Chapter 1

Psychometric Artificial General Intelligence: The Piaget-MacGuyver Room*

Selmer Bringsjord & John Licato

Department of Computer Science
Department of Cognitive Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

Psychometric AGI (PAGI) is the brand of AGI that anchors AGI science and engineering to explicit tests, by insisting that for an information-processing (i-p) artifact to be rationally judged generally intelligent, creative, wise, and so on, it must pass a suitable, well-defined test of such mental power(s). Under the tent of PAGI, and inspired by prior thinkers, we introduce the *Piaget- MacGyver Room* (PMR), which is such that, an i-p artifact can credibly be classified as general-intelligent if and only if it can succeed on *any* test constructed from the ingredients in this room. No advance notice is given to the engineers of the artifact in question, as to what the test is going to be; only the ingredients in the room are shared ahead of time. These ingredients are roughly equivalent to what would be fair game in the testing of neurobiologically normal Occidental students to see what stage within his theory of cognitive development they are at. Our proposal and analysis puts special emphasis on a kind of cognition that marks Piaget's Stage IV and beyond: viz., the intersection of hypothetico-deduction and analogical reasoning, which we call *analogico-deduction*).

2   *Selmer Bringsjord & John Licato*

1.1. Introduction

Psychometric AGI (PAGI; pronounced "pay guy"), in a nutshell, is the brand of AGI that anchors AGI science and engineering to explicit tests, by insisting that for an information-processing[a] (i-p) artifact to be rationally judged generally intelligent, creative, wise, and so on, the artifact must be capable of passing a suitable, well-defined test of such mental power(s), even when it hasn't seen the test before. (PAGI is built upon PAI, psychometric AI; see (Bringsjord and Schimanski, 2003).) For example, someone might claim that IBM's i-p artifact Deep Blue is really and truly intelligent, in light of the fact that if you test it by seeing whether it can prevail against the best human chessplayers, you will find that it can. And someone might claim that natural-language-processing artifact Watson, another i-p artifact from IBM (Ferrucci, Brown, Chu-Carroll, Fan, Gondek, Kalyanpur, Lally, Murdock, Nyberg, Prager, Schlaefer and Welty, 2010), is really and truly intelligent because it can vanquish human opponents in the game of *Jeopardy!*. However, while both of these artifacts are intelligent *simpliciter*, they most certainly aren't *general*-intelligent. Both Deep Blue and Watson were explicitly engineered to specifically play chess and Jeopardy!, nothing more; and in both cases the artifacts knew what their final tests would be.

Inspired by PAGI, and by a line of three thinkers (Descartes, Newell, and esp. Piaget) who gave much thought to the hallmarks of *general* intelligence, we define a room, the *Piaget-MacGyver Room* (PMR), which is such that, an i-p artifact can credibly be classified as general-intelligent if and only if it can succeed on *any* test constructed from the ingredients in this room. *No advance notice is given to the engineers of the artifact in question as to what the test is going to be.* This makes for rather a different situation than that seen in the case of both Deep Blue and Watson; for in both of these cases, again, the AI engineering that produced these i-p artifacts was guided by a thorough understanding and analysis, ahead of time, of the tests in question. In fact, in both cases, again, all along, the engineering was guided by repeatedly issuing pre-tests to both artifacts, and measuring their performance with an eye to making incremental improvements. This is particularly clear in the case of Watson; see (Ferrucci, Brown, Chu-Carroll, Fan, Gondek, Kalyanpur, Lally, Murdock, Nyberg, Prager, Schlaefer and Welty, 2010). Of course, we happily concede that both Deep Blue and Watson *are* intelligent; we just don't believe that either is *general*-intelligent.[b]

As we say, only the *ingredients* in PMR are shared ahead of time with the relevant engineers. These ingredients are equivalent to what would be fair game in the testing, by Piaget, of a neurobiologically normal Occidental student who has reached at least Piaget's

---

[a]By using 'information-processing' rather than 'computational' we leave completely open the level of information-processing power — from that of a standard Turing machine, to so-called "hypercomputers" — the artifact in question has. Note that we also for the most part steer clear of the term 'agent,' which is customary in AI. We do so because 'agent' is usually taken to imply a function that is Turing-computable or easier; e.g., see the use of 'agent' in (Russell and Norvig, 2002).

[b]Our attitude is anticipated e.g. by Penrose, who for instance pointed out that Deep Blue would be paralyzed if challenged on the spot to play variants of chess; see (Penrose, 1994). In the case of Watson, questions based on neologisms would paralyze the system. E.g., "Supposing the bloogering! is to take a prime and blooger it (add it to itself), and then blooger thrice more times, what is bloogering! 7?"

Stage III of cognitive development. If you will, Piaget is in control of the ingredients in the room, and, with a general understanding of the MacGyver television series,[c] and of course with an understanding of his own account of cognitive development, assembles from the ingredients in the room a test of an artificial agent that is purportedly general-intelligent. For example, Piaget might be "armed" with the ingredients shown in Figure 1.1.
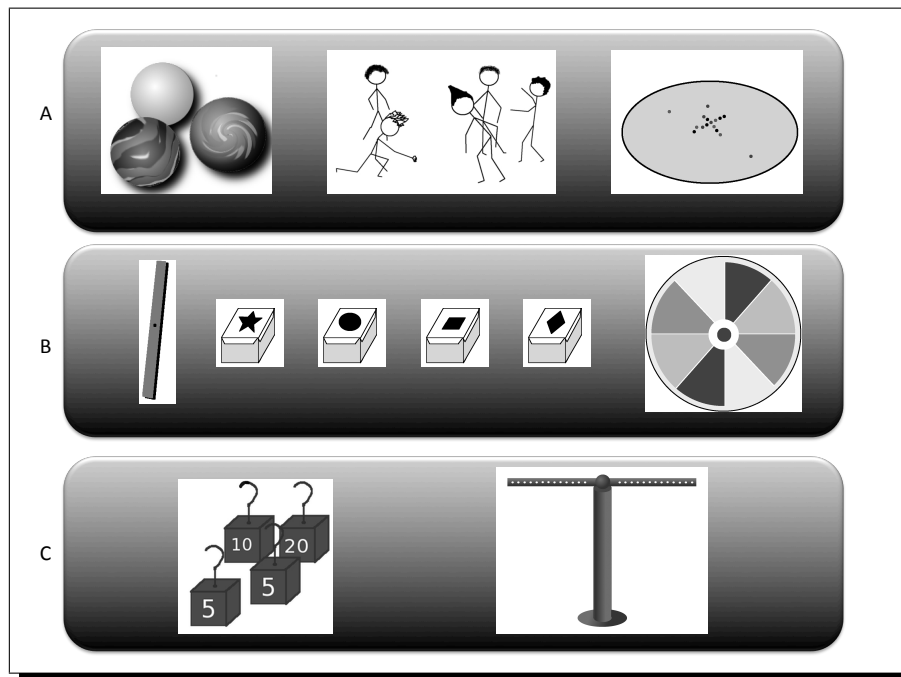

Fig. 1.1. A Possible Set of Ingredients From Which Piaget Can Work (weights, marbles (and its playing field), human confederates, familiar shapes, magnets, etc.)

