Note: This paper was presented to the AAAI 2015 Workshop on Beyond Turing Test.

Geometry, Drawings, Visual Thinking and Imagery: Towards a Visual Turing Test of Machine Intelligence

Ashok K. Goel
Design & Intelligence Laboratory, School of Interactive Computing, and
Center for Biologically Inspired Design, Georgia Institute of Technology
goel@cc.gatech.edu

Abstract: Psychometrics is the science of measurement of human intelligence, knowledge, aptitude and personality. By analogy, we need a science of computational psychometrics for measuring the intelligence and knowledge of intelligent machines. In psychometrics, the most common and reliable test of human intelligence is the Raven's Progressive Matrices Test that relies solely on a set of visual analogy problems. By analogy, we propose understanding visual drawings as a test of machine intelligence. Understanding drawings requires background knowledge, commonsense reasoning, as well as mental imagery, all of which appear central to human-level intelligence.

Background, Motivations and Goals: Any behavioral test of machine intelligence must have an input and a desired output. The Turing Test, for example, uses natural language as both the input and the output (Turing 1950). Many variants of the Turing Test, such as the so-called Winograd Schema test (Levesque, Davis & Morgenstern 2011), also use restricted forms of natural language as the input and output. This is not unreasonable: natural language understanding and generation typically require background knowledge, commonsense reasoning, and semantic analysis, not just syntactic parsing. However, using only natural language as the input and the output is neither necessary nor sufficient for measurement of machine intelligence.

Before we consider measurement of machine intelligence in detail, let us first briefly consider the measurement of human intelligence. Psychometrics is the science of measurement of human intelligence, knowledge, aptitude and personality. Psychometrics has developed a wide variety of intelligence tests such as the Wechsler's test. This is because human intelligence encompasses several kinds of intelligence, and because human intelligence develop and ages over time. Thus, no one test is sufficient for measuring the various kinds of human intelligence at various stages of development and aging.

It follows that as machines begin to reach human-level intelligence, there will be need to develop not one but several tests for measuring machine intelligence. Thus, by analogy to the science of psychometrics for measuring human intelligence, we need a new science for measuring machine intelligence. We call this science *computational psychometrics.*

Within psychometrics, the Raven's Progressive Matrices test of intelligence is a widely used test of general intelligence (Raven, Raven & Court 2003). Interestingly, the Raven's intelligence test consists only of a set of visual analogy problems. The Raven's test engages a variety of core cognitive abilities, including problem solving, background knowledge, commonsense reasoning, pattern abstraction, analogical transfer, as well as mental imagery (Hunt 1974). Although the Raven's test contains only visual analogy problems, it is considered to be a robust and reliable measurement of general intelligence, and its results correlate well with other tests of human intelligence.

By analogy to the Raven's test of human intelligence, we propose two kinds of tests for measuring machine intelligence: (1) Psychometrics tests that rely on visual analogy problems, and (2) Tests based on understanding design drawings. As Bringsjord & Schimanksi (2003) note, an advantage of the former is that it enables direct comparison of the performance of machines with normative data about human performance on the tests; a benefit of the latter is that they are open-ended.

Intelligence Tests Based on Visual Analogy Problems: Performance on psychometric tests of human intelligence is generally measured in terms of number of questions answered correctly, which can then be used as an index into normative test data to determine the score or ranking. However, patterns of errors on the test problems could be another important measure of machine intelligence on the psychometric tests.

Figure 1 illustrates a problem similar to problems on the Raven's test of intelligence. Kunda, McGreggor & Goel (2013) provide a computational model of the problem solving on the Raven's test. Kunda et al. (2013) provide an initial comparison of the errors made by the computational model with errors made by humans.

