

Running head: PSYCHOMETRIC ARTIFICIAL INTELLIGENCE

Psychometric artificial intelligence (PAI), its position in relation to mainstream AI and
implications for psychometrics

Tuuli Pöllänen

The intention of my seminar was to dive into the concept of psychometric artificial intelligence. As I thought out the theme, I was excited by the surprisingly large amount of literature I found, and by the idea of a concept utilizing the mathematical aspects of psychology and psychological assessment with automated computer models. My head was buzzing with all the potentials of the topic for all kinds of applications, from automated test item creation to gaining a better understanding of some of the most basic aspects of psychometric intelligence by implementing it with an automated model not burdened by individual or situational variation that often saturates the findings when researching it through a sample of test participants. But as I read through the articles, I gradually became more and more disenchanted by how little such an interesting concept had really been explored. Most of the publications were from a very small number of authors, and many of them displaying a surprisingly poor understanding of some basic psychometric concepts. But in a way, this gave me the opportunity to observe a field of computer science borrowing some of the most well-structured, immanent concepts of psychology, and gave me insight as to just how dire is the need for interdisciplinary collaboration.

So I changed the point of view I had on my theme, and decided to first discuss the very roots of psychometric AI, and its side role in the “triumph” of computational psychology (the rapid popularity it gained in 1970’s). I will then move on to discuss the characteristics of psychometric AI and one of its existing applications – or, to be completely honest, the only existing application I found I was really impressed by – and finally, I will discuss some of the dynamics between the main stream field of artificial intelligence and in PAI – one of its least popular constituents – as well as why I believe the position of PAI should be enhanced. Finally, I will reflect on some implications AI would have for psychometrics instead of only focusing on the opposite.

PAI and the triumph of computational psychology

Allen Newell's (1973) paper, "You can't play 20 questions with nature and win", is often cited as the turning point for computational psychology, at least when it comes to the vantage point of using cognitive architectures to model human cognition. What is often overlooked is that where cognitive architectures have been a popular subject in artificial intelligence (AI) ever since Newell's paper, it was not the only paradigm that first gained publicity in the 1973 publication. Even though psychometric artificial intelligence (PAI) has not been popular in the AI field until the turn of the millennium (without overlooking the contributions from Evans, 1968, who actually dealt with topics related to PAI already at the end of the 1960's), its roots are also in Newell's paper. It is curious that the text triggered such a paradigmatic change in computational models with cognitive architectures but did so little for psychometric AI – perhaps the scholars of that time considered the notion of building embodied artificial intelligence capable of solving psychometric intelligence tests utopian, or perhaps this is due to the relatively narrow scope of practical applications for PAI, which I will discuss later.

In any case, whether the discussion is of cognitive architectures or PAI, I find Newell's (1973) paper to be a good starting point for the discussion, as it represents the point where the story begins for both paradigms, and describes the problems computational psychology was first meant to counter.

You can't play 20 questions with nature and win

In 1973, Allen Newell was given a task to comment on every paper appearing in W.G. Chase's (Ed.) symposium on visual information processing. Rather ambitiously, Newell decided to not only comment on every paper, but also to attempt to draw a conclusion that would give a holistic picture of visual processing – a decision he claims he soon regretted (Newell, 1973). After realizing the impossibility of the task, Newell titled his paper (which is

now considered a well-read classic and which is quite an entertaining read – highly recommended) »You can't play twenty questions with nature and win«, and dove into the problems of cognitive psychology of his time (and largely what is still one of the key issues in cognitive psychology as well as AI in the more recent years).

Newell (1973) expressed his despair over how nearly all experimentalists of his time were dealing with their own little areas, performing what he thought were absolutely magnificent experiments and concluding interesting results, but driven by mutually exclusive binary distinctions (*nature or nurture, parallel or serial, central or peripheral*) and with very little interest in drawing together with other studies to create a thorough, holistic picture of human cognition, where processes could happen simultaneously on multiple systems level on several different ways, influenced and modified by several different factors. He pointed out that there were a few strands of theory that thrived for some quantitative explanation over a class of phenomena, but they were not common enough to »quiet his concern«. He argued that if something was not to change soon to provide the field a »path to clarity«, the situation was to get muddier and muddier. His key issue was stated as follows:

»Science advances in playing twenty questions with nature. The *proper* tactic is to frame a general question, hopefully binary, that can be attacked experimentally.