If the artifact passes what Piaget assembles, we can safely say that it's indeed general-intelligent; if it fails, we declare that it isn't. We shall allow a range of responses that fall into these two categories, since some with the general intelligence of a Feynman, after being given the test, might well be able to find an abundance of solutions.[d] As will be seen below, our proposal and analysis puts special emphasis on cognition that marks (Piagetian) Stage IV and beyond: viz., the intersection of hypothetico-deduction and analogical reasoning (which we call *analogico-deduction*). In hypothetico-deduction one creates hypotheses $h_1, h_2, \ldots, h_n$, conditionals of the form $h_i \rightarrow r_i$, and then tests to see whether the

---

[c]For a description of the series, see: http://en.wikipedia.org/wiki/MacGyver. The hero and protagonist, Mac-Gyver, is stunningly resourceful, and hence hard to trap, seriously injure, or kill; he always manages to carry out some physical manipulation that solves the problem at hand. In the harder of Piaget's tests, a high degree of resourcefulness is a *sine qua non*; and the tests invariably call a la MacGyver for *physical manipulation* that confirms the resourcefulness.

[d]See http://blogs.msdn.com/b/ericlippert/archive/2011/02/14/what-would-feynman-do.aspx).

results r_i do indeed obtain, following upon an instantiation of h_i. If r_i doesn't obtain, *modus tollens* immediately implies that h_i is to be rejected. When analogical reasoning undergirds either the generation of the hypotheses or the conditionals, or the negation of a result r_i, the overall process falls under analogico-deduction. In order to focus matters we shall restrict our attention to not only such reasoning, but to such reasoning applied to a representative test fashioned by Piaget: his ingenious magnet test.

The plan of the chapter is as follows. We first (§1.2) explain in a bit more detail what PAGI is, and why the roots of this brand of AGI are to be explicitly found in the thinking of Newell, and before him, in two tests described by Descartes. We then (§1.3) look at these two tests in a bit more detail. Next, in section 1.4, we give a barbarically quick overview of Piaget's view of thinking, and present his magnet challenge. We then (§1.5) briefly describe the LISA system for modeling analogical reasoning. Following this, we model and simulate, using the linked information-processing system Slate+LISA (Slate is an argument-engineering environment that can be used in the purely deductive mode for *proof* engineering), human problem-solving cognition in the magnet challenge (§1.6). A brief pointer to the next research steps in the PAGI research program in connection with PMR (which, ultimately, we have every intention of building and furnishing), wraps up the chapter.

1.2. More on Psychometric AGI

Rather long ago, Newell (1973) wrote a prophetic paper: "You Can't Play 20 Questions with Nature and Win." This paper helped catalyze both modern-day computational cognitive modeling through cognitive architectures (such as ACT-R, NARS, Soar, Polyscheme, etc.), and AI's — now realized, of course — attempt to build a chess-playing machine better at the game than any human. However, not many know that in this paper Newell suggested a *third* avenue for achieving general machine intelligence, one closely aligned with psychometrics, and one — as we shall see — closely aligned as well with the way Piaget uncovered the nature of human intelligence. In the early days of AI, at least one thinker started decisively down this road for a time (Evans 1968); but now the approach, it may be fair to say, is not all that prominent in AI. We refer to this approach as *Psychometric AGI*, or just PAGI (rhymes with "pay guy").

1.2.1. *Newell & the Neglected Route Toward General Machine Intelligence*

In the "20 Questions" paper, Newell bemoans the fact that, at a symposium gathering together many of the greatest psychologists at the time, there is nothing whatsoever to indicate that any of their work is an organized, integrated program aimed seriously at uncovering the nature of intelligence as information processing. Instead, Newell perceives a situation in which everybody is carrying out work (of the highest quality, he cheerfully admits) on his or her own specific little part of human cognition. In short, there is nothing that, to use Newell's phrase, "pulls it all together." He says: "We never seem in the experimental literature to put the results of all the experiments together." (1973: 298) After making clear

that he presupposes that "man is an information processor," and that therefore from his perspective the attempt to understand, simulate, and replicate human intelligence is by definition to grapple with the challenge of creating machine intelligence, Newell offers three possibilities for addressing the fragmentary nature of the study of mind as computer.

The first possibility Newell calls "Complete Processing Models." He cites his own work (with others; e.g., Simon; the two, of course, were to be a dynamic duo in AI for many decades to come) based on production systems, but makes it clear that the production-system approach isn't the only way to go. Of course today's cognitive architectures [e.g., NARS (Wang, 2006); SOAR (Rosenbloom, Laird and Newell, 1993); ACT-R (Anderson, 1993; Anderson and Lebiere, 1998; Anderson and Lebiere, 2003); Clarion (Sun, 2001); and Polyscheme (Cassimatis, 2002; Cassimatis, Trafton, Schultz and Bugajska, 2004)] can be traced back to this first possibility.

The second possibility is to "Analyze a Complex Task." Newell sums this possibility up as follows.

> A second experimental strategy, or paradigm, to help overcome the difficulties enumerated earlier is to accept a single complex task and do all of it ... the aim being to demonstrate that one has a significant theory of a genuine slab of human behavior. ... A final example [of such an approach] would be to take chess as the target super-task. (Newell 1973: 303–304)

This second possibility is one most people in computational cognitive science and AI are familiar with. Though Deep Blue's reliance upon standard search techniques having little cognitive plausibility perhaps signaled the death of the second avenue, there is no question that, at least for a period of time, many researchers were going down it.

The third possibility, "One Program for Many Tasks," is the one many people seem to have either largely forgotten or ignored. Newell described it this way:

> The third alternative paradigm I have in mind is to stay with the diverse collection of small experimental tasks, as now, but to construct a single system to perform them all. This single system (this model of the human information processor) would have to take the instructions for each, as well as carry out the task. For it must truly be a single system in order to provide the integration we seek. (Newell 1973: 305)

For those favorably inclined toward the test-based approach to AI or AGI, it's the specific mold within Newell's third possibility that is of acute interest. We read:

> A ... mold for such a task is to construct a single program that would take a standard intelligence test, say the WAIS or the Stanford-Binet. (Newell 1973: 305)

We view this remark as a pointer to PAGI, and to a brief explication of this brand of AGI we now turn.

6   *Selmer Bringsjord & John Licato*

1.2.2. *So, What is Psychometric AGI?*

What is AI? We'd be willing to wager that many of you have been asked this question — by colleagues, reporters, friends and family, and others. Even if by some fluke you've dodged the question, perhaps you've asked it yourself, maybe even perhaps (in secret moments, if you're a practitioner) *to* yourself, without an immediate answer coming to mind. At any rate, AI *itself* repeatedly asks the question — as the first chapter of many AI textbooks reveals. Unfortunately, many of the answers standardly given don't ensure that AI tackles head on the problem of *general* intelligence (whether human or machine). For instance, Russell and Norvig (2002) characterize AI in a way (via functions from percepts to actions; they call these functions *intelligent agents*) that, despite its many virtues, doesn't logically entail any notion of generality whatsoever: An agent consisting solely in the factorial function qualifies as an intelligent agent on the R-N scheme. Our answer, however, is one in line with Newell's third possibility, and one in line with a perfectly straightforward response to the "What is AI?" question.

