Comparing humans and AI agents

Javier Insa-Cabrera[1] David L. Dowe[2] Sergio España-Cubillo[3]
M.Victoria Hernández-Lloreda[4] José Hernández-Orallo[1]

[1] DSIC, Universitat Politècnica de València, Spain. {jinsa, jorallo}@dsic.upv.es
[2] Clayton School of Information Technology, Monash University, Australia.
david.dowe@monash.edu
[3] ProS Research Center, Universitat Politècnica de València, Spain.
sergio.espana@pros.upv.es
[4] Departamento de Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Spain. vhlloreda@psi.ucm.es

Abstract. Comparing humans and machines is one important source of information about both machine and human strengths and limitations. Most of these comparisons and competitions are performed in rather specific tasks such as calculus, speech recognition, translation, games, etc. The information conveyed by these experiments is limited, since it portrays that machines are much better than humans at some domains and worse at others. In fact, CAPTCHAs exploit this fact. However, there have only been a few proposals of general intelligence tests in the last two decades, and, to our knowledge, just a couple of implementations and evaluations. In this paper, we implement one of the most recent test proposals, devise an interface for humans and use it to compare the intelligence of humans and Q-learning, a popular reinforcement learning algorithm. The results are highly informative in many ways, raising many questions on the use of a (universal) distribution of environments, on the role of measuring knowledge acquisition, and other issues, such as speed, duration of the test, scalability, etc.

Keywords: Intelligence measurement, universal intelligence, general vs. specific intelligence, reinforcement learning, IQ tests.

1 Introduction

It is well-known that IQ tests are not useful for evaluating the intelligence of machines. The main reason is not because machines are not able to 'understand' the test. The real reason is scarcely known and poorly understood, since available theories do not manage to fully explain the empirical observations: it has been shown that relative simple programs can be designed to score well on these tests [11]. Some other approaches such as the Turing Test [15] and Captchas [17] have their niches, but they are also inappropriate to evaluate AGI systems.

In the last fifteen years, several alternatives for a general (or universal) intelligence test (or definition) based on Solomonoff's universal distributions [12] (or related ideas such as MML, compression or Kolmogorov complexity) have been appearing on the scene [1, 3, 7, 8, 5], claiming that they are able to define or evaluate (machine) intelligence. In this paper we use one of these tests, a prototype based on the anytime intelligence test presented in [5] and the environment class introduced in [4], to evaluate one easily accessible biological system (*Homo sapiens*) and one off-the-shelf AI system, a popular reinforcement algorithm known as Q-learning [18]. In order to do the comparison we use the same environment class for both types of systems and we design hopefully non-biased interfaces for both. We perform a pilot experiment on a reduced group of individuals.

From this experiment we obtain a number of interesting findings and insights. First, it is possible to do the same test for humans and machines without being anthropomorphic. The test is exactly the same for both and it is founded on a theory derived from sound computational concepts. We just adapt the interface (what way rewards, actions and observations look like) depending on the type of subjects. Second, humans are not better than Q-learning in this test, even though the test (despite several simplifications) is based on a universal distribution of environments over a very general environment class. Third, since these results are consistent to those in [11] (which show that machines can score well in IQ tests), this gives additional evidence that a test which is valid for humans or for machines separately might be useless to distinguish or to place humans and machines on the same scale, so failing to be a universal intelligence test.

The following section overviews the most important proposals on defining and measuring machine intelligence to date, and, from them, it describes the intelligence test and the environment class we will use in this paper. Sections 3 and 4 describe the testing setting, the two types of agents we evaluate (Q-learning and humans) and their interfaces. Section 5 includes the comparison of the experimental results, analysing them by several factors. Finally, section 6 examines these results in a deeper way and draws several conclusions about the way universal intelligence tests should and should not be.

