelligence evaluation?

  - We only have partial approaches in some AI competitions and, of course, some variants and incarnations of the Turing Test.

# The comparative approach

- Turing Test:
  - The *imitation game* was not really conceived by Turing as a *test*, but as a compelling argument.




  - Problems of using the imitation game as a test of intelligence.
    - Humanity (and not intelligence) is taken as a reference.
    - Evaluation is subjective: evaluators are also humans.
    - Too focussed on (teletype) *dialogue*.
    - Not based on reproducible tasks but on particular, unrepeatable conversations.
    - Not really scalable far below or beyond human intelligence.
    - Not clear how it behaves for collective intelligence (with one teletype communicator).

Is there an alternative principled way of measuring intelligence?

# Computational measurement of intelligence

- During the past 15 years, there has been a discreet line of research advocating for a formal, computational approach to intelligence evaluation.
  - Issues:
    - Humans cannot be used as a reference.
      - No arbitrary reference is chosen. Otherwise, comparative approaches would become circular.
    - Intelligence is a gradual (and most possibly factorial) thing.
      - It must be graded accordingly.
    - Intelligence as performance on a diverse tasks and environments.
      - Need to define these tasks and environments.
    - The difficulty of tasks/environments must be assessed.
      - Not on populations (psychometrics), but from computational principles.

# Computational measurement of intelligence

- Problems this line of research is facing at the moment.
  - Most approaches are based on tasks/environments which represent patterns that have to be discovered and correctly employed.
  - These tasks/environments are not representative of what an intelligence being may face during its life.
  - Environments lack on evaluate some skills that discriminates better between different systems.

(Social) intelligence is the ability to perform well in an environment full of other agents of similar intelligence

# Computational measurement of intelligence

- This definition of Social intelligence prompted the definition of a different distribution of environments:
  - **Darwin-Wallace distribution** (Hernandez-Orallo et al. 2011): environments with intelligent systems have higher probability.
    - It is a *recursive* (but not circular) distribution.
    - Use agents' intelligence to create new social environments.
    - While resembles artificial evolution, it is guided and controlled by intelligence tests, rather than selection due to other kind of fitness.

# Reunion: bridging antagonistic views

- The setting of the Darwin-Wallace distribution suggests:
  - Comparative approaches may not only be useful but necessary.
  - The Turing Test might be more related to social intelligence than other kinds of intelligence.
- This motivates a reunion between the line of research based on computational, information-based approaches to intelligence measures with the Turing Test.
  - However, this reunion has to be made without renouncing to one of the premises of our research: the elimination of the human reference.

**Use (Turing) machines, and not humans, as references.**
**Make these references meaningful by recursion**

# Base case: the TT for TMs

- The Turing Test makes some particular choices:
  - Takes the human reference from a distribution: adult homo sapiens.
  - Takes the judges from a distribution (also adult homo sapiens) but they are also instructed on how to evaluate.

- But other choices can be made.
  - Informally?
    - A Turing Test for Nobel laureates, for children, for dogs or other populations?
  - Formally? Generally?
    - Nothing is more formal and general than a Turing Machine.

# Base case: the TT for TMs

- Let us generalise the TT with TMs:

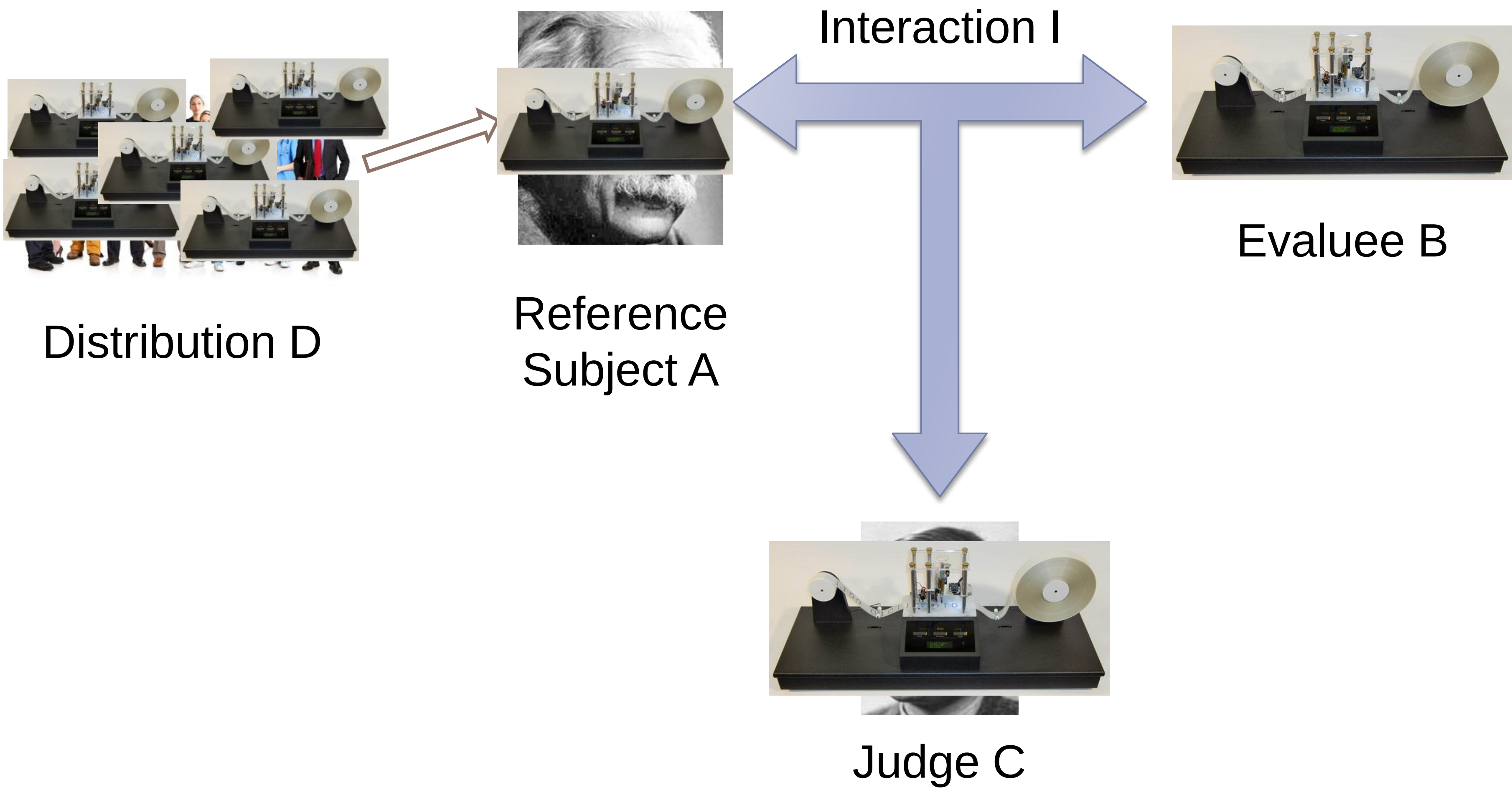
**Definition 1** *The imitation game for Turing machines is defined as a tuple* $\langle D, B, C, I \rangle$

- *The reference subject* $A$ *is randomly taken as a TM using a distribution* $D$.
- *Subject* $B$ *(the evaluee) tries to emulate* $A$.
- *The similarity between* $A$ *and* $B$ *is 'judged' by a criterion or judge* $C$ *through some kind of* interaction *protocol* $I$. *The test returns this similarity.*