To move toward our answer, note first that presumably the 'A' part of 'AGI' isn't the challenge: We seem to have a fairly good handle on what it means to say that something is an artifact, or artificial. It's the 'G' and the 'I' parts that seem to throw us for a bit of a loop. First, what's intelligence? *This* is the first of the two big, and hard, questions. Innumerable answers have been given, but many outside the test-based approach to AI seem to forget that there is a particularly clear and straightforward answer available, courtesy of the field that has long sought to operationalize the concept in question; that field is psychometrics. Psychometrics is devoted to systematically measuring psychological properties, usually via tests. These properties include the ones most important in the present context: both intelligence, and general intelligence. In a nutshell, the initial version of a psychometrics-oriented account of general intelligence (and this definition marks an answer to the second big question: What's *general* intelligence?) is simply this: Some i-p artifact is intelligent if and only if it can excel at *all* established, validated tests of neurobiologically normal cognition, even when these tests are new for the artifact.

Of course, psychometrics is by definition devoted to specifically defining and measuring *human* intelligence; the SAT, part of the very fabric of education in the United States (with its counterparts in widespread use in other technologized countries), is for example administered to humans, not machines. Some hold that it's neither possible nor necessary for AI to be identical to human intelligence. After all, it seems possible for an AI to have a vision system that covers a different frequency range of light, compared to of a normal human. Consequently, such a system may fail some tests that require color recognition. In this context, someone might object: "What makes the two of you think, then, that intelligence tests can be sensibly applied to i-p artifacts?"

Full analysis and rebuttal of this worry would occupy more space than we have; we must rest content with a brief response, via three points:

(1) Judgments regarding whether i-p artifacts are intelligent are already informally, but firmly, rooted in the application of tests from the human sphere. We know that Kasparov is quite an intelligent chap; and we learned that Deep Blue, accordingly, is intelligent; a parallel moral

emerged from the victory of Watson. We are simply, at bottom, extending and rigorizing this already-established human-centric way of gauging machine intelligence.

(2) 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. But in PAGI, the bottom line is "getting the job done," and in fact we assume that i-p machines will, "under the hood," depart from human techniques.

(3) The third point in our answer flows from the second, and is simply a reminder that while in the human sphere the scoring of tests of mental ability is indeed constrained by comparison to other human test-takers (an IQ "score," after all, is meaningless without relative comparison to other humans who take the relevant test), PAGI is founded upon a much more coarse-grained view of intelligence tests — a view according to which, for instance, a perfect score on the part of an i-p artifact indicates that it's intelligent *simpliciter*, not that it's intelligent within some human-centric continuum. This general point applies directly to PMR: For example, prowess in PMR specifically requires sensorimotor prowess, but not *human* sensorimotor adroitness. We assume only that one side of general intelligence, as that concept covers both human and i-p machine, is perceiving and moving, in planful ways, physical objects.

We anticipate that some will insist that while intelligence tests are sensibly applicable to i-p artifacts in principle, the fact remains that even broad intelligence tests are still just too narrow, when put in the context of the full array of cognitive capacities seen in *homo sapiens*. But one can understand general intelligence, from the standpoint of psychometrics, to include many varied, indeed for that matter all, tests of intellectual ability. Accordingly, one can work on the basis of a less naïve definition of PAGI, which follows.[e]

> Psychometric AGI is the field devoted to building i-p artifacts capable of at least solid performance on all established, validated tests of intelligence and mental ability, without having seen these tests beforehand at all; the class of tests in play here includes not just the rather restrictive IQ tests, but also tests of the many different forms of intelligence seen in the human sphere.[f]

This definition, when referring to tests of mental ability, is pointing to much more than IQ tests. For example, following Sternberg (1988), someone with much musical aptitude would count as brilliant even if their scores on tests of "academic" aptitude (e.g., on the SAT, GRE, LSAT, etc.) were low. Nonetheless, even if, hypothetically, one were to restrict attention in PAGI to intelligence tests, a large part of cognition would be targeted. Along this line, in choosing the WAIS, Newell knew what he was doing.

To see this, we begin by going back to the early days of AI, specifically to a time when Psychometric AI was at least implicitly entertained. For example, in the mid 1960s, the largest Lisp program on earth was Evans' (1968) ANALOGY program, which could solve problems like those shown in Figure 1.2. Evans himself predicted that systems able

---

[e]For more on PAI, which of course forms the foundation for PAGI, readers can consult a recent issue of the *Journal of Experimental and Theoretical Artificial Intelligence* devoted to the topic: **23.3**.

[f]The notion that intelligence includes more than *academic* intelligence is unpacked and defended by numerous psychologists. E.g., see (Sternberg, 1988).

8   *Selmer Bringsjord & John Licato*
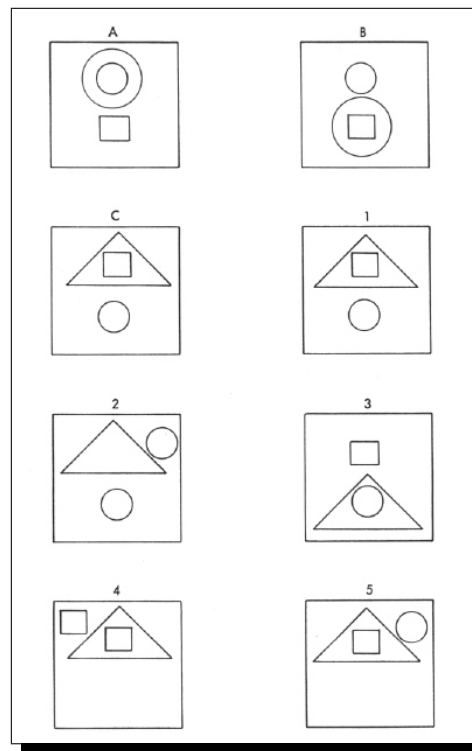

Fig. 1.2.   Sample Problem Solved by Evan's (1968) ANALOGY Program. Given sample geometric configurations in blocks A, B, and C, choose one of the remaining five possible configurations that completes the relationship: *A is to B as C is to …?*. Subjects asked to prove that their answers are correct must resort to analogico-deduction.

to solve such problems would "be of great practical importance in the near future," and he pointed out that performance on such tests is often regarded to be the "touchstone" of human intelligence. However, ANALOGY simply hasn't turned out to be the first system in a longstanding, comprehensive research program (Newellian or otherwise). Why is this? Given our approach and emphasis, this question is a penetrating one. After all, we focus on analogical reasoning, and ANALOGY certainly must be capable of such reasoning. (There is no deduction required by a program able to solve problems in the class in question, but if the artifact was asked to rigorously justify its selection, deduction would unstoppably enter the picture.) So again: Given that Evans was by our own herein-advertised lights on the right track, why the derailment?

    We think the main reason is summed up in this quote from Fischler & Firschein (1987):

> If one were offered a machine purported to be intelligent, what would be an appropriate method of evaluating this claim? The most obvious approach might be to give the machine an IQ test. … However, [good performance on tasks seen in IQ tests would not] be completely satisfactory *because the machine would have to be specially prepared for any specific task that it was asked to perform.* The task could not be described to the machine in a normal conversation (verbal

or written) if the specific nature of the task was not already programmed into the machine. Such considerations led many people to believe that the ability to communicate freely using some form of natural language is an essential attribute of an intelligent entity. (Fischler & Firschein 1987, p. 12; emphasis ours)

1.2.3. *Springboard to the Rest of the Present Paper*

Our response to this response is three-fold. One, there is nothing here that tells against the suspicion that the marriage of analogical and deductive reasoning, which is specifically called for by problems of the sort that the ANALOGY system solved, is at the heart of general intelligence, whether that intelligence is embodied in the mind of a person or machine. Two, a test-based approach to AI can, despite what F&F say, take full account of the requirement that a truly intelligent computing machine must not simply be pre-programmed. Indeed, this is one of the chief points of the PMR. And finally, three, a test-based approach to uncovering the nature of human intelligence, when broadened in the manner of Piaget, provides a suitable guide to engineering aimed at producing artificial general intelligence.

