Exploring the Role of Analogico-Deductive Reasoning in the Balance-Beam Task

John Licato[1] & Selmer Bringsjord[2] & John E. Hummel[3]

Rensselaer AI and Reasoning (RAIR) Laboratory
Rensselaer Polytechnic Institute (RPI)[1,2]
Relational Reasoning Laboratory
University of Illinois at Urbana-Champaign[3]
licatj@RPI.edu[1]• Selmer@RPI.edu[2]• jehummel@illinois.edu[3]

Abstract

Piaget's classic experiments enabled, among many other things, identification of the types of reasoning children use in different developmental stages. His work identified both hypothetico-deductive and analogical reasoning as strong indicators of normal development. Although examples of each can be seen early on, consistent application and control of these types of reasoning emerges somewhere between Stages III and IV. Given their importance in the developmental process, it can only help the study of cognitive development to have a deeper understanding of how these types of reasoning are carried out. We focus in this paper on ways in which they interact (we call this ADR, for *analogico-deductive reasoning*); specifically on the interaction of generation, development, and testing of hypotheses. We present a model and show how it can explain the reasoning processes of a Stage-III subject working on Inhelder and Piaget's (1958) seminal balance-beam task.

Analogy and Analogico-Deductive Reasoning

Analogical reasoning (AR) has always played an important role in Piagetian theories. While Piaget suspected that AR was a mark of formal reasoning (Piaget, Montangero, and Billeter 2001), Goswami showed that at least some analogical ability resides in younger children (Goswami and Brown 1990). Of the Neo-Piagetians, Halford was perhaps the most passionate proponent of AR, citing it as a key feature of his relational complexity theory (Halford and McCredden 1998; Halford et al. 2002). Some go as far as to suggest that AR is the core of all cognition (Gentner 2003; Hofstadter 2001).

However, relatively recently, some researchers have started to take deeper, more detailed looks at how AR interacts with the hypothetico-deductive process, mostly through simulations using computational models. Christie and Gentner (2010) examined whether analogical reasoning was involved in the generation of new hypotheses, and their results suggested that children tend to use relational hypotheses only when the experiment design invites side-by-side comparison. In (Bringsjord and Licato forthcoming), the present authors applied analogical reasoning to the Piagetian magnet test, proposing a model to simulate what we called *analogico-deductive reasoning*.

Analogico-deductive reasoning (ADR) is the intersection of hypothetico-deduction and analogical reasoning. In particular, we focus on a subset of ADR that involves the formation of hypotheses h_i and the generation of results r_i which must be true if the h_i are true. Analogical reasoning can be responsible for either the formation of the hypotheses, or the conditionals of the form $h_i \rightarrow r_i$. Through further reasoning and experimentation in order to determine the validity of the r_i, the h_i can be either supported or rejected, ultimately by inference rules like *modus tollens* and *reductio ad absurdum*.

This type of ADR is frequently used by individuals trying to understand new or unfamiliar concepts. Holyoak et al. (2001), for example, explain that as the wave theory of sound became better understood, it became used as a basis to describe a wave theory of light. This was an analogical mapping which would have implied the existence of a medium through which light would travel (the luminiferous aether), just as sound waves travel through air. Compare this with a particle theory of light that sees photons as analogous to billiard balls, which would not require a medium to travel through, and would produce a different set of experimentally verifiable predictions.

However, the nondeterministic nature of analogical reasoning makes it difficult to use any model of ADR as a way to predict behavior. AR is often used to map knowledge from a domain with which a reasoner is familiar (the source analog) to a domain with which the reasoner is not (the target analog). But since the source analog can be almost any domain with which the reasoner is familiar, the analogical mappings drawn may be based on something as arbitrary as what thoughts happen to be active in the reasoner's mind at the time. Unless the source domain is provided, it is difficult if not impossible to accurately predict beforehand which analogs a reasoner will retrieve. What we *do* know, however, is that analog retrieval and selection tends to be related to surface similarity; for more information,

see Gentner et al. (1993), Holyoak (1987), Ross (1989), and Gentner and Forbus (2011).