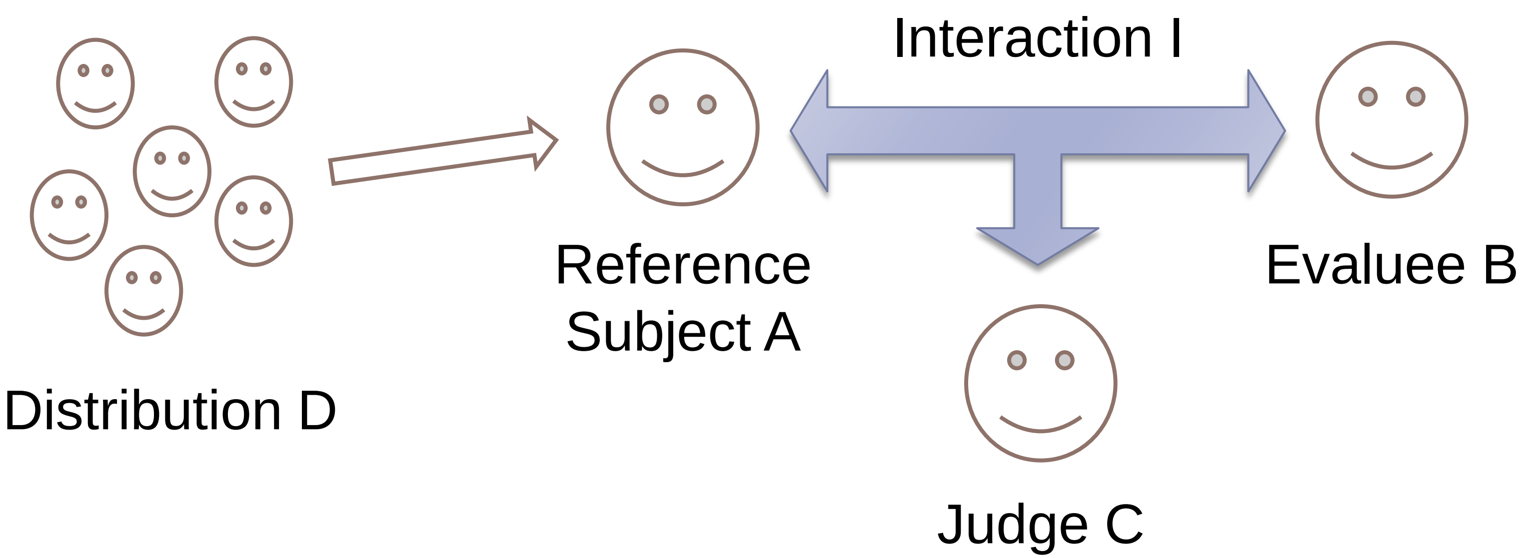
# Base case: the TT for TMs

- The use of *Turing* machines for the reference is relevant:
  - We can actually define formal distributions on them (this cannot be done for humans, or animals or "agents").
- It is perhaps a convenience for the judge.
  - Any formal mechanism would suffice.
- It is not exactly a generalisation, because in the TT there is an *external reference*.
  - the judge compares both subjects with his/her knowledge about human behaviour.