Having settled that bits-worth, one can proceed to the next. The policy appears optimal – one never risks much, there is feedback from nature at every step, and progress is inevitable. Unfortunately, the questions never seem to be really answered, the strategy does not seem to work.« (Newell, 1973, p. 290).

In his paper, Newell diagnoses the problems he found in an attempt to make a synthesis on the paper, and writes down two injunctions on how to approach cognitive experimentation:

1. Know the method your subject is using to perform the experimental task. Behavior is programmable. This means that it is under the control of the subject as means to his own ends. To predict behavior, you must know his goals, the structure of the task environment, and the invariant structure of his processing mechanisms. From these, you can predict what methods are available to the subject.

This injunction sums the general intention of computational psychology. If behavior is programmable, under the control of the subject as means to ends, then behavior is predictable, given realistic constraints and if the parameters of the situations are known.

2. Never average over methods. Uncertainty of which method the subject uses takes substantial space in discussion of experimental results, but averaging methods leads to either “garbage, or, even worse: spurious regularity” (Newell, 1973, p. 295).

Newell (1973) argued that the task in psychology is first to discover the structure of the subject, which is fixed and invariant so that we can theoretically infer the method it would choose for solving a task. Then, once we know the subject’s goal and the task environment, we generate a small collection of methods that are likely to be used by him in the task, and with careful experimental design or by suitable post hoc analyses, we can find out what method he selected. He explained that the reason for numerous correctional experiments in cognitive psychology was likely the fact that subjects opted for new methods to approach the problem – by predicting likely methods beforehand it would be possible to exit the loop of repeating correctional experiments. Newell (1973) argued that where experimentalists in cognitive psychology often did attempt to pick a couple of methods their subject could have used to approach the task, or portrayed the entire set of processing stages in a flow diagram (a novelty at his time), they often left open the total method being used, and did not operate within a frame, a *control structure*, which would *constrain* what other methods might also be

evoked to perform the same task. This brings us to computational cognitive models, as Newell argues that the control structures are best illustrated by programming languages. A rather simplified account of Newell's explanations is that programmed algebraic structures work well for making natural constraints (or control structures) in that they create a space where different methods can be tested *if* they conform to the limitations of space and time.

In order to remedy the problems he pointed out, Newell (1973) suggested three approaches, or paradigms, to experimentation:

1. Complete processing models

Newell (1973) suggested that scientists should attempt to construct complete processing models, rather than partial ones as was the tradition. He also argued that the control processes employed here should be psychologically relevant. This is the suggested paradigm that opened the door for cognitive architectures (specifically complete processing models), with the idea of creating systems that would work as blueprints for creating and evaluating intelligent entities with human-like problem solving strategies.

2. Analyzing complex tasks

This paradigm had the basic idea of focusing on a single complex task instead of designing specific small experiments to try and settle specific small questions. By making sure that experiments employing those complex tasks could be sown together and each contributed their own detail to a bigger picture, Newell wanted to fix the problem he saw in psychology simply being too fragmented as a science. This paradigm was popular for quite a while, and is the one most people in computational cognitive science and AI are familiar with (Bringsjord, 2011), however with time – and in my opinion, perhaps with increasing processing capacities in computers – the applicative value of this concept became questionable. The Deep Blue chess machine is often cited as an example as to why this

paradigm so swiftly lost its popularity (e.g. Bringsjord & Licato, 2012; Bringsjord, 2011).

Chess playing is often considered a good example of a complex task that could be analyzed and modeled with this approach, but the Deep Blue displayed that winning is possible simply with employing standard search techniques in massive data searches through the entire game space, with very little reliance on any cognitive plausibility. Simply – what is the point in analyzing complex tasks when increased memory capacity in computing has made it possible just to use relatively simple search algorithms and brute force to gain better results?

3. One program for many tasks: A third solution would be to construct a single system to perform a diverse collection of small experimental tasks. This suggestion contains the concept of both cognitive architectures, as well as psychometric artificial intelligence, as Newell proposes that an alternative mold would be to construct a single program that could take a standard intelligence test (he recommended WAIS).