Perhaps because analogies are so hard to forecast, most models of Piagetian tasks do not focus on the analogies that subjects use to understand them. This has certainly been the case throughout the history of the Balance-beam task, the central task of this paper. Yet while such a focus has limited use as a predictive model, it deserves a closer look. Attempting to model ADR by closely analyzing the analogies given by individual subjects may shed some light on some processes that other models tend to neglect, but which are of central importance to human thought: insight, hypothesis generation, experiment creation, and hypothesis testing, to name a few. In this paper, we focus on the ADR used by a subject described by Inhelder and Piaget (1958); the subject used an analogically-inspired insight to produce solutions to the balance-beam task (BBT).

We begin by describing the BBT and its history, particularly the history of computational simulations of the BBT. We then introduce the subject whose analogy we will study, and describe our approach to modeling it. A discussion of the model and lessons learned follows.

The Balance-Beam Task

The balance-beam task (BBT), first described by Inhelder and Piaget (1958), is one of the most well-known and studied of the classical Piagetian tasks. In one variation, the subject is presented with a scale containing a number of holes on each side (Figure 1). On these holes the subject can hang a number of weights. In one common variation of the BBT, the subject is shown a configuration with weights already hung on each side, and is asked to predict whether the scale will balance, tip to the right, or tip to the left.

The torque produced on each side is equal to the sum of each weight multiplied by that weight's distance from the fulcrum (the 'torque rule'). Knowing this allows for an easy way to make an accurate prediction. However, the ways in which younger subjects attempt to muddle through is very instructive, and a wealth of literature has been produced from reflection on this seemingly simple problem. In order to understand the current state of computational modeling of the BBT, we will briefly review some of the relevant literature here. For more thorough summaries of the current state of research on the BBT, please see van Rijn et al. (2003) and van der Maas and Raijmakers (2009).

Models of the BBT

One of the earliest and perhaps most-cited post-Piagetian analyses of the BBT comes from Siegler (1976; 1981). Siegler introduces a decision-tree model based on the idea that children will follow one of four *rules* when attempting to solve the BBT. Each of these rules could be represented by a single decision tree, and each rule is more complex than the last. For example, Rule I consisted of a single decision: Is the weight on the two sides the same? If so, the subject predicts the scale will balance; if not, the subject predicts the side with the greater weight will go down. Rule IV, on the other hand, has decisions contingent on more complicated cases (e.g., *Is the side with the greater weight the same side as that with the greater distance?*). Subjects were then classified based on which of these rules they followed. It is worth noting that the increasing complexity of the rules seems to fit nicely with the idea of working memory as described by Halford (1998; 2002): the older subjects get, the more variables they can simultaneously reason over (and the more complex the rules they can follow).

Although Siegler found that many subjects could indeed be neatly classified as following one of these rules, the amount of unclassified subjects was not insignificant. This problem was addressed by another rule-based model, that of Wilkening and Anderson (1982). They determined that Siegler's decision-tree methodology lacked the ability to represent algebraic integration rules. For example, the decisions of a subject using an "adding-type rule" which would add the weights to their distances from the fulcrum on each side, and then determine which side was greater, would not be modeled correctly using binary decision trees. Some of these criticisms were supported by Normandeau et al. (1989), who found that Siegler's Rule III underestimated the ability of many adolescents. Their behaviors were better explained with a rule-based model that incorporated both Wilkening and Anderson's adding-type rule, and a "QP" (qualitative proportionality) rule. Users of the QP rule would predict that a heavy weight at a small distance on one side balances with a light weight at a great distance on the other, demonstrating just the type of integrative reasoning (since rather than examining weight or distance separately, it was done in parallel) which was missing from Siegler's original rules.

As much as these rule-based models allowed for some explanation of the behaviors of subjects taking the BBT, they did a poor job at explaining where the rules came from. Furthermore, rule-based models seemed to have trouble explaining the TD (torque-distance) effect (Ferretti and Butterfield 1986), which seemed to indicate an intuitive understanding of torque in subjects who did not know the equations to properly describe it. Subjects have an easier time predicting the outcome of the scale when the torque difference is larger, even if there are fewer weights on the side with the higher torque. It is important to note that the torque difference must be relatively large, otherwise the TD effect is not observed in empirical data (Jansen and van der Maas 1997).