# Base case: the TT for TMs

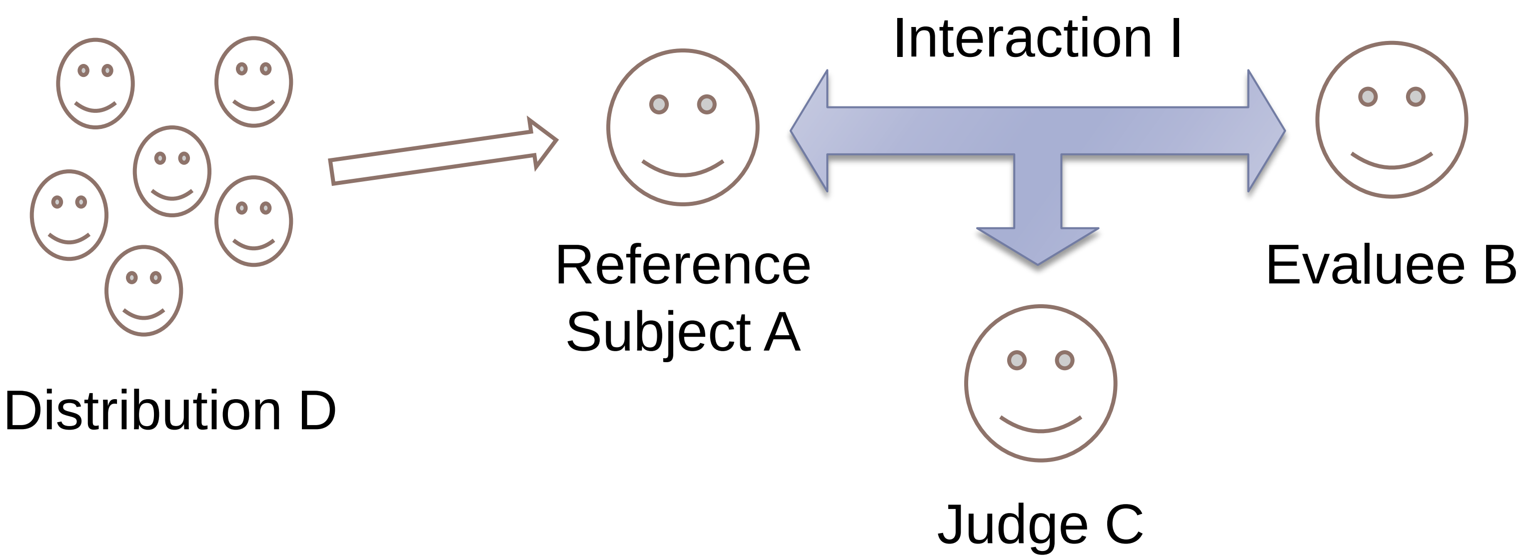


# Base case: the TT for TMs




- The C-test can be seen as a special case of the TT for TMs:
  - The reference machines have no input (they are static)
  - The distribution gives high probability to sequences of a range of difficulty (Levin's Kt complexity).
  - The judges/evaluation just look for an exact matching between the reference outputs and the evaluee.

# Base case: the TT for TMs




- Legg & Hutter's Universal Intelligence can be seen as a special case of the TT for TMs:
  - The reference machines are interactive and issue rewards.
  - The distribution gives high probability to TMs with low Kolmogorov complexity.
  - The judges/evaluation just look for high rewards.

# Base case: the TT for TMs

- Other more ‘orthodox’ versions could be defined:
  - Question-answer setting:
    - Judges just issue questions from a distribution (they are string-generating TM).
    - Reference A is another TM which receives the input and issues an output.
    - The evaluee learns from the input-outputs over A and tries to imitate.
  - However, the original version of the TT was adversarial.
    - Reference subjects were instructed to play against the evaluee (and vice versa). Both wanted to be selected as *authentic*.
      - However, we do not have an external reference.

# Base case: the TT for TMs

- The simplest adversarial Turing Test:
  - Symmetric roles:
    - Evaluee B tries to imitate A. It plays the *predictor* role.
    - Reference A tries to evade B. It plays the *evader* role.
  - This setting is exactly the matching pennies problem.
    - Predictors win when both coins are on the same side.
    - Evaders win when both coins show different sides.

| | | Player 2 | |
|---|---|---|---|
| | | Heads | Tails |
| Player 1 | Heads | 1,-1 | -1,1 |
| | Tails | -1,1 | 1,-1 |

# Base case: the TT for TMs

- Interestingly,
  - Matching pennies was proposed as an intelligence test (adversarial games) (Hibbard 2008, 2011).
- Again, the distribution of machines D is crucial.
  - Machines with very low complexity (repetitive) are easy to identify.
  - Machines with random outputs have very high complexity and are impossible to identify (a tie is the expected value).

Can we derive a more realistic distribution?

# Recursive TT for TMs

- The TT for TMs can start with a base distribution for the reference machines.
  - Whenever we start giving scores to some machines, we can start updating the distribution.
    - Machines which perform well will get higher probability.
    - Machines which perform badly will get lower probability.
  - By doing this process recursively:
    - We get a controlled version of the Darwin-Wallace distribution.
    - It is meaningful for some instances, e.g., matching pennies.