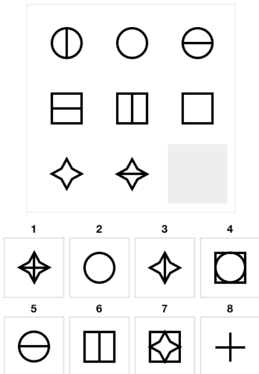

Figure 1: A problem similar to the problems on the Raven's test of general intelligence. (The correct answer is #3.)

Figure 2 illustrates an example from the so-called Dehaene test of core geometry (Dehaene et al. 2006). The task here is to select the odd one out, i.e., the drawing that does not fit the general pattern shown in the six images. McGreggor & Goel (2013) provides a computational model of problem solving on the Dehaene's test and compares it with human performance.

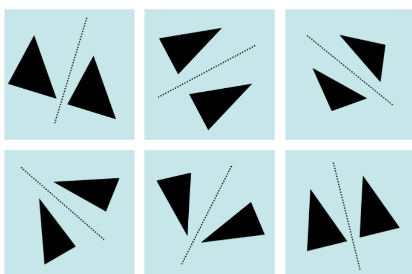

Figure 2: An example from Dehaene test of core geometry. (The correct answer is #1 because it is not symmetric around the depicted axis.)

Turing Test Based on Understanding Visual Designs: A second test of machine intelligence that builds on the above psychometric tests, and is closer in spirit to the Turing test, pertains to understanding of visual designs. Let us consider, for example, the visual design illustrated in Figure 4. Most humans can understand the basics of the design: they can recognize the main components, abstract the main function of the design, and reason about the causal behaviors of the system. Although the problem requires some background knowledge, one does not need correct or complete knowledge of the statics and dynamics systems to give commonsense answers to the basic questions about the pulley system depicted in the figure. Yaner & Goel (2007) provide a computational model of understanding design drawings.

An advantage of an intelligence test based on an understanding of visual designs such as the one depicted in Figure 3 is that is both open-ended and constrained in the right ways. It is open-ended because we can easily generate

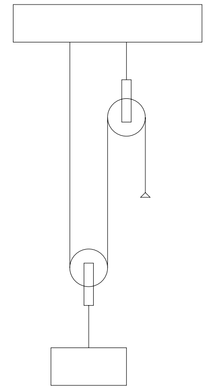

Figure 3: A Drawing of Simple Pulley Design.

a large number of drawings of simple designs of systems humans encounter in everyday life. It is constrained because a machine cannot give arbitrary answers to questions about the design that can deceive a judge. This is because the questions and answers in this version of the Turing Test, though expressed in natural language, are grounded in the visual drawing. This addresses some of the limitations of the original Turing Test.

References

Bringsjord, S., & Schimanski, B. (2003) What is artificial intelligence? Psychometric AI as an answer. In *Procs IJCAI-2003*, pp. 887-893.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, 311(5759), 381-384.

Hunt, E. (1974) Quote the Raven? Nevermore. In L.Gregg (Ed.) *Knowledge and Cognition,* Erlbaum, Hillsdale, NJ, pp. 129-158.

Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the RPM intelligence test using iconic visual representations. *Cognitive Systems Research*, *22-23*, 47-66.

Kunda, M., Souleires, I. Rozgа, A., & and Goel, A. (2013) Methods for Classifying Errors on the Raven's Standard Progressive Matrices Test. In *Procs. 35th Annual Meeting of the Cognitive Science Society*, Berlin, July 31-August 3.

Levesque, H., Davis, E., & Morgenstern, L. (2011) The Winograd Schema Challenge. In Procs. AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning.

McGreggor, K., & Goel, A. (2013) Fractal Representations and Core Geometry. In *Procs. Second Conference on Advances in Cognitive Systems*, Baltimore.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Turing, A. (1950) Computing machinery and intelligence, *Mind* 59, 433–460.

Wechsler, D. (2008) *Wechsler Adult Intelligence Scale.* San Antonio, TX: Pearson.

Yaner, P., & Goel, A. (2007) Understanding Diagrams by Compositional Analogy. In Procs. IJCAI-07, pp. 1131-1137.