Connectionists use the inability of rule-based models to explain the TD effect as one of the main arguments for the superiority of their models, the most successful of which were perhaps those of McClelland (McClelland 1989; McClelland 1995) and the cascade-correlation models of Shultz (Shultz, Mareschal, and Schmidt 1994;

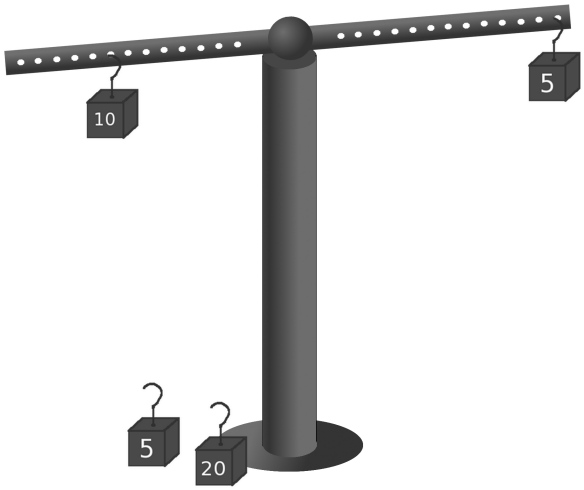

Figure 1: A Balance Scale Like That Used in Inhelder and Piaget (1958)

Shultz 2003). These models were able to duplicate many of the behaviors described in the empirical data. Furthermore, it seemed that these models provided a lower-level explanation of where the rule-following behavior actually *came from*, something which was sorely lacking in the rule-based models. There were of course some limitations of the connectionist models; among these is the fact that they are unable to satisfactorily demonstrate the highest level of performance on the BBT: the discovery and use of the torque law (van der Maas and Raijmakers 2009). Furthermore, the connectionist models perform poorly when subjected to latent class analysis (LCA); for a technical discussion see van der Maas and Raijmakers (2009), who discuss the limitations of connectionist models and discuss their own hybrid approach, which attempts to combine symbolic and sub-symbolic representation.

To close out the discussion on connectionist models we note that they suffer from another major weakness. Consider that under classical reasoning, subjects can give high-level explanations and arguments for how they believe things work. In fact, making use of this sort of justification is a major component of Piaget's *méthode clinique*. With the abstract distributed representation employed by most purely connectionist models, it is not clear how high-level reasoning, beliefs, and arguments fit in.

Though it seems that much ground has already been covered in the existing literature, at least one approach to computational models of the BBT is conspicuously absent—that of modeling the reasoning which leads to discovery of the law. This is no doubt due to the scope of the aforementioned computational models, which attempt only to duplicate the behavior of subjects who are limited to predicting the outcome of a fixed instance of the BBT. However, in the more complex version of the BBT, which allows subjects to freely place and remove weights (we will refer to this as the "non-predictive BBT," or NBBT, as opposed to the restricted version used by Siegler and others, which we will refer to as the RBBT), the subjects quickly act in ways that go beyond what can be described by simple behavioral computational models. For example, if a child following one of Siegler's rules observes that an instance of the NBBT behaves in a way contrary to what he predicted, he might then attempt to discover a new rule **R** to follow. He could then design experiments that would test the behavior of the scale, and either support or refute **R**.

But where does **R** *come from*? This is hardly an easy question to answer from the the perspective of rule-based models. The connectionist models at least seem to offer a partial answer: most connectionist models can be seen as systems which take in a certain set of inputs, and produce outputs based on the current values of the system's parameters (which are usually the weights of the connections). The design of the system thus defines a set (which may be infinite in size) of possible hypotheses the model will follow to determine the output(s), which are adjusted over time as the system learns and the parameters are adjusted.[1] Still, how such processes would produce the kinds of hypotheses that require complex reasoning (such as the torque rule) is yet to be seen.

Analogico-deductive reasoning may be able to address this trade-off, by showing how new rules can be generated from analogical inference, and supported or refuted using a combination of analogical and deductive reasoning. We demonstrate this by applying our model to a case described in Inhelder and Piaget (1958).

[1] It is a bit more complicated with the cascade-correlation models (Shultz, Mareschal, and Schmidt 1994; Shultz 2003), since their learning process involves qualitative additions to the neural network itself.