At this point the reader has sufficient understanding of PAGI to permit us to move on.[g]

1.3. Descartes' Two Tests

Descartes was quite convinced that animals are mechanical machines. He felt rather differently about persons, however: He held that persons, whether of the divine variety (e.g., God, the existence of whom he famously held to be easily provable) or the human, were *more* than mere machines.

Someone might complain that Descartes, coming before the likes of Turing, Church, Post, and Gödel, could not have had a genuine understanding of the concept of a *computing* machine, and therefore couldn't have claimed the human persons are more than such machines. There are two reasons why this complaint falls flat. One, while we must admit that Descartes didn't *exactly* have in the mind the concept of a computing machine in the manner of, say, of a universal Turing machine, or a register machine, and so on, what he did have in mind would subsume such modern logico-mathematical devices. For Descartes, a machine was overtly mechanical; but there is a good reason why recursion theory has been described as revolving around what is *mechanically solvable*. A Turing machine, and ditto for its equivalents (e.g., register machines) are themselves overtly mechanical.

Descartes suggested two tests to use in order to separate mere machines from human persons. The first of these directly anticipates the so-called "Turing Test." The second test is the one that anticipates the Piaget-MacGyver Room. To see this, consider:

> If there were machines which bore a resemblance to our body and imitated our actions as far as it was morally possible to do so, we should always have two very certain tests by which to recognize that, for all that, they were not real men. The first is, that they could never use speech or other signs as we do when placing

---

[g]For more on PAI, the foundation for PAGI, readers can consult a recent issue of the *Journal of Experimental and Theoretical Artificial Intelligence* devoted to the topic: 23.3.

> our thoughts on record for the benefit of others. For we can easily understand
> a machine's being constituted so that it can utter words, and even emit some
> responses to action on it of a corporeal kind, which brings about a change in
> its organs; for instance, if it is touched in a particular part it may ask what we
> wish to say to it; if in another part it may exclaim that it is being hurt, and so
> on. But it never happens that it arranges its speech in various ways, in order to
> reply appropriately to everything that may be said in its presence, as even the
> lowest type of man can do. And the second difference is, that although machines
> can perform certain things as well as or perhaps better than any of us can do,
> they infallibly fall short in others, by which means we may discover that they
> did not act from knowledge, but only for the disposition of their organs. For
> while reason is a universal instrument which can serve for all contingencies, these
> organs have need of some special adaptation for every particular action. From this
> it follows that it is morally impossible that there should be sufficient diversity in
> any machine to allow it to act in all the events of life in the same way as our
> reason causes us to act. (Descartes 1911, p. 116)

We now know all too well that "machines can perform certain things as well or perhaps
better than any of us" (witness Deep Blue and Watson, and perhaps, soon enough, say,
auto-driving cars that likewise beat the pants off of human counterparts); but we also know
that these machines are engineering for specific purposes that are known inside and out
ahead of time. PMR is designed specifically to test for the level of proficiency in using
what Descartes here refers to as a "universal instrument." This is so because PMR inherits
Piaget's focus on general-purpose reasoning. We turn now to a brief discussion of Piaget
and this focus.

1.4. Piaget's View of Thinking & The Magnet Test

Many people, including many outside psychology and cognitive science, know that Piaget
seminally — and by Bringsjord's lights, correctly — articulated and defended the view
that mature human reasoning and decision-making consists in processes operating for the
most part on formulas in the language of classical extensional logic (e.g., see Inhelder and
Piaget, 1958b).[h] You may yourself have this knowledge. You may also know that Piaget
posited a sequence of cognitive stages through which humans, to varying degrees, pass; we
have already referred above to Stages III and IV. How many stages are there, according to
Piaget? The received answer is: four; in the fourth and final stage, *formal operations*, neu-
robiologically normal humans can reason accurately and quickly over formulas expressed
in the logical system known as first-order logic, $\mathscr{L}_I$. This logic allows for use of relations,
functions, the universal and existential quantifiers, the familiar truth-functional connectives
from the propositional calculus, and includes a so-called "proof theory," that is, a mechan-

---

[h]Many readers will know that Piaget's position long ago came under direct attack, by such thinkers as Wason
and Johnson-Laird (Wason, 1966; Wason and Johnson-Laird, 1972). In fact, unfortunately, for the most part
academics believe that this attack succeeded. Bringsjord doesn't agree in the least, but this isn't the place to visit
the debate in question. Interested readers can consult (Bringsjord, Bringsjord and Noel, 1998; Rinella, Bringsjord
and Yang, 2001). Piaget himself retracted any claims of *universal* use of formal logic: (Piaget, 1972).

ical method for deriving some formulas from others.[i] One cornerstone of every classical proof theory, as the reader will likely well know, is *modus ponens*, according to which the formula ψ can be derived from the formulas ϕ and $\phi \rightarrow \psi$ (read: if ϕ then ψ).
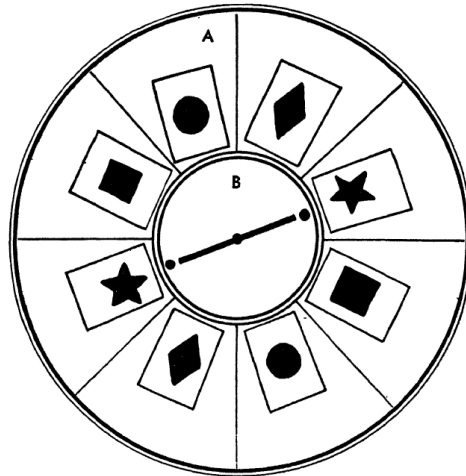

Fig. 1.3.   Piaget's famous "rigged" rotating board to test for the development of Stage-III-or-better reasoning in children. The board, A, is divided into sectors of different colors and equal surfaces; opposite sectors match in color. B is a rotating disk with a metal rod spanning its diameter—but the catch is that the star cards have magnets buried under them (hidden inside wax), so the alignment after spinning is invariably as shown here, no matter how the shapes are repositioned in the sectors (with matching shapes directly across from each other). This phenomenon is what subjects struggle to explain. Details can be found in (Inhelder and Piaget, 1958b).

Judging by the cognition taken by Piaget to be stage-III or stage-IV (e.g., see Figure 1.3, which shows one of the many problems presented to subjects in (Inhelder and Piaget, 1958b)), the basic scheme is that an agent $\mathscr{A}$ receives a problem P (expressed as a visual scene accompanied by explanatory natural language), represents P in a formal language that is a superset of the language of Ł$_I$, producing $[P]$, and then reasons over this representation (along with background knowledge, which we can assume to be a set Γ of formal declarative statements) using at least a combination of some of the proof theory of Ł$_1$ and "psychological operators."[j] This reasoning allows the agent to obtain the solution $[S]$. To ease exposition, we shall ignore the heterodox operations that Piaget posits (see note j) in favor of just standard proof theory, and we will moreover view $[P]$ as a triple (ϕ, C, Q), where ϕ is a (possibly complicated) formula in the language of $\mathscr{L}_I$, C is further information that provides context for the problem, and consists of a set of first-order

---

[i]A full overview of logic, $\mathscr{L}_I$ included, in order to model and simulate large parts of cognition, can be found in (Bringsjord, 2008).

