ble for download from our lab website. The Raven test images as scanned, however, are copyrighted and thus are not available for download.

Abstractions, Metrics, and Calculations

The images associated with each problem, in general, had a maximum pixel dimension of between 150 and 250 pixels. We chose a partitioning scheme which started at the maximum dimension, then descended in steps of 10, until it reached a minimum size of no smaller than 4 pixels, yielding 14 to 22 levels of abstraction for each problem.

At each level of abstraction, we calculated the similarity value for each possible answer, as proscribed by the Confident Ravens algorithm. For those calculations, we used the Tversky contrast ratio formula (1977), and set α to 1.0 and β equal to 0.0, conforming to values used in the coincidence model by Bush and Mosteller (1953), yielding an asymmetric similarity metric preferential to the problem matrix's relationships. From those values, we calculated the mean and standard deviation, and then calculated the deviation and confidence for each answer. We made note of which answers provided a confidence above our chosen level, and whether for each abstraction level the answer was unambiguous or ambiguous, and if ambiguous, in what manner.

As we were exploring the advent and disappearance of ambiguity and the effect of confidence, we chose to allow the algorithm to run fully at all available levels of abstraction, rather than halting when an unambiguous answer was determined.

Performance on the SPM test: 54 of 60

On the Raven's Standard Progressive Matrices (SPM) test, the Confident Ravens algorithm detected the correct answer at a 95% or higher level of confidence on 54 of the 60 problems. The number of problems with detected correct answers per set were 12 for set A, 10 for set B, 12 for set C, 8 for set D, and 12 for set E. Of the 54 problems where the correct answers detected, 22 problems were answered ambiguously.

Performance on the APM test: 43 of 48

On the Raven's Advanced Progressive Matrices (APM) test, the Confident Ravens algorithm detected the correct answer at a 95% or higher level of confidence on 43 of the 48 problems. The number of problems with detected correct answers per set were 11 for set A, and 32 for set B. Of the 43 problems where the correct answers detected, 27 problems were answered ambiguously.

Performance on the CPM test: 35 of 36

On the Raven's Coloured Progressive Matrices (CPM) test, the Confident Ravens algorithm detected the correct answer at a 95% or higher level of confidence on 35 of the 36 problems. The number of problems with detected correct answers per set were 12 for set A, 12 for set AB, and 11 for set B. Of the 35 problems where the correct answers detected, 5 problems were answered ambiguously.

Performance on the SPM Plus test: 58 of 60

On the Raven's SPM Plus test, the Confident Ravens algorithm detected the correct answer at a 95% or higher level of confidence on 58 of the 60 problems. The number of problems with detected correct answers per set were 12 for set A, 11 for set B, 12 for set C, 12 for set D, and 11 for set E. Of the 58 problems where the correct answers detected, 23 problems were answered ambiguously.

Confidence and Ambiguity, Revisited

We explored a range of confidence values for each test suite of problems, and illustrate these findings in Table 2.

Note that as confidence increases from 95% to 99.99%, the test scores decrease, but so too does the ambiguity. Analogously, as the confidence is relaxed from 95% down to 60%, test scores increase, but so too does ambiguity. By inspection, we note that there is a marked shift in the rate at which test scores and ambiguity change between 99.9% and 95%, suggesting that 95% confidence may be a reasonable choice.

confidence threshold	SPM 60		APM 48		CPM 36		SPMPlus 60	
	correct	ambiguous	correct	ambiguous	correct	ambiguous	correct	ambiguous
99.99%	41	1	28	1	24	0	44	2
99.9%	49	4	38	8	30	0	53	5
99%	53	14	42	16	33	1	58	14
95%	54	22	43	27	35	5	58	23
90%	55	29	45	31	36	9	59	32
80%	57	36	45	38	36	9	59	37
60%	58	42	47	45	36	14	60	45

Table 2. The Effect of Confidence on Score and Ambiguity

Our findings indicate that at 95% confidence, those problems which are answered correctly but ambiguously are vacillating almost in every case between two choices (out of an original 6 or 8 possible answers for the problem). This narrowing of choices suggests to us that ambiguity resolution might entail a closer examination of just those specific selections, via re-representation as afforded by the fractal representation, a change of representational framework, or a change of algorithm altogether.

Comparison to other computational models

As we noted in the introduction, there are other computational models which have been used on some or all problems of certain tests. However, all other computational accounts report scores when choosing a single answer per problem, and do not report at all the confidence with which their algorithms chose those answers. As such, our reported totals must be considered as a potential high score for Confident Ravens if the issues of ambiguity were to be sufficiently addressed.

Also as we noted earlier, this paper presents the first computational account of a model running against all four variants of the RPM. Other accounts generally report scores on the SPM or the APM, and no other account exists for scores on the SPM Plus.

Carpenter et al. (1990) report results of running two versions of their algorithm (FairRaven and BetterRaven) against a subset of the APM problems (34 of the 48 total). The subset of problems chosen by Carpenter et al. reflect those whose rules and representations were deemed as inferable by their production rule based system. They report that FairRaven achieves a score of 23 out of the 34, while BetterRaven achieves a score of 32 out of the 34.

Lovett et al (2007, 2010) report results from their computational model's approach to the Raven's SPM test. In each account, only a portion of the test was attempted, but Lovett et al project an overall score based on the performance of the attempted sections. The latest published account by Lovett et al (2010) reports a score of 44 out of 48 attempted problems from sets B through E of the SPM test, but does not offer a breakdown of this score by problem set. Lovett et al. (2010) project a score of 56 for the entire test, based on human normative data indicating a probable score of 12 on set A given their model's performance on the attempted sets.