The Case of Rog

Given the vast collection of empirical data and models to predict how a child is likely to reason through the RBBT, it is interesting that there is one subject described in the original text by Inhelder and Piaget which does not fit into the predictions made by any of the previously mentioned models. The subject, a child in Piaget's Stage-IIIA, and referred to as 'Rog,' provides an explanation for his reasoning that is clearly analogical, although its modeling is not as simple as it first seems. The entirety of the interaction as provided by Inhelder and Piaget is as follows (**E** denotes the experimenter, and text in italics is spoken by neither the experimenter nor Rog):

For a weight [weighing 16-units] placed at the very tip of one arm [28 holes], he puts [a 3-unit weight and a 5-unit weight] in the middle of the other arm, measures the distances, and says:
Rog: "That makes 14 holes. It's half the length. If the weight [8-units] is halved, [the 16-unit weight] duplicates."
E: "How do you know that you have to bring the weight toward the center?"
Rog: "The idea just came to me, I wanted to try. If I bring it in half way, the value of the weight is cut in half. I know, but I can't explain it. I haven't learned."
E: "Do you know other similar situations?"
Rog: "In the game of marbles, if five play against four, the last one of the four has the right to an extra marble."
He also discovers that for two distances of 1 and 1/4 you have to use weights 1 and 4; that for two distances of 1 and 1/3 you need weights 1 and 3, etc.
Rog: *(later)* "You put the heaviest weight on the portion that stands for the lightest weight [*which corresponds to the lightest weight*], going from the center." (Inhelder and Piaget 1958, p.173-174)

Inhelder and Piaget took this to be a clear indicator of the presence of proportional reasoning, and they were likely correct. They suggest the existence of an "anticipatory schema" that is "taken from notions of reciprocity or of compensation" (Inhelder and Piaget 1958), where those notions were themselves taken from Rog's experiences with the game of marbles. Rog is, in effect, using analogical inference to generate a theory about how the balance beam works: It is as if each side of the scale corresponds to a team of marble-players. Like the previously mentioned analogies regarding the nature of light, Rog's theory comes with hypotheses that could then be tested, as he does when he places eight units of weight on the scale.

The exact procedure used with Rog is not clear from the text; for example, we do not know how much experimentation Rog was allowed to do before the above transcription begins, nor do we know if Rog was told that he was not allowed to move the 16-unit weight (or if he decided to place that first weight on his own). We do know that Inhelder and Piaget allowed some subjects "to hang the weights simultaneously on both sides of the balance," or to allow "successive and alternate suspensions of the weights" (Inhelder and Piaget 1958, p.173). We will assume here that Rog was presented with the subproblem defined as follows: Given a set **W** of weights, and a scale set up so that one 16-unit weight (which cannot be moved) is placed on the 28th hole of one arm, what weights must be placed at which distance on the other arm in order to balance the scale?

Some of the obvious answers must be ruled out. We assume that there are no weights greater than or equal to 16-units in **W**, and no weights in **W** can be combined to be equal to or greater than 16-units. Otherwise, Rog might have easily just placed the proper combination of weights on the 28th hole, balancing the scale. Again, it is not clear that these were the conditions in Rog's case, but since our purposes here are to use modeling in an exploratory manner, we clarify them in the interest of simplicity.

Modeling the Analogy

We can now turn to our approach to modeling the type of reasoning that allowed Rog to reach his conclusion. We start by noting that even if Rog's reasoning process was at least in part analogical, it is worthwhile to briefly consider that it may not have been done in the way we suppose. Rog's knowledge of the game of marbles may not have been the source in the analogical mapping, but rather a target analog that was generated just to provide an after-the-fact justification of his actions to the experimenter. There are at least two other possibilities which seem plausible:

- **P1** The analogical mapping was made in the other direction: The game of marbles was the target, and the BBT was the source. Rog's understanding of how to balance the BBT came from some other process (which may or may not have been analogico-deductive).
- **P2** There are separate, but similar, analogical mappings from a more general "reciprocity or compensation" schema to both the game of marbles and the BBT.

It is entirely plausible that either of these possibilities were actualized in Rog's case. However, they do not provide any answers to the question of the origin of the source schema, and are therefore not discussed in this paper. They are only mentioned here in order to clarify the direction in which we are exploring, and may be useful topics for future work. We will therefore assume for now that the process which leads to Rog's theory is something like an analogical mapping drawn from his knowledge of the game of marbles.