[j] The psychological operators in question cannot always be found in standard proof theories. For example, Piaget held that the quartet I N R C of "transformations" were crucial to thought at the formal level. Each member of the quartet transforms formulas in certain ways. E.g., N is *inversion*, so that N$(p \vee q) = \neg p \wedge \neg q$; this seems to correspond to DeMorgan's Law. But R is *reciprocity*, so R$(p \vee q) = \neg p \vee \neg q$, and of course this isn't a valid inference in the proof theory for the propositional calculus or $\mathscr{L}_I$.

Selmer Bringsjord & John Licato

formulas, and Q is a query asking for a proof of ϕ from $C \cup \Gamma$. So:

$$[P] = (\phi, C, Q = C \cup \Gamma \vdash \phi?)$$

At this point a reader might be puzzled about the fact that what we have so far described is exclusively deductive, given that we have said that our focus is reasoning that includes not just deduction, but also analogical reasoning; the key term, introduced above, is *analogico-deduction*. To answer this, and to begin to give a sense of how remarkably far-reaching Piaget's magnet challenge is, first consider how this deduction-oriented scheme can be instantiated.

To begin, note that in the invisible magnetization problem shown in Figure 1.3, which requires stage-III reasoning in order to be solved, the idea is to explain how it is that $\phi^{\star\star}$, that is, that the rotation invariably stops with the two stars selected by the rod. Since Piaget is assuming the hypothetico-deductive method of explanation made famous by Popper (Popper, 1959), to provide an explanation is to rule out hypotheses until one arrives deductively at $\phi^{\star\star}$. In experiments involving child subjects, a number of incorrect (and sometimes silly) hypotheses are entertained—that the stars are heavier than the other shaped objects, that the colors of the sections make a difference, and so on. Piaget's analysis of those who discard mistaken hypotheses in favor of $\phi^{\star\star}$ is that they expect consequences of a given hypothesis to occur, note that these consequences fail to obtain, and then reason backwards by *modus tollens* to the falsity of the hypotheses. For example, it is key in the magnet experiments of Figure 1.3 that "for some spins of the disk, the rod will come to rest upon shapes other than the stars" is an expectation. When expectations fail, disjunctive syllogism allows $\phi^{\star\star}$ to be concluded. However, the reasoning patterns so far described are only those at the "top level," and even at that level exclude the *generation* of hypotheses. Beneath the top level, many non-deductive forms of reasoning are perfectly compatible with Piaget's framework, and one thing that is crystal clear on a reading of his many experiments is that subjects draw from past experience to by analogy rule our hypotheses, and to generate hypotheses in the first place.

Hence the magnet challenge, like other famous challenges invented and presented by Piaget, is a portal to a remarkably wide landscape of the makings of general intelligence. This is confirmed not just by taking account of the magnet challenge in the context of Piaget's framework, and more generally in the context of deliberative reasoning and decision-making; it's also confirmed by placing the magnet challenge (and counterparts that can be fashioned from the raw materials for PMR) in the context of broad characterizations of intelligence offered even by AI researchers more narrowly oriented than AGI researchers. For example, the magnet challenge taps many elements in the expansive, more-than-deduction view of rational intelligence laid out by Pollock (1989), and likewise taps much of the functionality imparted to the more sophisticated kinds of agents that are pseudo-coded in (Russell and Norvig, 2009).

As will soon be seen, our modeling and simulation of the magnet challenge reflects its requiring much more than straight deduction. But before moving in earnest to that modeling and simulation, we provide a rapid overview of the system we use for analogical reasoning: LISA.

1.5. The LISA model

LISA (*Learning and Inference with Schemas and Analogies*) model is the formidable fruit of an attempt to create a neurally-plausible model of analogical reasoning by using a hybrid connectionist and symbolic architecture (Hummel and Holyoak, 2003a; Hummel and Holyoak, 2003b). We here provide only a very brief summary of some relevant features of LISA; for a more detailed description the reader is directed to (Hummel & Holyoak 2003) and (Hummel & Landy 2009).

LISA allows for explicit representation of propositional knowledge, the arguments of which can be either token objects or other propositions.[k] Propositional knowledge is organized into *analogs*, which contain the proposition nodes, along with other related units: the sub-propositional units which help to bind relational roles within propositions to their arguments, nodes representing the objects (one object unit corresponds to a token object across all propositions within an analog), predicate units which represent the individual roles within a proposition, and higher-level groupings of propositions (Hummel and Landy, 2009). Semantic units, which are outside of and shared by all of the analogs, connect to the object and predicate units.

In self-supervised learning, LISA performs analogical inference by firing the propositional units in a preset order, which propagates down to the semantic units. This allows for units in different analogs to be temporarily mapped to each other if they fire in synchrony, and for new units to be recruited (or *inferred*) if necessary. Of course, many details are left out here in the interests of space; for more, see (Hummel & Holyoak 2003).

1.6. Analogico-Deductive Reasoning in the Magnet Test

The ways in which analogical and deductive reasoning interact in a typical human reasoner are, we concede, complex to the point of greatly exceeding any reasoning needed to excel in the narrower-than-real-life PMR; and, in addition, these ways no doubt vary considerably from person to person. A model such as the one we present here can thus only hope to be a simulation of a *possible* way that a reasoner might solve a problem on the order of the magnet challenge and its relatives.

This said, and turning now to the Piagetian task on which we focus, we first note again that analogical reasoning can often be useful in generation of hypotheses and theories to explain unfamiliar phenomena. For example, Holyoak et al. (2001) explain that the wave theory of sound, as it became better understood, was the basis for creating an analogy that described the wave theory of light. Such an analogical mapping would presumably be responsible for inferring the existence of a medium through which light would travel, just as sound needs air or something like it (indeed, the luminiferous aether was of course proposed to be this very medium). In contrast, Newton's particle theory of light would provide an analogical mapping that would not require a medium. Thus, we have two different analogical mappings; and each then suggests slightly different groups of hypotheses, members

---

[k]E.g., *knows*(*Tom*, *loves*(*Sally*, *Jim*)).

of which, in both cases, could in turn be tested with a combination of experimentation and deductive reasoning.

Now let's get more specific. Analogico-deductive reasoning in the Piagetian hidden-magnet experiment can be modeled using LISA and Slate together; specifically, a dialogue between an experimenter and a subject referred as 'Gou' provides an interesting basis for doing so (Inhelder and Piaget, 1958a). Gou, who is developmentally in Piaget's concrete operations stage (Stage III), after being presented with the hidden-magnets challenge, does from the start suspect that magnets are responsible — but quickly abandons this hypothesis in favor of the one claiming that the weight of the objects is what leads the needle to repeatedly stop on the stars. The experimenter then asks Gou what he would have to do in order to "prove that it isn't the weight," to which Gou responds by carrying out a series of small experiments designed to prove that weight isn't responsible for the bar's stopping. One of these experiments involves removing the star and diamond boxes, and checking to see if the bar still stops on the heaviest of the remaining boxes. Predictably (given *our* understanding of the background's mechanisms), it does not; this provides Gou with empirical evidence that weight is not causally responsible for the bar's stopping as it invariably does (although he continues to subsequently perform small experiments to further verify that weight is not responsible).