2 Measuring intelligence universally

Measuring machine intelligence or, more generally, performance has been virtually relegated to a philosophical or, at most, theoretical issue in AI. Given that state-of-the-art technology in AI is still far from truly intelligent machines, it seems that the Turing Test [15] (and its many variations [10]) and Captchas [17] are enough for philosophical debates and practical applications respectively. There are also tests and competitions in restricted domains, such as competitions in robotics, in game playing, in machine translation and in reinforcement learning (RL), most notably the RL competition. All of them use a somewhat arbitrary and frequently anthropomorphic set of tasks.

An alternative, general proposal for intelligence and performance evaluation is based on the notion of universal distribution [12] and the related algorithmic information theory (a.k.a. Kolmogorov complexity) [9]. Using this theory, we can define a universal distribution of tasks for a given AI realm, and sort them according to their (objective) complexity. There are some early works which develop these ideas to construct intelligence tests. First, [1] suggested the introduction of inductive inference problems in a somehow *induction-enhanced* or *compression-enhanced* Turing Test [15]. Second, [3] derived intelligence tests (C-tests) as sets of sequence prediction problems which were generated by a universal distribution, and the result (the intelligence of the agent) was a sum of performances for a range of problems of increasing complexity. The complexity of each sequence was derived from its Kolmogorov complexity (a Levin variant was used). This kind of problem (discrete sequence prediction), although typical in IQ tests, is a narrow AI realm. In fact, [11] showed that relatively simple algorithms could score well at IQ tests (and, as a consequence, at C-tests). In [3] the suggestion of using interactive tasks where "rewards and penalties could be used instead" was made. Later, Legg and Hutter (e.g. [7],[8]) gave a precise definition to the term "Universal Intelligence", also grounded in Kolmogorov complexity and Solomonoff's prediction theory, as a sum (or weighted average) of performances in all the possible RL-like environments. However, in order to make a feasible test by extending from (static) sequences to (dynamic) environments, several issues had to be solved first. In [5], they address the problem of finding a finite sample of environments and sessions, as well as appropriate approximations to Kolmogorov complexity, the inclusion of time, and the proper aggregation of rewards. The theory, however, has not been put into practice until now in the form of a real test, in order to evaluate artificial and biological agents, and, interestingly, to compare them. In this paper, we use a (simplified) implementation of this test (non-anytime) [5] using the environment class introduced in [4] to compare Q-learning with *Homo sapiens*.

From this comparison we want to answer several questions. Are these tests general enough? Does the complexity of the exercises correlate with the success rate of Q-learning and humans? Does the difference correspond to the real difference in intelligence between these two kinds of agents? What implications do the results have on the notion of universal intelligence and the tests that attempt to measure it? Answering all these questions is the goal of this paper.

The choice of a proper environment class is a crucial issue in any intelligence test. This is what [4] attempts, a hopefully unbiased environment class (called Λ) with spaces and agents with universal descriptive (Turing-complete) power. Basically, this environment considers a space as a graph with a different (and variable) topology of actions. Objects and agents can be introduced using Turing-complete languages to generate their movements. Rewards are rational numbers in the interval $[-1, 1]$ and are generated by two special agents *Good* and *Evil*, which leave rewards in the cells they visit. *Good* and *Evil* have the same pattern for behaviour except for the sign of the reward (+ for *Good*, − for *Evil*).

The environment class Λ is shown in [4] to have two relevant properties for a performance test: (1) their environments are always balanced (a random agent has expected reward 0), and (2) their environments are reward-sensitive (there is no sequence of actions such that the agent can be stuck in a heaven or hell situation, where rewards are positive or negative independently of what the agent may do). As argued in [5], these two properties are very important for the environments to be discriminative and comparable (and hence the results being properly aggregated into a single score, a performance or intelligence score). No other properties are imposed, such as (e.g.) environments being Markov processes or being ergodic. For more details of the environment class Λ, see [4].