Although computational cognitive models were certainly already around before Newell's critique of cognitive psychological experiments in 1973, I (together with many others, such as Bringsjord, 2011; Schimanski, 2004; Langley, Laird, Rogers & Sun, 2008) believe Newell's contribution helped raise them more towards the main streams of cognitive research, where in the sixties, computational cognitive models were the subject of only a handful of scientists (including Newell himself, rather unsurprisingly). Langley et al. (2008) refer to Newell's (1973) text as a "call to arms", which elicited a response in the scientific community and triggered new lines of research and modeling, mainly the specific class of cognitive architectures called *product systems*, and which eventually (with a delay of 30 years or so) also significantly contributed to the development of psychometric artificial intelligence as a relatively underground paradigm within mainstream AI.

PAI, and what psychometrics have to offer for AI

The idea behind PAI is very simple, and articles introducing it often begin with a relatively long account on the concept of psychometric intelligence, with which I assume the target audience here is well acquainted, so I believe I can skip this introductory part and jump right into the core of the theme. The intention of PAI is to construct and evaluate artificial intelligent entities capable of solving psychometric tests. The idea is *not* to model human-like problem solving or processes as in cognitive architectures, but to also employ “any means necessary”-methods, brute force and extensive data searches to reach the goal. In the following paragraphs I will explain why psychometrics and AI make an interesting cocktail.

Generally, psychometrics enable AI people to create models that really genuinely *do* represent intelligence. To me, it feels fair to argue that without the contributions of psychological methodology and psychometrics in actively working to describe, explain and assess intelligence, the concept of intelligence would be nothing but subject to abstract philosophical discussions. If it wasn't for psychology, the general consensus around intelligence might very well be intelligence will never be a subject to concrete conceptualization. It follows, then, that if AI really works with intelligence, they would quickly run out of ingredients without contributions from psychology. However, people (authors, software engineers...) who actually deal with AI are rarely familiar with psychological notions of intelligence, let alone intelligence tests and psychometrics. Really, the field of artificial intelligence provides interesting applications of search heuristics and algorithmic thought, integrated with machine learning mechanisms that guarantee something that appears relatively intelligent behavior in the contexts and functions they were created for. But in my opinion, it is questionable whether it is justifiable to claim that this is artificial *intelligence*. This way, in my opinion, psychometric artificial intelligence represents the core of the AI paradigm as it justifies its very name. Granted that building and evaluating models

capable of psychometric intelligence is more challenging (and in many ways also has less uses) than building other types of models, as well as more problematic as it requires mandatory interdisciplinary collaboration, it is still important specifically for applications where AI attempts to model human intelligence, and should have a more wide spread status in the domain of artificial intelligence.

The notion behind PAI is this: We wish to create and evaluate intelligent entities. Here, we rely on the concept, the definition of intelligence. We can engage in a deep philosophical reflection on the nature and the very existence of intelligence, or we can borrow its definition from a field that routinely works with the notion, which brings us to psychometrics. Some go as far as to argue that since psychometrics deal with tests, a machine can only be intelligent if it can take a psychometric intelligence test. This is rather naive and narrows down the scope of AI, PAI *and* psychometrics: the scope of psychometrics is *not* exclusively testing, as both PAI and psychometrics can be applied for constructs other than intelligence (examples include things such as creativity and mechanical ability), and neither human nor machine intelligence can be boiled down to something as simple as test performance, as displayed by the numerous prominent theories of intelligence in psychology, which include contributions such as those of Sternberg and Gardner, who create a real challenge for assessment.

On top of providing a concrete definition of intelligence that can be actualized computationally, psychometrics has several other useful characteristics for computer modeling. The field is highly methodical and mathematical, making it relatively easy to implement without having to worry about altering the nature of the concept during the implementation, and since it deals with assessment, it naturally provides tools for model creation – all the model really has to be capable of is completing tasks on a test.