In our overall model of Gou's reasoning as being of the analogico-deductive variety, we of course must make use of both deduction and analogical reasoning, woven together. The overall reasoning abides by the deductive framework known as *proof by cases*, which is straightforward and bound to be familiar to all our readers. The core idea is that if one knows that a disjunction

$$\phi_1 \vee \phi_2 \vee \ldots \vee \phi_n$$

holds, and knows as well that one or more of the disjuncts ϕ_i fail to hold, then one can infer a new disjunction lacking the false disjuncts. In the case at hand, in light not only of Gou's thinking, but that of many other subjects, there are four hypotheses in play, as follows (with underlying S-expressions in FOL given in each case).[1]

H1 Weight accounts for the invariance.

- `(Initially boardWeighted)`

H2 Color accounts for the invariance.

- `(Initially boardColored)`

H3 Magnets account for the invariance.

- `(Initially boardMagnetized)`

H4 Order accounts for the invariance.

- `(Initially boardOrdered)`

---

[1]Modeling and simulating the *generation* of the full quartet of hypotheses is outside our scope, and we thus commence our analysis in earnest essentially at the point when this quartet is being entertained.

The overall proof, which makes use of LISA to carry out analogical reasoning to rule out the hypothesis that weight is the causally responsible element in the test, is shown in Figure 1.4, which we encourage the reader to take a few minutes to assimilate.
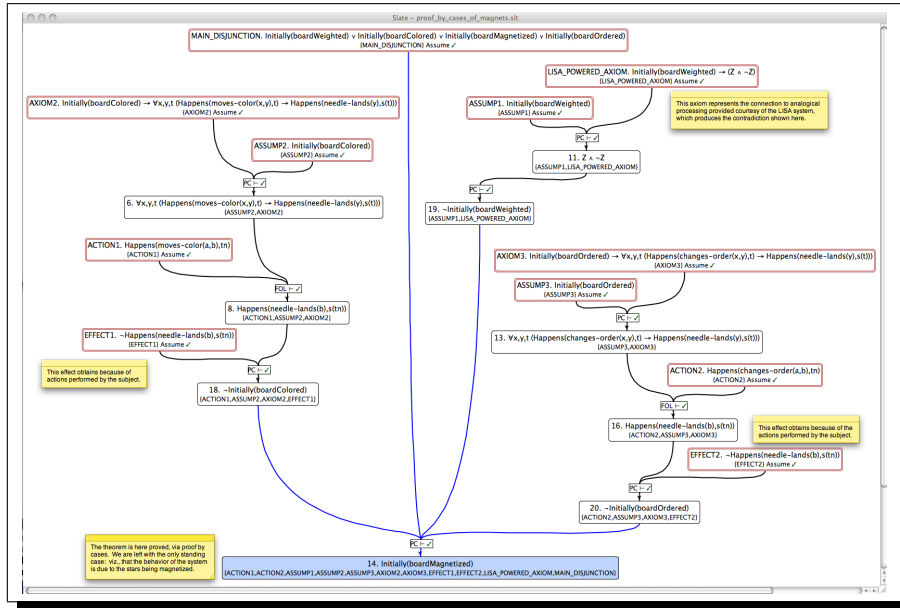

Fig. 1.4.   The Top-Level Reasoning Strategy

We can model Gou's reasoning process, by first assuming that he already understands that there is some force or property P_{stop} that causes the bar to stop. We can model this by invoking a predicate *More_P*(x,y), which is true iff a pair of boxes x is more likely to stop the rotating bar than another pair of boxes y. Gou does know that some boxes are heavier than others, which can be represented by predicates of the form *Heavier*(x,y). We will assume that Gou has some knowledge of the transitivity of weight. Finally, the causal relationship suggested to Gou by the experimenter — that weight is causally responsible for the bar's stopping — is represented using the special-group proposition *Causes*(*Heavier*(x,y),*More_P*(x,y)).[m]

In the first stage of this simulation, the set **D**, consisting of both propositional knowledge held by Gou and semantic knowledge about the objects in **D**, is represented in Slate's memory. Semantic knowledge is represented using a special predicate *Semantic_Prop*, which simply connects an object to its relevant semantic unit. For example, *Semantic_Prop*(*bill*,*tall*) and *Semantic_Prop*(*jim*,*tall*) connect the objects *bill* and *jim* to the semantic unit *tall*.

D is then subjected to reasoning by analogical inference. To do this, **D** must first be divided into two subsets: **D**$_{source}$ and **D**$_{target}$. Note that these two subsets need not be

---

[m]Proposition groupings are treated differently in LISA than regular propositions (Hummel and Holyoak, 2003a).
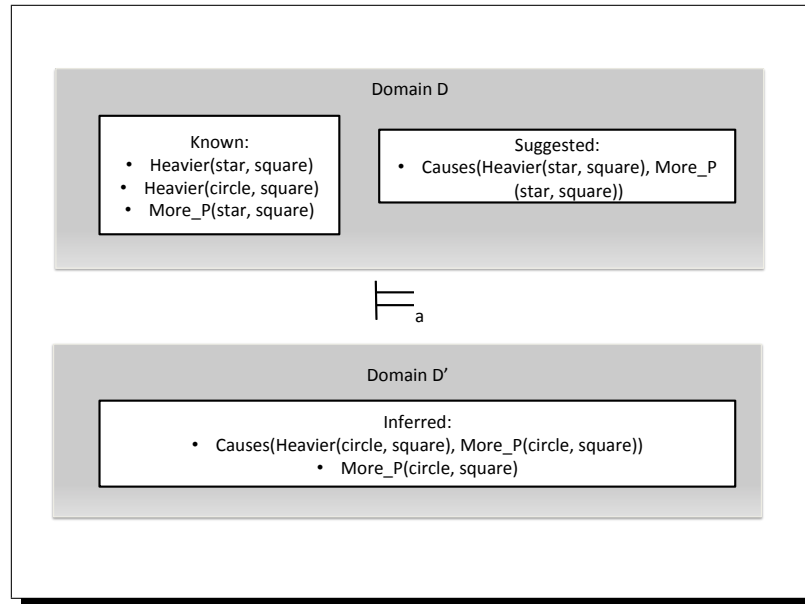
Fig. 1.5.   Propositional knowledge used in LISA for the Gou example. Domain D' is inferred from D using LISA's analogical inferencing capability. Propositions representing semantic connections are not pictured here.

mutually exclusive or collectively exhaustive — they only need to each be subsets of **D**. Choosing which propositions to include in **D**$_{\text{source}}$ and **D**$_{\text{target}}$ may be an iterative process, the details of which we do not provide at this time. For now, we can assume that in a relatively simple problem such as this, a useful division such as the one we will describe shortly will occur.