3 Test setting and administration

Following the definition of the environment class Λ, we perform some simplifications to generate each environment. For instance, speed is not considered thus being a non-anytime version of the test presented in [5]. In addition, we do not use a Turing-complete algorithm to generate the environments. Spaces are generated by first determining the number of cells n_c, which is given by a number between 2 and 9, using an geometric / 'unary' distribution (i.e. $prob(n) = 2^{-n}$, and normalising to sum up to 1). Similarly, the number of actions n_a is defined with a uniform distribution between 2 and n_c. Both cells and actions are indexed with natural numbers. There is a special action 0 which connects every cell with itself (it is always possible to stay at the cell). A cell which is accessible from another cell using one action is called a 'neighbouring' or adjacent cell. The connections between cells are created by using a uniform distribution for each pair of cell and action, which assigns the destination cell for each pair. We consider the posibility that some actions may be disabled. Fig. 1 shows an example of a randomly generated space.

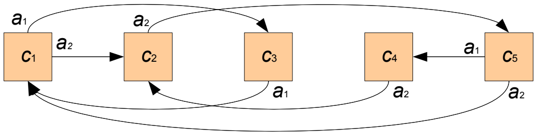

Fig. 1. A space with 5 cells and 3 actions (a_0, a_1, a_2). Reflexive action a_0 is not shown.

The number of cells and actions is, of course, related to the complexity of the space, but not monotonically related to its Kolmogorov complexity (or a computable variant such as Levin's Kt). Nonetheless, most of the actual grading of environments comes from the behaviour of *Good* and *Evil*. The sequence of actions for *Good* and *Evil* is defined by using a uniform distribution for each element in the sequence, and a unary (geometric) distribution to determine whether to stop the sequence, by using a probability of stopping (p_{stop}). An example of sequence for the space in Fig. 1 is 201210200, which means the execution of actions a_2, a_0, a_1, a_2, etc. Consider, e.g., that *Good* is placed at cell c_5. Since the pattern starts with '2', *Good* will move (via a_2) to cell c_1. The agents *Good* and *Evil* take one action from the sequence and execute it for each step. When the actions are exhausted, the sequence is started all over again. If an action is not allowed at a particular cell, the agent does not move.

Initially, each agent is randomly (using a uniform distribution) placed in a cell. Then, we let *Good*, *Evil* and the evaluated agent interact for a certain number of steps m. We call this an exercise (or episode). For an exercise we average the obtained rewards, so giving a score of the agent in the environment.

A test is a sequence of exercises or episodes. We will use 7 environments, each with a number of cells (n_c) from 3 to 9. The size of the patterns for *Good* and *Evil* will be made proportional (on average) to the number of cells, using $p_{stop} = 1/n_c$. In each environment, we will allow $10 \times (n_c - 1)$ steps so the agents have the chance to detect any pattern in the environment (exploration) and also have some further steps to exploit the findings (in case a pattern is actually conceived). The limitation of the number of environments and steps is justified because the tests is meant to be applied to biological agents in a reasonable period of time (e.g., 20 minutes) and we estimate an average of 4 seconds per action. Table 1 shows the choices we have made for the test:

Env. #	No. cells (n_c)	No. steps (m)	p_{stop}
1	3	20	1/3
2	4	30	1/4
3	5	40	1/5
4	6	50	1/6
5	7	60	1/7
6	8	70	1/8
7	9	80	1/9
TOTAL	-	350	-

Table 1. Setting for the 7 environments which compose the test.

Although [4] suggests a partially-observable interface, here we make it fully-observable, so agents see all the cells, the actions and their contents. The agents do not know in advance who *Good* is and who *Evil* is. They have to guess that.

4 Agents and interfaces

4.1 An AI agent: Q-learning

The choice of Q-learning is, of course, one of many possible choices for a reinforcement learning algorithm. The reason is deliberate because we want a standard algorithm to be evaluated first, and, most especially, because we do not want to evaluate (at the moment) very specialised algorithms for ergodic environments or algorithms with better computational properties (e.g. delayed Q-learning [13] would be a better option if speed were an issue). We use an off-the-shelf implementation of Q-learning, as explained in [18] and [14].