Additionally, throughout the 20th century, the ultimate goals for intelligent agents in AI have been to pass the Turing test (Turing, 1950), or the Total Turing Test (Harnad, 1991). The Turing test tests the machine's ability to exhibit intelligent behavior indistinguishable from that of an actual human. The idea is that a human judge engages in a natural language conversation with the machine and a human. If the judge cannot reliably tell which entity is the human and which is the machine, the machine is said to have passed the test. The Total Turing Test is even more demanding, not requiring only human-like communicative capabilities, but also sensorimotor abilities. Where these two tests have essentially been considered the peak of machine intelligence, AI systems rarely (if ever, really) pass them, which is why testing system quality has essentially taken the back seat to engineering, specifically since AI models come in such a large scope of applications (Schimanski, 2004). Psychometrics offers an abundant range of tests with different levels of difficulty, for different application domains, and thus brings testing goodness of the models back to the central role.

Generally, psychometrics offers an opportunity to widen the field of AI, its scope and applications, and an opportunity for interdisciplinary collaboration (which is as an absolute necessity is really both a curse and an opportunity), but its use in AI also has some questionable characteristics. The usefulness of PAI models is questionable in that compared to concepts such as product system cognitive architectures they have a very narrow scope of practical applicative value. The idea behind PAI is *not* to model human-like problem solving like in cognitive architectures (which makes the architectures excellent for things such as workplace and task optimization and interface design) but to simply construct and evaluate intelligent entities capable of taking psychometric tests.

The core issue is this: Is it meaningful to make computer simulations of a something that is the key subject of a school that has developed around assessing psychological traits in humans – entities that are in some ways very different from computers – without actually

attempting to model human-like problem solving? It really depends on the nature of intelligence, and whether or not human thought and computer processing really are that different. It would be interesting to test the view of intelligence as the extent and tendency of the individual to engage in algorithmic, rather than heuristic, problem solving. If this was the case, psychometric artificial intelligence would be only so much more topical. This discussion, however, should take place some other time as it is not the intention of this seminar, nor does it fall in its scope.

One interesting point brought up by Bringsjord and Licato (2012) is that even though PAI is really an extension of the scope of human intelligence tests, there are obvious differences between human and machine intelligence, even when assessed with the same instrument. Where psychometrics essentially deal with comparing intelligent entities (humans) with each other, expressions of psychometric intelligence in intelligent machines is a fundamentally different concept. While in humans, scoring of mental ability is constrained by comparison to other test-takers (where test scores often portray the individual's position in a distribution of the relative population); PAI is founded on a much more coarse-grained view of intelligence tests. A perfect score on a test completed by the machine indicates that it is intelligent *simpliciter*, but it does not indicate any relative position of that test completion in relation to some continuum of other test completions. This implies that certain metric characteristics of tests (such as discriminativeness) can essentially be waned in the case that the model will not be used for things such as test item construction, and shows that psychometrics based on human test participants *do* have implications for AI models, if we can lax some of criteria of quality they must pass when used with human test takers. I wonder, then, with the increasing popularity of AI, if they will once develop an assessment paradigm of their own, similar to psychometrics, but where the goal is exclusively to test intelligent machines.

PERI (Bringsjord & Schimanski 2004)

I mentioned earlier that I would present one relatively recent application of PAI. I chose to present the PERI (Psychometric Experimental Robotic Intelligence) project, as it is virtually the only project I found that strictly adhered to the definition of psychometric AI.

The ambitious goal of the project is to construct a robot, PERI, which would first be capable of solving the entire WAIS intelligence test. PERI is a full system capable of logic and reasoning, vision, physical manipulation, speech and hearing (Schimanski, 2004; Bringsjord & Schimanski, 2004). It is not a system based on cognitive modeling (architecture), but cognitive robotics. PERI was not designed to simulate how a human thinks, or tackle problems in a human-like way, as is the implicit goal of models based on cognitive architectures, but often not the goal in PAI. This is actually a pretty striking characteristic about the paradigm. One would then assume that the domain of psychometric artificial intelligence would be to work with *cognitive architectures*, which are blue prints for creating and assessing intelligent agents with human-like problem solving capabilities – this would make sense, since intelligence tests are obviously designed to assess intelligence in entities with human-like problem solving capabilities (humans, or perhaps humanoid aliens if such creatures were to be found available and willing for testing) – yet PERI, as the very first really ambitious PAI project, deals with robotics. This is interesting, since characteristics of an efficient robot are not human-like, whereas cognitive architectures aim at modeling very human-like characteristics and behavior, and intelligence is by definition a human-like trait. On one hand I'm fascinated by the idea of automated machines solving intelligence tests with data searches and methods of »any means necessary«, but on the other hand, I have a feeling that if the goal of the project was to discover something about human intelligence, then combining the notions of PAI and cognitive architectures would be more adequate.