   D$_{\text{source}}$ and **D**$_{\text{target}}$ are then sent to LISA, where they are each used to create an analog. Analogical inference then produces the structure seen in Figure 1.6. Note that the semantic connections from the inferred predicate are mapped to the relevant semantic values as a result of the analogical inference process (Hummel and Holyoak, 2003a). The inferred predicates and semantic connections are then collected as the set **D**$'$ (Figure 1.5), which is returned to Slate, where it is then subjected to further deductive reasoning. This reasoning over **D** $\cup$ **D**$'$ may ideally derive one of two things: a testable hypothesis, which a reasoner would then empirically verify or refute; or a contradiction. A failure to derive either can result in either a repeat of the analogical process with different subsets chosen for **D**$_{\text{source}}$ and **D**$_{\text{target}}$, or a general failure condition. In the present example, Figure 1.5 shows that **D**$'$ contains the proposition *More_P*(*circle*, *square*). Gou's experiment, however, shows $\neg$*More_P*(*circle*, *square*). This leads to a contradiction exploited in Slate, and hence the possibility that weight is causally responsible for whatever force is stopping the metal bar is rejected.

   One might ask at this point whether analogical reasoning is necessary to carry out this process. The answer is clearly "No." But the question is wrong-headed. After all,
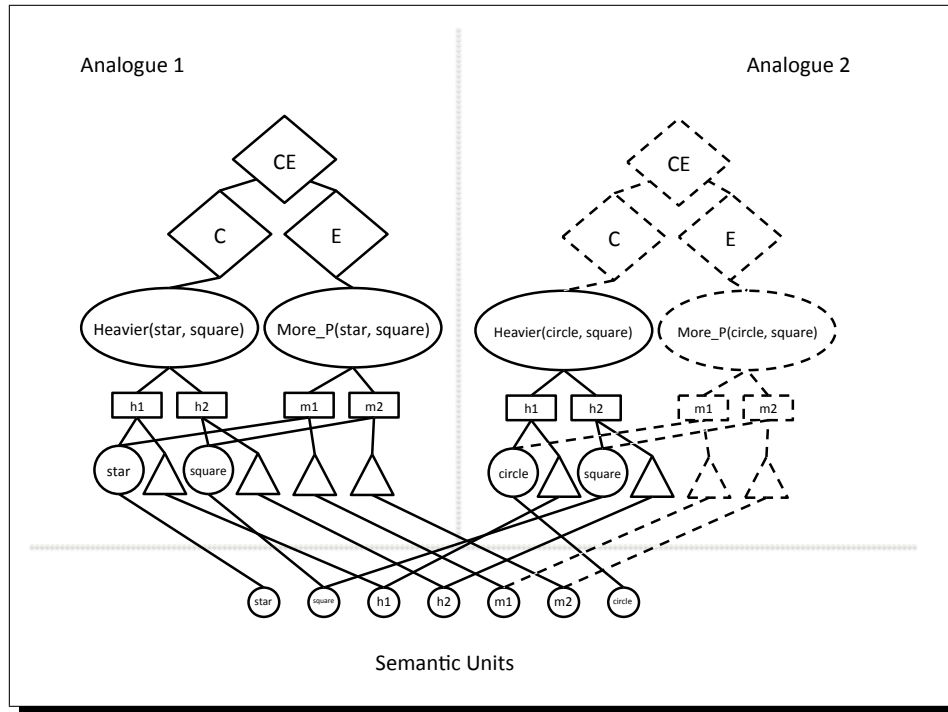
Fig. 1.6.   Structured knowledge used in LISA for the Gou example. The units in dotted lines are generated using analogical inference.

every elementary logic textbook covering not just deduction, but also induction, abduction, and analogical reasoning,[n] presents the alert reader with formal facts that allow her to see that, *in principle*, deduction can be used to arrive at whatever conclusion is produced in heterogeneous fashion — if additional premises are added. (This is actually an easily proved theorem, given that the commonality to all forms of reasoning is that the content reasoned over is relational and declarative.) Accordingly, while we are not arguing that the precise procedure we have chronicled exactly models the thought processes all reasoners go through, it seems that analogical reasoning produces a plausible explanation.

In fact, consider the following. After Gou is asked to investigate what force or property P_{stop} was responsible for stopping the bars, he might then perform some experiments on the assumption that P_{stop} is transitive. For example, he might think that if the *star* boxes are heavier than the *square* boxes, and a set of boxes b existed that were heavier than the *star* boxes, then the b boxes should be more likely to stop the bar than the *star* boxes. However, it doesn't follow from deductive reasoning alone that P_{stop} is transitive. After all, it may be the case that stacking two boxes on top of each other would cancel out their relative contributions to P_{stop}, or that the boxes together would have no stronger effect on stopping the rotating bar than they would have alone. He may have suspected that P_{stop}

---

[n]E.g., (Copi, Cohen and MacMahon, 2011)

behaved in a similar way to forces familiar to him; forces like gravity or magnetism. If so, analogical ability neatly explains how he would have mapped the properties of magnetism — for example, its ability to pull on some objects more than others — on to P_{stop}. This process suggests to us that he previously understood the transitivity of weight, analogically inferred that P_{stop} was similarly transitive, and formed an easily testable hypothesis.

Note that in the previous paragraph we say "suggests *to us*." Although we have stated that complete psychological plausibility is not a primary goal of our simulation (it focuses more on possible ways in which analogical and deductive reasoning can interact), we should note here that Piaget himself was suspicious of the existence of analogical reasoning in children who have not yet reached Stage III. A series of experiments he carried out with Montangero and Billeter seemed to suggest that young children are not capable of consistently performing stable analogical reasoning, and that they instead tend to reason using surface similarity in analogical problems (Piaget, Montangero and Billeter, 2001). Goswami and Brown (1990) recreated a similar experiment with items and relations more likely to be familiar to small children; she demonstrated that they indeed had more analogical ability than Piaget suspected. Further experimentation by other researchers showed analogical ability in pre-linguistic children as well (Goswami, 2001). In any case, these results point to the complexity and ever-changing nature of the ways in which analogical and deductive reasoning mix.

Recent work by Christie & Gentner (2010) suggests that at least in the case of young children, analogical reasoning is not likely to be used in generating hypotheses — unless the relevant stimuli are presented simultaneously, in a manner that invites side-by-side comparison and higher-level relational abstraction. Instead, the magnet experiment's format would encourage hypotheses based on surface similarity, which presumably would lack the depth to provide a satisfactory set of testable hypotheses. (We see this with most of Piaget's younger subjects: after a while, they simply *give up* (Inhelder and Piaget, 1958a).) The example we presented here does not have the child initially using analogy to generate a theory about weight. Instead, the mapping from weight is triggered by a suggestion from the experimenter himself. Analogico-deductive reasoning is then used to elaborate on this suggestion, and ultimately refute its validity.