We use the description of cell contents as a state. We choose Q-learning's parameters as $\alpha = 0.05$ *learning rate* and $\gamma = 0.35$ *discount factor*. The parameters have been chosen by trying 20 consecutive values for α and γ between 0 and 1. These 400 combinations have been evaluated for 1,000 sessions each using random environments of different size and complexity and episodes of 10,000 steps. This choice is, of course, beneficial for Q-learning's performance in the tests.

Since we have rewards between -1 and 1, the elements in the Q matrix are set to 2.0 initially (rewards are normalised between 0 and 2 to always be positive).

4.2 A biological agent: *Homo sapiens*

We took 20 humans from a University Department (PhD students, research and teaching staff) with ages ranging between 20 and 50.

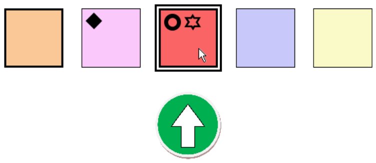

Fig. 2. A snapshot of the interface for humans. The agent has just received a positive reward, shown with the circle with an upwards arrow. The image also shows the agent located in cell 3, and *Evil* and *Good* are placed in cells 2 and 3 respectively. The agent can move to cell 1 and cell 3. Cell 3 is highlighted since the mouse pointer is over it.

The interface for humans has been designed with the following principles in mind: i) the signs used to represent observations should not have an implicit meaning for the subject, to avoid bias in favour of humans (e.g. no skull-and-bones for the Evil agent), ii) actions and rewards should be easily interpreted by the subject, to avoid a cognitive overhead that would bias the experiment in favour of Q-learning. This way, the following design decisions have been made (Fig. 2 shows a snapshot of the interface). At the beginning of the test, the subject is presented the task instructions, which strictly contain what the user should know. The cells are represented by coloured squares. Agents are represented by symbols that aim to be 'neutral' (e.g., ♦ stands for *Evil* and ✶ stands for *Good* in the third environment, and ◯ represents the subject in every environment). Accessible cells have a thicker border than non-accessible ones. When the subject rolls the mouse pointer over an accessible cell, this cell is highlighted using a double border and increasing the saturation of the background colour. Positive, neutral and negative rewards are represented by an upwards arrow in a green circle, a small square in a grey circle, and a downwards arrow in a red circle, respectively. The test and its interface for humans can be downloaded from `http://users.dsic.upv.es/proy/anynt/human1/test.html`.

5 Results

We performed 20 tests (with 7 exercises each) with the setting shown in Table 1 and we administered each of them to a human and to Q-learning.

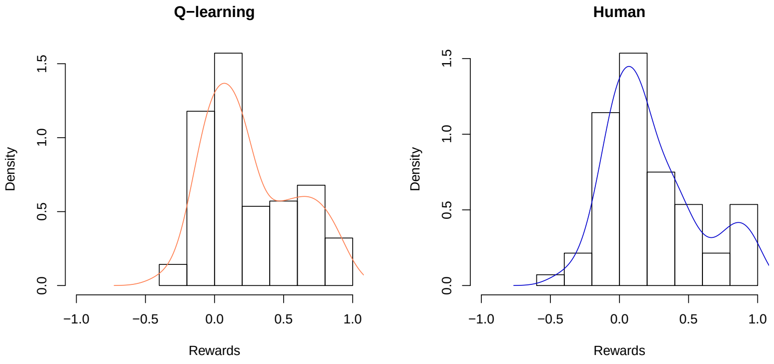

Fig. 3. Histograms of the (20 × 7 =) 140 exercises for Q-learning (left) and humans (right). Lines show the probability densities.

The first observation from this paired set of results comes from the means. While Q-learning has an overall mean of 0.259, humans show a mean of 0.237. The standard deviations are 0.122 and 0.150 respectively. Figure 3 shows the histograms and the probability densities (estimated by the R package).