Cirillo and Ström (2010) report that their system was tested against Sets C through E of the SPM and solved 8, 10, and 10 problems, respectively, for a score of 28 out of the 36 problems attempted. Though unattempted, they predict that their system would score 19 on the APM (a prediction of 7 on set A, and 12 on set B).

Kunda et al. (2013) reports the results of running their ASTI algorithms against all of the problems on both the SPM and the APM tests, with a detailed breakdown of scoring per test. They report a score of 35 for the SPM test, and a score of 21 on the APM test. In her dissertation, Kunda (2013) reports a score of 50 for the SPM, 18 for the APM, and 35 on the CPM.

McGreggor et al. (2011) contains an account of running a preliminary version of their algorithm using fractal representations against all problems on the SPM. They report a score of 32 on the SPM, 11 on set A, 7 on set B, 5 on set C, 7 on set D, and 2 on set E. They report that these results were consistent with human test taker norms. Kunda et al. (2012) offers a summation of the fractal algorithm as applied to the APM, with a score of 38, 12 on set A, and 26 on set B.

The work we present here represents a substantial theoretical extension as well as a significant performance improvement upon these earlier fractal results.

Conclusion

In this paper, we have presented a comprehensive account of our efforts to address the entire Raven's Progressive Matrices tests using purely visual representations, the first such account in the literature. We developed the Confident Ravens algorithm, a computational model which uses features derived from fractal representations to calculate Tversky similarities between relationships in the test problem matrices and candidate answers, and which uses levels of abstraction, through re-representing the visual representation at differing resolutions, to determine overall confidence in the selection of an answer. Finally, we presented a comparison of the results of running the Confident Ravens algorithm to all available published accounts, and showed that the Confident Ravens algorithm's performance at detecting the correct answer is on par with those accounts.

The claim that we present throughout these results, however, is that a computational model may provide both an answer as well as a characterization of the confidence with which the answer is given. Moreover, we have shown that insufficient confidence in a selected answer may be used by that computational model to force a reconsideration of a problem, through re-representation, representational shift, or algorithm change. Thus, we suggest that confidence is hereby well-established as a motivating factor for reasoning, and as a potential drive for an intelligent agent.

Acknowledgments

This work has benefited from many discussions with our colleague Maithilee Kunda and the members of the Design and Intelligence Lab at Georgia Institute of Technology. We thank the US National Science Foundation for its support of this work through IIS Grant #1116541, entitled "Addressing visual analogy problems on the Raven's intelligence test."

References

Barnsley, M., and Hurd, L. 1992. *Fractal Image Compression*. Boston, MA: A.K. Peters.

Bringsjord, S., and Schimanski, B. 2003. What is artificial intelligence? Psychometric AI as an answer. *International Joint Conference on Artificial Intelligence,* 18: 887–893.

Bush, R.R., and Mosteller, F. 1953. A Stochastic Model with Applications to Learning. *The Annals of Mathematical Statistics*, 24(4): 559-585.

Carpenter, P., Just, M., and Shell, P. 1990. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review,* 97(3): 404-431.

Cirillo, S., and Ström, V. 2010. An anthropomorphic solver for Raven’s Progressive Matrices (No. 2010:096). Goteborg, Sweden: Chalmers University of Technology.

Haugeland, J. ed. 1981. *Mind Design: Philosophy, Psychology and Artificial Intelligence*. MIT Press.

Hofstadter, D., and Fluid Analogies Research Group. eds. 1995. *Fluid concepts & creative analogies: Computer models of the fundamental mechanisms of thought.* New York: Basic Books.

Hunt, E. 1974. Quote the raven? Nevermore! In L. W. Gregg ed., *Knowledge and Cognition* (pp. 129–158). Hillsdale, NJ: Erlbaum.

Kunda, M. 2013. Visual Problem Solving in Autism, Psychometrics, and AI: The Case of the Raven's Progressive Matrices Intelligence Test. Doctoral dissertation, Georgia Institute of Technology.

Kunda, M., McGreggor, K. and Goel, A. 2011. Two Visual Strategies for Solving the Raven’s Progressive Matrices Intelligence Test. *Proceedings of the 25th AAAI Conference on Artificial Intelligence*.

Kunda, M., McGreggor, K. and Goel, A. 2012. Reasoning on the Raven’s Advanced Progressive Matrices Test with Iconic Visual Representations. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan.

Kunda, M., McGreggor, K., & Goel, A. K. 2013. A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22-23, pp. 47-66.

Levesque, H. J., Davis, E., & Morgenstern, L. 2011. The Winograd Schema Challenge. *In AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning.*

Lovett, A. Forbus, K., and Usher, J. 2007. Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. *Proceedings of the 29th Annual Conference of the Cognitive Science Society.*

Lovett, A., Forbus, K., and Usher, J. 2010. A structure-mapping model of Raven’s Progressive Matrices. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

Mandelbrot, B. 1982. *The fractal geometry of nature.* San Francisco: W.H. Freeman.

McGreggor, K. (2013). Fractal Reasoning. Doctoral dissertation, Georgia Institute of Technology.

McGreggor, K., Kunda, M., & Goel, A. K. (2011). Fractal Analogies: Preliminary Results from the Raven's Test of Intelligence. In Proceedings of the Second International Conference on Computational Creativity (ICCC), Mexico City. pp. 69-71.

McGreggor, K., & Goel, A. 2011. Finding the odd one out: a fractal analogical approach. *Proceedings of the 8th ACM conference on Creativity and cognition* (pp. 289-298). ACM.

McGreggor, K., and Goel, A. 2012. Fractal analogies for general intelligence. *Artificial General Intelligence.* Springer Berlin Heidelberg, 177-188.

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

Rasmussen, D., and Eliasmith, C. 2011. A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science,* 3(1), 140-153.

Tversky, A. 1977. Features of similarity. *Psychological Review,* 84(4), 327-352.