1.7. Next Steps

Alert readers will have observed that under the assumption that Piaget can draw from the ingredients in Figure 1.1 to construct a PAGI challenge in PMR, rather more than the magnet challenge is possible. Our next foray into PAGI via analogico-deduction, now underway, involves another of Piaget's challenges: the balance problem. In this challenge, subjects are presented with a scale like that shown in Figure 1.7. To crack this puzzle, the subject must reason to the general rule *r* that balance is achieved under weight differentials when distance from the vertical post for hanging weights is proportional to the amount of weight in question. Victorious problem-solvers here, like MacGyver, manage in relatively short order to figure out that weights of different sizes can nonetheless by hung so that balance
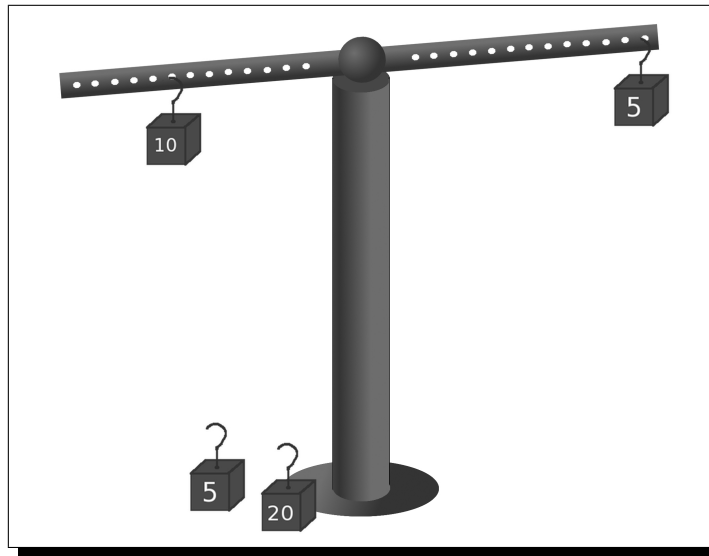
Fig. 1.7. Rendering of the Scale in Piaget's Famous Balance Challenge

is achieved, as long as *r* is apprehended, and followed in the physical manipulation. In our work-in-progress, *r* is represented by a formula in FOL, and is arrived at via — no surprise here — analogico-deduction. Use of such reasoning is supported by what is seen in the subjects; for example in the case of a child who finds the secret to the balance puzzle in the game of marbles, which, if you look carefully, you will indeed see listed in Figure 1.1 as raw material for PMR.

References

Anderson, J. and Lebiere, C. (2003). The Newell test for a theory of cognition, *Behavioral and Brain Sciences* **26**: 587–640.

Anderson, J. R. (1993). *Rules of Mind*, Lawrence Erlbaum, Hillsdale, NJ.

Anderson, J. R. and Lebiere, C. (1998). *The Atomic Components of Thought*, Lawrence Erlbaum, Mahwah, NJ.

Bringsjord, S. (2008). Declarative/Logic-Based Cognitive Modeling, *in* R. Sun (ed.), *The Handbook of Computational Psychology*, Cambridge University Press, Cambridge, UK, pp. 127–169.

Bringsjord, S. and Schimanski, B. (2003). What is artificial intelligence? Psychometric AI as an answer, *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI–03)*, Morgan Kaufmann, San Francisco, CA, pp. 887–893.

Bringsjord, S., Bringsjord, E. and Noel, R. (1998). In Defense of Logical Minds, *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Mahwah, NJ, pp. 173–178.

Cassimatis, N. (2002). *Polyscheme: A Cognitive Architecture for Integrating Multiple Representation and Inference Schemes*, PhD thesis, Massachusetts Institute of Technology (MIT).

Cassimatis, N., Trafton, J., Schultz, A. and Bugajska, M. (2004). Integrating cognition, perception and action through mental simulation in robots, *in* C. Schlenoff and M. Uschold (eds), *Pro-*

ceedings of the 2004 AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems, AAAI, Menlo Park, pp. 1–8.

Christie, S. and Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment, *Journal of Cognition and Development* **11**(3): 356–373.

Copi, I., Cohen, C. and MacMahon, K. (2011). *Introduction to Logic*, Prentice-Hall, Upper Saddle River, NJ. This is the 14th (!) edition of the book.

Descartes, R. (1911). *The Philosophical Works of Descartes, Volume 1. Translated by Elizabeth S. Haldane and G.R.T. Ross*, Cambridge University Press, Cambridge, UK.

Evans, G. (1968). A program for the solution of a class of geometric-analogy intelligence-test questions, *in* M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, MA, pp. 271–353.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, W., Nyberg, E., Prager, J., Schlaefer, N. and Welty, C. (2010). Building Watson: An Overview of the DeepQA Project, *AI Magazine* pp. 59–79.

Fischler, M. and Firschein, O. (1987). *Intelligence: The Eye, the Brain, and the Computer*, Addison-Wesley, Reading, MA.

Goswami, U. (2001). Analogical reasoning in children, *in* D. Gentner, K. J. Holyoak and B. N. Kokinov (eds), *The Analogical Mind: Perspectives from Cognitive Science*, The MIT Press.

Goswami, U. and Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations, *Cognition* **35**(1): 69–95.

Holyoak, K. J., Gentner, D. and Kokinov, B. N. (2001). Introduction: The place of analogy in cognition, *in* D. Gentner, K. J. Holyoak and B. N. Kokinov (eds), *The Analogical Mind: Perspectives from Cognitive Science*, The MIT Press, chapter 1.

Hummel, J. E. and Holyoak, K. J. (2003a). Relational reasoning in a neurally-plausible cognitive architecture: An overview of the lisa project, *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* **10**: 58–75.

Hummel, J. E. and Holyoak, K. J. (2003b). A symbolic-connectionist theory of relational inference and generalization, *Psychological Review* **110**: 220–264.

Hummel, J. E. and Landy, D. H. (2009). From analogy to explanation: Relaxing the 1:1 mapping constraint...very carefully, *in* B. Kokinov, K. J. Holyoak and D. Gentner (eds), *New Frontiers in Analogy Research: Proceedings of the Second International Conference on Analogy*, Sofia, Bulgaria.

Inhelder, B. and Piaget, J. (1958a). *The Growth of Logical Thinking : From Childhood to Adolescence*, Basic Books, Inc.

Inhelder, B. and Piaget, J. (1958b). *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, NY.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium, *in* W. Chase (ed.), *Visual Information Processing*, New York: Academic Press, pp. 283–308.

Penrose, R. (1994). *Shadows of the Mind*, Oxford, Oxford, UK.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood, *Human Development* **15**: 1–12.

Piaget, J., Montangero, J. and Billeter, J. (2001). The formation of analogies, *in* R. Campbell (ed.), *Studies in Reflecting Abstraction*, Psychology Press.

Pollock, J. (1989). *How to Build a Person: A Prolegomenon*, MIT Press, Cambridge, MA.

Popper, K. (1959). *The Logic of Scientific Discovery*, Hutchinson, London, UK.

Rinella, K., Bringsjord, S. and Yang, Y. (2001). Efficacious Logic Instruction: People are not Irremediably Poor Deductive Reasoners, *in* J. D. Moore and K. Stenning (eds), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 851–856.

Rosenbloom, P., Laird, J. and Newell, A. (eds) (1993). *The Soar Papers: Research on Integrated Intelligence*, MIT Press, Cambridge, MA.

Russell, S. and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ. Third edition.

Sternberg, R. (1988). *The Triarchic Mind: A New Theory of Human Intelligence*, Viking, New York, NY.

Sun, R. (2001). *Duality of the Mind*, Lawrence Erlbaum Associates, Mahwah, NJ.

Wang, P. (2006). The logic of intelligence, *in* B. Goertzel and C. Pennachin (eds), *Artificial General Intelligence*, Springer, New York, NY, pp. 31–62.

Wason, P. (1966). Reasoning, *New Horizons in Psychology*, Penguin, Hammondsworth, UK.

Wason, P. and Johnson-Laird, P. (1972). *Psychology of Reasoning: Structure and Content*, Harvard University Press, Cambridge, MA.