To see the results in more detail in terms of the exercise, Figure 4 (left) shows the results aggregating by exercise (there is one exercise for each number of cells between 3 and 9, so totalling 7 exercises per test). This figure shows the mean, median and dispersion of both Q-learning and humans for each exercise. Looking at the boxplots for each space size we also see that there is no significant difference in terms of how Q-learning and humans perform in each of the seven exercises. While means are around 0.2 and 0.3, variances are smaller the larger the number of cells is. This is explained because the exercise with higher number of cells has a higher number of iterations (see Table 1).

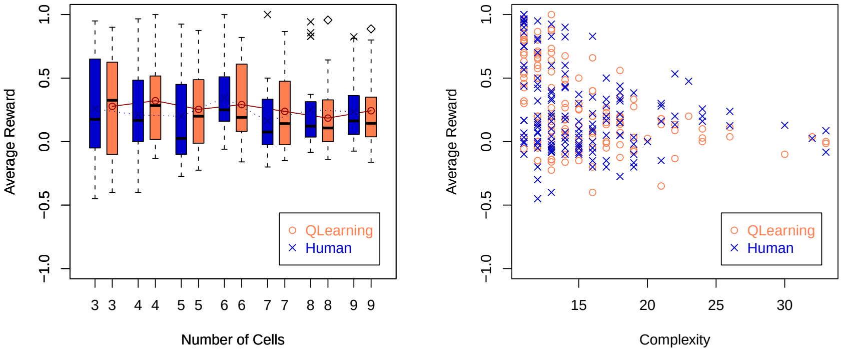

Fig. 4. Left: Box (whisker) plots for the seven exercises depending on the agent. Medians are shown in the box as a short black segment. Means are connected by a continuous line for Q-learning and a dashed line for humans. Right: the average reward results for the $20 \times 7 \times 2 = 280$ exercises using K^{approx} as a measure of complexity.

We applied two-way repeated measures ANOVA (agent × number of cells). ANOVA showed no statistically significant effects neither for agent ($F_{1,19} = .461$, $P = .506$), nor for the number of cells ($F_{6,114} = .401$, $P = .877$). No statistically significant *interaction* effect was found ($F_{6,114} = .693$, $P = .656$) either.

Finally, since the size of the space is not a measure of complexity, we explored the relation with the complexity of the environments. In order to approximate this complexity, we used the size of the compressed pattern for *Good* and *Evil*, denoted by P. More formally, given an environment μ, we calculate an approximation to its (Kolmogorov) complexity, denoted by K^{approx} as follows:

$$K^{approx} = LZ(P))$$

For instance, if a pattern is P=“2012222002222220022222002”, we compress the string (using the *memCompress* function in R, with a GNU project implementation of Lempel-Ziv coding). The length of the compressed string is 19.

Figure 4 (right) shows each of the $20 \times 7 = 140$ exercises for each kind of agent. Again we see a higher dispersion for humans than for Q-learning (the 20 humans are different, while Q-learning is exactly the same algorithm for each of the 20 tests). We calculate the Pearson correlation coefficient between complexity and reward. Now we do find a statistically significant correlation both for humans ($r = -.257$, $n = 140$, $P = .001$) and for Q-learning ($r = -.444$, $n = 140$, $P < .001$). We also analyse these correlations by number of cells, as shown in Table 2. This table shows Pearson correlation coefficients and associated significance levels (one tailed test) between "complexity" and "reward" by "numbers of cells" for each agent. All $n = 20$.

Agent	3 cells	4 cells	5 cells	6 cells	7 cells	8 cells	9 cells
Human	-.474 (.017)	-.134 (.286)	-.367 (.056)	-.515 (.010)	-.282 (.114)	-.189 (.213)	-.146 (.270)
Q-learning	-.612 (.002)	-.538 (.008)	-.526 (.009)	-.403 (.039)	-.442 (.026)	-.387 (.046)	-.465 (.019)

Table 2. Pearson correlation coefficients and p values (in parentheses) between "complexity" and "reward" by "numbers of cells".