# Recursive TT for TMs

**Definition 2** *The recursive imitation game for Turing machines is defined as a tuple $\langle D, C, I \rangle$ where tests and distributions are obtained as follows:*

1. *Set $D_0 = D$ and $i = 0$.*
2. *For each agent $B$ in a sufficiently large set of TMs*
3. *Apply a sufficiently large set of instances of definition 1 with parameters $\langle D_i, B, C, I \rangle$.*
4. *$B$'s intelligence at degree $i$ is averaged from this sample of imitation tests.*
5. *End for*
6. *Set $i = i + 1$*
7. *Calculate a new distribution $D_i$ where each TM has a probability which is directly related to its intelligence at level $i - 1$.*
8. *Go to 2*

# Recursive TT for TMs

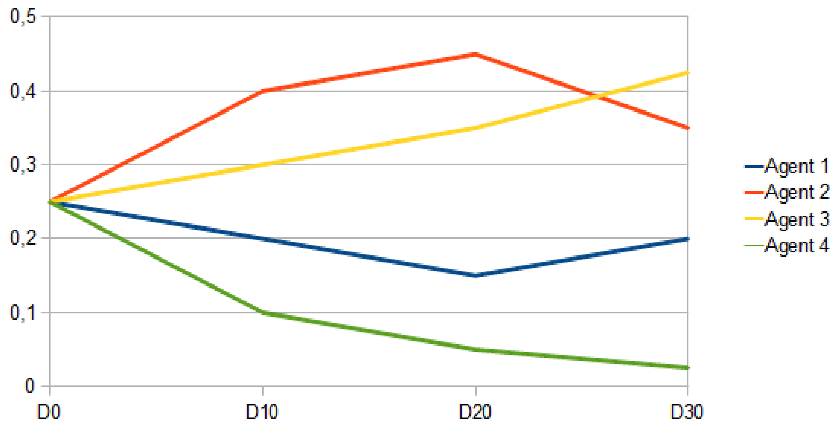
- The previous definition has many issues.
  - Divergent?
  - Intractable.
- But still useful conceptually.

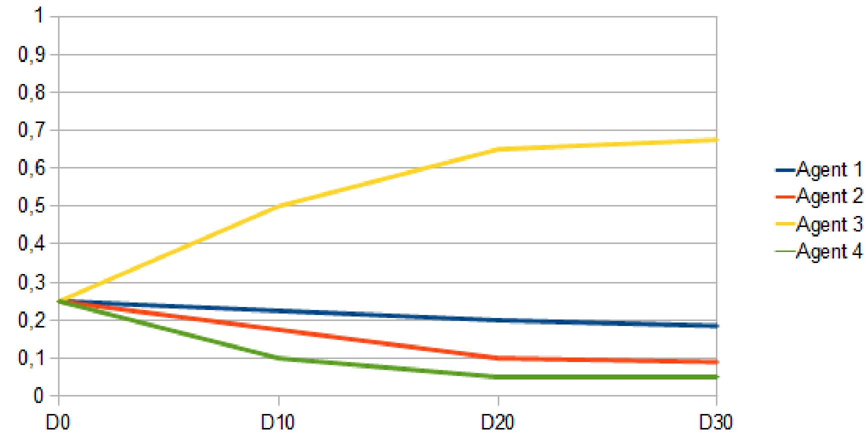
- In practice, it can be substituted by a (sampling) ranking system:
  - (e.g.) Elo's rating system in chess.

- Given an original distribution, we can update the distribution by randomly choosing pairs and updating the probability.

# Possible resulting distributions

- Depending on the agents and the game where they are evaluated, the resulting distribution can be different.



# Discussion

- The notion of Turing Test with Turing Machines is introduced as a way:
  - To get rid of the human reference in the tests.
  - To see very simple social intelligence tests, mainly adversarial.
- The idea of making it recursive tries to:
  - escape from the universal distribution.
  - derive a different notion of difficulty.

# Discussion

- The setting is still too simple to make a feasible test, but it is already helpful to:
  - Bridge the (until now) antagonistic views of intelligence testing using the Turing Test or using computational formal approaches using Kolmogorov Complexity, MML, etc.
  - Link intelligence testing with (evolutionary) game theory.

# Thank you!

Some pointers:

- Project: **anYnt** (Anytime Universal Intelligence)

  http://users.dsic.upv.es/proy/anynt/