The LISA System

The system we use for analogical reasoning is LISA (*Learning and Inference with Schemas and Analogies*), a neurally-plausible model of analogical reason-

ing which uses 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 2003a;2003b) and (Hummel & Landy 2009).

LISA is a "structured connectionist" architecture, which allows for both explicit, localist representation of propositional knowledge, and distributed semantic knowledge. Propositional knowledge, the arguments of which can be either token objects or other propositions,[2] is localist (meaning a single node is created for each instance of a proposition). All 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. Higher-order groupings of propositions are also allowed, such as cause-effect groups.

When visually presenting data in LISA, we maintain consistency with (Hummel and Holyoak 2003a), as in figure 2. We divide the image up into three sections: the schemas of analog 1 and analog 2, and the semantic units shared by both analogs. Semantic units are pictured as circles at the bottom (although note that in Figure 2 the individual semantic units are not shown); object units are also in circles but are parts of the analogs. The sub-propositional units are rectangles, and the predicate units which connect those sub-propositional units to semantic units are shown as triangles. Propositional units are shown as large ovals, and the group nodes organizing those propositions are shown as diamonds on the top of the screen. Additionally, when necessary, we draw units using dotted lines if they were created as a result of analogical inference.

LISA makes the theoretical assumption that the same underlying processes are responsible for two necessary functions of analogical ability: analog retrieval, which retrieves relevant source analogs from long-term memory; and analogical mapping, which aligns knowledge units in the source analog to units in the target. This is a key difference from models such as the Companions cognitive architecture (Forbus, Klenk, and Hinrichs 2009; Gentner and Forbus 2011), which uses SME (Gentner 1983) for analogical mapping and MAC/FAC (Forbus, Gentner, and Law 1995) for analogical retrieval. LISA uses its detailed semantic representations to bring relevant analogs into working memory (Hummel and Holyoak 1997).

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 (alternately: *inferred*) if necessary. The recruiting of new propositional units in the source analog is the ideal result of analogical inference, and will be the primary functionality of LISA used in the present paper.

Because the mapping process is initiated by the firing of propositional units, the way they fire has an effect on what mappings are formed in two ways. First, the order can sometimes cause erroneous mappings to occur, and when they occur early enough in the mapping process they can heavily influence the final results (Hummel and Holyoak 2003a; Hummel and Holyoak 2003b). Secondly, the number of propositional units fired simultaneously (which is called the *phase set*[3]) is limited by a number of factors; one of the most important of these is driver inhibition, or the ability of sub-propositional units to inhibit each other when firing simultaneously.

This enables LISA to simulate some normal effects of cognitive development and working memory. LISA also contains an option for unlimited working memory, which enables a phase set of virtually unlimited size. These parameters are useful when examining robustness, but we will be using the default values unless otherwise noted.

Simulating Rog's Analogico-Deductive Reasoning

Inspired by the description of Rog's analogy, we attempt to create a simple model that can carry out the ADR required to reach the same or similar conclusions. We are not here particularly interested in how a source analog is chosen; rather, we focus on the ADR which follows the selection of a suitable analog. The goal of such research is to extract lessons from that attempt, and apply them toward a more general model of analogico-deductive reasoning.

As mentioned earlier, we will assume that the process that leads to Rog's theory is something like an analogical mapping drawn from his knowledge of the game of marbles. What might such a mapping look like, using the data structures available in LISA? It seems safe to start with the objects that Rog directly mentions: the number of players on each team, the last player on the team of less players, and the extra marble given to this player.

Listing the predicates that Rog seems to use is slightly more complicated. Clearly, he has some conception of balance, in the abstract sense that can describe both a state of equilibrium on the scale and a sort of fairness that occurs when both teams in the game of

[2] E.g., *Knows(tom, Loves(sally, jim))*.

[3] The phase set can also be defined, as in Hummel and Holyoak (1997), as the "number of distinct role bindings (SPs) that can be considered together prior to updating the mapping connections."