We see that correlations are stronger and always significant for Q-learning, while they are milder (and not always significant) for humans. This may be explained because humans are not reset between exercises. In general, we would need more data (more tests) to confirm or refute this hypothesis.

6 Discussion

In section 2 we outlined several questions. One question is whether the test is general enough. It is true that we have made many simplifications to the environment class, in such a way that *Good* and *Evil* do not react to the environment (they just execute a cyclical sequence of actions as a pattern), and we have used a very simple approximation to complexity instead of better approximations to Kolmogorov complexity or Levin's Kt. In addition, and the parameters for Q-learning have been chosen to be optimal for these kinds of spaces and patterns. Besides, humans are not (cannot be) reset between exercises. Despite all these issues (most of) which are in favour of Q-learning, we think (although this cannot be concluded in an absolute way) that the tests are not general enough. Q-learning is not the best AI algorithm available nowadays (in fact we do not consider Q-learning very intelligent). So, the results are not representing the real difference in intelligence between humans and Q-learning.

A possibility is that our sample size is perhaps too small. Having more environments of higher complexity and letting the agents interact longer with each of them may perhaps portray a different picture. Nonetheless, it is not clear that humans can scale up well in this kind of exercise, especially if no part of previous exercises can be reused to other exercises. First, some of the patterns which appeared in the most complex exercises were considered very difficult by humans. Second, Q-learning requires many interactions to converge, so perhaps this would only exaggerate the difference in favour of Q-learning. In any case, this should be properly analysed with further experiments.

A more fundamental issue is whether we are testing on the wrong sort of environments. The environment class is a general class which includes two symmetrical agents, *Good* and *Evil*, which are in charge of rewards. We do not think that this environment class is, in any case, biased against humans (the contrary can be argued, though). In the end, the question of whether a test is biased is difficult to answer, since any single choice implies a certain bias. So, in our opinion, the problem might be found in the environment distribution. Choosing the universal distribution gives high probability to very simple examples with very simple patterns, but more importantly, makes any kind of rich interaction impossible even in environments of high Kolmogorov complexity. So, a better environment distribution (and perhaps class) should give more probability to incremental knowledge acquisition, social capabilities and more reactivity.

This goal towards more knowledge-intensive tasks has the risk of focussing on knowledge and language, or to embark on Ttests without any theoretical background, such as Jeopardy-like contests. The generality of these tasks may be high, although the adaptability and the required learning abilities might be low. This is something recurrent in psychometrics, where it is important (but difficult) to distinguish between knowledge acquisition capabilities and knowledge application. And it is also a challenge for RL-like evaluations and systems, where knowledge acquisition usually starts from scratch and is not incremental.

So, one of the things that we have learnt is that the change of universal distributions from passive environments (as originally proposed in [1] and [3]) to interactive environments (as also suggested in [3] and fully developed in [7, 8]) is in the right direction, but it is not the solution yet. It is clear that it allows for a more natural interpretation of the notion of intelligence as performance in a wide range of environments, and it eases the application of tests outside humans and machines (children, apes, etc.), but there are some other issues we have to address to give an appropriate definition of intelligence and a practical test. The proposal for an adaptive test [5] introduces many new ideas about creating practical intelligence tests, and the universal distribution is substituted by an adaptive distribution, so allowing a faster convergence to complexity levels which are more appropriate for the agent. Nonetheless, we think that the priority is in defining new environment distributions which can give higher probability to environments where intelligence can show its full potential (see, e.g. [6]).