Psychometric AI in a structure created to model human-like problem solving would obviously be a better description of the typical test taker.

Nevertheless, the authors are (or were) hoping that the robot might once give insight to human intelligence - its complexity and versatility as well as the problems it encounters and can solve. Schimanski (2004) explains that the project is not meant to end when the entire WAIS has been solved, but the rather ambitious goal of the team is to move on to all other established tests (which makes me wonder whether they have actually thought how many thousands of psychometric tests of varying quality there are).

The team chose an embodied model, or a physical implementation (rather than a simulation) in the form of a physical robot, as it allows them to simulate the test taking just as a human respondent would complete the test. Schimanski (2004) explains that she perceived that there is a potentially greater sense of *intelligence* in a physical entity executing some action, rather than watching solutions appear on a computer screen (here she cites Sternberg who argued that intelligence is displayed in actions, not only in thoughts), but she also speculates that while the point where embodiment of the model begins to have an influence on the model's performance is not clear, such a point is likely to exist. I find her explanation for implementing the model as a robot rather than in a simulation grossly vague (if I were to spend several tens of thousands on a robotic arm I'd like to be sure I need it before buying it), but I can think of other benefits from using an embodied model. In applications with cognitive architectures, embodied models have the benefit of interacting directly with interfaces, allowing direct evaluation of the interface as well as better predictions and explanations for user behavior (Ritter & Young, 2001). They have similar implications for psychometric AI, but also allow us to discover other information, such as identification of the extent to which certain sensory processes are necessary for successful completion of a task.

The last update on the PERI project's web page was in 2005. The robot was in a promising stage at that point and was already capable of solving several tasks on WAIS (granted that the arithmetic or digit span tasks really should not represent a problem for a computer). The source code (in Lisp) and a list of the items needed to build the robot are listed on the project website and in Schimanski, 2004, should anyone want to take over and continuing where they left off.

Critique of the current state of the (P)AI paradigm

What I found alarming when running through the literature was that the authors often displayed only a coarse understanding of the structure of theory of intelligence they were working with (for example, Bringsjord and Licato, 2012, attribute the concept of general intelligence and G-factor to Piaget rather than, say, Spearman). Many times they seemed to miss some of the key characteristics of the constructs and only cherry pick the parts that seem suitable to tinker with. Same goes for psychometrics: they seemed to miss the central point of psychometrics as drafting instruments that can be used to assess some kind of a psychological variable, where the domain is obviously typically a person, or humans in general (which obviously represents the constraints for creating such instruments), but rather considered the tests as »care-free tools« for creating their models of machine intelligence. This poses a problem, as uncritical use of the tests gives a skewed image of the type of intelligence (or other construct, be it creativity or something else) the authors attempt to model.

It is unclear as to what exactly many of the researchers consider the domain of psychometrics when they borrow its elements in their applications, but it is clear that some of them take the field more seriously than others. Those more familiar with the field actually attempt to take into account some of its key features (by selecting higher quality tests), whereas others treat the domain as some kind of a playground where they can take some

principles and ignore others without influencing the integrity of the solutions they create that way.

The key concept computer scientists appear to miss when dealing with PAI is construct validity. I believe that behind this problem lies the fact that the definition of model fit or adequacy is so different in computing and in psychology. Grossly simplified, the problem is this: In computer science, if the code compiles and the program computes results that appear mathematically adequate and can realistically be assumed to be a representation of the underlying construct, everything is working fine. For example, in Bringsjord and Licato (2012), the authors took the concept of general intelligence and built a computerized space (the Piaget-McGyver room) where their aim was to test the capability of their computer model's »general-intelligent behavior« by seeing how well it fared in Piagetian tasks of logical thought. Where the idea is ingenious (and creates a good alternative to the overambitious but omnipresent Turing test), it appears that the authors took a psychological construct (general intelligence), created their own model around it, utilizing things they had heard about Piaget's cognitive tasks and theory of cognitive development and processes and did not pay enough attention to other measures – perhaps those designed specifically for the construct of general intelligence – they could have utilized build the model. Fundamentally, there is nothing with this approach, except the fact that the authors claim it a model of general intelligence without citing any pre-existing sources that would draw any parallels between Piaget's cognitive theories and psychological theories of general intelligence based on decades of empirical research. This way, instead of taking a general intelligence model with verified construct validity, the authors created a model that may and may not really model general intelligence, without even understanding the necessity of testing such a claim. What I find the juiciest part of this is that the authors (Bringsjord & Licato, 2012) actually *acknowledge* the

issue of construct validity, but go as far as to argue that *it is not relevant* for psychometric artificial (general) intelligence:

»The field of psychometrics is in reality constrained by the need for construct validity, but in PAGI this constraint is cheerfully defenestrated. Tests that are construct-valid are such that, when successfully taken, ensure that the relevant underlying structures and processes have been active “inside” the agent in question.«
(Bringsjord & Licato, 2012, p. 7).

While I agree that the explicit need to separately implement construct validity can be waned in PAI (assuming that the intelligence tests that are used have construct validity), the authors ignore the fact that they took one construct (general intelligence, or the g factor), and equated it with a *different* theory of cognitive function (Piaget's cognitive development) without even citing the necessary literature to prove that drawing parallels with the two is justifiable. The concept of construct validity in AI may be much more lax than it is in psychometrics, but that does not justify taking the name of one construct, replacing its entire contents with something else, and still claiming that the resulting model represents the original construct, as, in my opinion, was done in Bringsjord and Licato (2012).

Authors of PAI models *also* have to be critical about construct validity, in selecting high quality instruments with empirical backing, just as psychologists must. Simply, they *must* familiarize with the instrument they're about to use, and the psychological construct it's based on, before beginning with the construction of their model. While my opinion about having a minimal understanding of the basic concepts of psychological intelligence before attempting to build computer models to solve intelligence tests may have the fault of being overzealous, I really think that familiarizing with at least fundamental concepts of the psychological models underlying the psychometric instruments used to construct the

computer programs (such as construct validity) should be of key interest to computer scientists aspiring to model psychological constructs with support from decades of empirical work. Models based on such constructs are simply worth more than models based on similar but untested constructs.

This brings me to the point of interdisciplinary collaboration. Where computer scientists have typically worked with cognitive scientists on projects related to modeling some kind of human cognition (including intelligence), the problem here may be that since we're dealing with psychometrics, without a psychologist who is well acquainted with the methodology behind test construction, the team working on the project may lack the expertise to assess whether or not a given instrument is suitable for their application. Having a psychologist in the team would allow them access to most tests, and help the team decide what metric characteristics can be wanted, depending on the intention of the project. This way, PAI represents an area where tight interdisciplinary networking between computer and cognitive scientists and psychologists is simply an absolute necessity. Taking on rigorous study of the concept of psychometrics before taking on the project of using a given test for model construction opens new avenues for computer modelers, and making sure the test they have selected is really fitting for the phenomenon they want to model guarantees that the authors truly can relax the notion of construct validity.

A final issue I'd like to discuss in relation to problems in computational psychology and artificial intelligence is that where many of the paradigms present in the field today were created as a response to Newell's critique on how experimental cognitive psychology was in a crisis due to its fragmentation, Newell's critique is still topical. But this time, rather than being directed at psychology, it should be directed at the field of artificial intelligence. As AI and cognitive science have matured, they have fragmented into several well-defined sub-disciplines, with different interests, goals and independent criteria for evaluation (Langley,

Laird, Rogers & Sun, 2008; Schimanski, 2004), despite the fact that their applications require increasingly integrated intelligent systems, not just improved system components. Newell's critique against setting binary distinction is also relevant to the field, as different schools disagree on whether the model should be theoretical or empirical, embodied or abstract, if processing should be parallel, whether or not it should model neurons and brain function as a base for every application, or if it is arbitrary how the system is constructed, as long as it is a good-enough model of its intelligent agent. In a way, only the platform of Newell's critique has changed, from experimental cognitive psychology to cognitive modeling and AI. While not drawing parallels to Newell's (1973) critique and the current state of AI, Schimanski (2004) also supports the notion that PAI is a necessary component in pulling together the field of artificial intelligence currently comprised of »many different modules that make up flashy independent programs designed to solve independent problems«.

What AI has to offer for psychometrics

AI has some interesting implications for psychometrics – or rather, it could have numerous interesting implications, if the field of PAI was more developed and widespread. I'm going to discuss a few of them next, but keeping in mind the current state of the paradigm, it is obvious that many of these things are outright utopian.

If we were to have a working model capable of solving a pre-existing test with suitable metric characteristics, empirical backing and construct validity, the same or a similar intelligent machine could be altered for the means of item creation, an exhausting task in psychometrics as it calls for resources such as time and money (the use of both of which could be greatly reduced with automation). The machine could come up with a massive number of new items in a relatively short time and offer some primitive types of item analysis even before their administration in a pilot sample, all for virtually no cost on top of that needed to construct the machine – this would work especially well for the so-called culture-free types of

test items. Having some primitive information about the items before even administering them to a pilot sample would save huge amounts of money while potentially allowing for a significant increase in the item pool.

As I suggested, the machine could also offer rough estimations of the metric characteristics of the new items. This could be done with the help of some parameter tweaking. An example (though I cannot say exactly how realistic an example) of such a process would be to employ data space searches executing a given algorithm to find a solution for the task at hand, until the correct solution is found. The system would then be reset and this would be repeated several times. This would allow us to calculate the average number of iterations of the algorithm needed to find a solution to gain a rough estimation of the item's difficulty index. The machine might also select incorrect answers (especially if using a heuristic strategy rather than algorithm), in which case we would gain insight about both the difficulty index and the attractiveness of distractors. Giving the machine data about metric characteristics of already existing and well-researched test items is also a possibility. The machine could then compare the iterations and techniques needed to solve the existing items to those needed to solve newly created items, which might give us insight to how the items would behave in the participant sample that would be used when finally pilot testing the new items.

We could also use different model parameters to simulate different cognitive processes implied in solving tasks on intelligence tests. The parameters could then be altered to display contributions of individual aspects of information processing and e.g. object manipulation in successful task performance. This would have great implications for developing tasks for special needs groups, as it would allow us to isolate items with good metric characteristics but which do not necessarily employ great efforts from a given system of information processing, enabling us to create relatively reliable items for people who typically represent a challenge

for psychometric assessment (e.g. those with a neurological impairment that influences some aspect of their cognitive processing but not others, such as sequence planning or attention). Similarly, extending this principle to embodied models would allow us to isolate task components that demand given sensorimotor skills.

None of this is really possible with the current state of PAI and AI in general. One could say that the entire field of AI is in a crisis of fragmentation. Perhaps it is time for another Allen Newell to step up and suggest new paradigms to try and mitigate the lack of holistic direction in computer modeling. The previous critique led to the triumph of product systems and other cognitive architectures – perhaps this time would give birth to something equally interesting and applicable but with a different orientation.

References

- Bringsjord, S & Licato, J. (2012). *Psychometric Artificial General Intelligence: The Piaget-MacGyver Room*. Atlantis Press Preview, Rensselaer Polytechnic Institute (RPI), Troy NY 12180 USA.
- Bringsjord, S. & Schimanski, B. (2004) *Psychometric AI: "Pulling it All Together" at the Human Level*. AAAI Fall 2004 Symposium Series: Achieving Human-Level Intelligence Through Integrated Systems and Research.
- Bringsjord, S. (2011). Psychometric artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 23(3), 271-277.
- Evans, G. (1968). A program for the solution of a class of geometry-analogy intelligence-test questions, in M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, AM, pp. 271-353.
- Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old
- Langley, P., Laird, J.E., Rogers, S. & Sun, R. (2008). Cognitive architectures: research issues and challenges. *Cognitive Systems Research*, doi:10.1016/j.cogsys.2006.07.004
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp.283-308). New York: Academic Press
- philosophical problem, *Minds and Machines* 1(1), 43–54.
- Ritter, F. E. & Young, M. R. (2001). Embodied models as simulated users: introduction to the special issue on using cognitive models to improve interface design. *International Journal of Human-Computer Studies*, 55, 1-14

Schimanski, B. (2004). *Introducing Psychometric AI, and the robot PERI, its first implementation*. Rensselaer Polytechnic Institute, Troy, New York.

Turing, A. (1950). Computing machinery and intelligence, *Mind LIX*(236), 433–460.