Summing up, while there has been some work on comparing humans and machines on some specific tasks, e.g., humans and Q-learning in [2], this paper may start a series of experimental research comparing several artificial agents (such as other algorithms in reinforcement learning, MonteCarlo AIXI [16], etc.) and other biological agents (children, other apes, etc) for *general* tasks. This might be a highly valuable source of information about whether the concept of universal intelligence evaluation works, by trying to construct more and more general (and universal) intelligence tests. This could lead eventually to a new discipline, for which we already suggest a name: “universal psychometrics”.

Acknowledgments

We thank the anonymous reviewers for their helpful comments. We also thank José Antonio Martín H. for helping us with several issues about the RL competition, RL-Glue and reinforcement learning in general. We are also grateful to all the subjects who took the test. We also thank the funding from the Spanish MEC and MICINN for projects TIN2009-06078-E/TIN, Consolider-Ingenio CSD2007-00022 and TIN2010-21062-C02, for MEC FPU grant AP2006-02323, and Generalitat Valenciana for Prometeo/2008/051.

References

1. D. L. Dowe and A. R. Hajek. A non-behavioural, computational extension to the Turing Test. In *Intl. Conf. on Computational Intelligence & multimedia applications (ICCIMA'98), Gippsland, Australia*, pages 101–106, 1998.
2. D. Gordon and D. Subramanian. A cognitive model of learning to navigate. In *Proc. 19th Conf. of the Cognitive Science Society, 1997*, volume 25, page 271. Lawrence Erlbaum, 1997.
3. J. Hernández-Orallo. Beyond the Turing Test. *J. Logic, Language & Information*, 9(4):447–466, 2000.
4. J. Hernández-Orallo. A (hopefully) non-biased universal environment class for measuring intelligence of biological and artificial systems. In M. Hutter et al., editor, *Artificial General Intelligence, 3rd Intl Conf*, pages 182–183. Atlantis Press, Extended report at http://users.dsic.upv.es/proy/anynt/unbiased.pdf, 2010.
5. J. Hernández-Orallo and D. L. Dowe. Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence*, 174(18):1508 – 1539, 2010.
6. J. Hernández-Orallo, D.L. Dowe, S. España-Cubillo, M.V. Hernández-Lloreda, and J. Insa-Cabrera. On more realistic environment distributions for defining, evaluating and developing intelligence. In J. Schmidhuber and K.R. Thórisson (eds), editors, *Artificial General Intelligence 2011.* LNAI series, Springer, 2011.
7. S. Legg and M. Hutter. A universal measure of intelligence for artificial agents. In *Intl Joint Conf on Artificial Intelligence, IJCAI*, volume 19, page 1509, 2005.
8. S. Legg and M. Hutter. Universal intelligence: A definition of machine intelligence. *Minds and Machines*, 17(4):391–444, 2007.
9. M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and its applications (3rd ed.).* Springer-Verlag New York, Inc., 2008.
10. G. Oppy and D. L. Dowe. The Turing Test. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy.* Stanford University, 2011. http://plato.stanford.edu/entries/turing-test/.
11. P. Sanghi and D. L. Dowe. A computer program capable of passing IQ tests. In *4th Intl. Conf. on Cognitive Science (ICCS'03), Sydney*, pages 570–575, 2003.
12. R. J. Solomonoff. A formal theory of inductive inference. Part I. *Information and control*, 7(1):1–22, 1964.
13. A.L. Strehl, L. Li, E. Wiewiora, J. Langford, and M.L. Littman. PAC model-free reinforcement learning. In *Proc. of the 23rd Intl Conf on Machine learning*, ICML '06, pages 881–888, New York, 2006.
14. R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction.* The MIT press, 1998.
15. A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.
16. J. Veness, K.S. Ng, M. Hutter, and D. Silver. A Monte Carlo AIXI Approximation. *Journal of Artificial Intelligence Research, JAIR*, 40:95–142, 2011.
17. L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. *Communications of the ACM*, 47(2):56–60, 2004.
18. C.J.C.H. Watkins and P. Dayan. Q-learning. *Mach. learning*, 8(3):279–292, 1992.