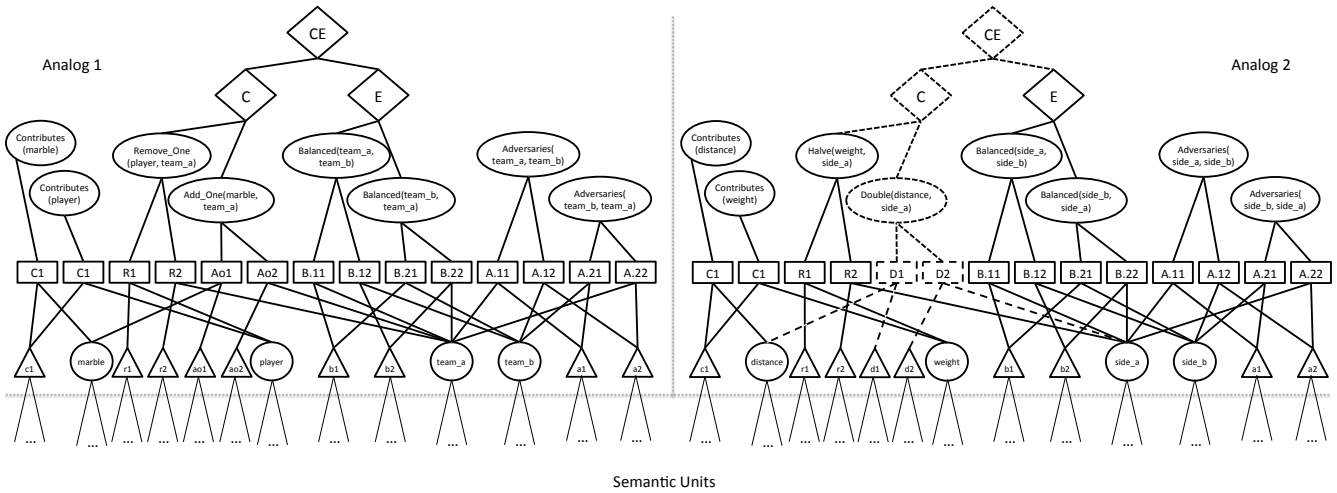

Figure 2: The Analogs Used in Our Model. For simplicity, not all individual semantic units, propositions, or other units are pictured here (see the appendix for full details on which units were used). Dotted lines represent units that were recruited as a result of analogical inference.

marbles have an equal opportunity to win. Complementary to this understanding of balance is a likely notion of an adversarial relationship between the two teams in the game and the two sides of the balance beam. He understands the concept of halving and doubling numbers, and that they are, to use Piagetian terminology, reversible.[4] He also understands that incrementing the lesser of two items corrects a difference of one, an operation that is reversed by removing one.

These objects and predicates suggest a possible set of schemas we can use to create an analogical inference. We can imagine a possible mapping by first describing Rog's knowledge about the game of marbles as the source analog (analog 1) and his BBT-related knowledge as the target (analog 2). Some mappings seem intuitive, and a possible list of these mappings are seen in table 1.

This mapping, however, introduces a problem that must be addressed before we continue. Whereas most of the mappings should be relatively uncontroversial, the final two pairs are not so straightforward. We know that he understands the operations of *adding one, removing one, doubling,* and *halving*; therefore, either both analogs have access to these operations' relevant predicates, or we must explain why they do not. Assuming Rog starts with the idea of halving the weight on one side (which, recall, is the sub-problem we have decided to tackle in this paper), he must solve something like the analogy:

$$\textit{Removing one} : \textit{Adding one} :: \textit{Halving} : \mathbf{O?}$$

where the operation **O?** is some operation available to him in analog 2. Solving this sort of problem requires a relation **P** between *Halving* and *Doubling* which is similar to, or the same as, the relation between *Adding one* and *Subtracting one*, which we can at least partially model through specially placed semantic connections. But if all four operations are available to both analogs, and **P** cannot be explicitly represented (as it is a second-order relation), how can we ensure that the predicate *Doubling*, rather than *Adding one*, is chosen? After all, whatever similarities *Doubling* and *Adding one* have (e.g., they both increase numerical value), nothing is more similar to a predicate than itself.

At least two solutions seem plausible. In the first, the filtering process that decides what predicates and propositional knowledge to include in the analogs *before* analogical mapping and inference begins, must include the predicate *Doubling* and exclude *Adding one*. This might be as a result of similarity-based retrieval, which has been shown to be the primary process used in analog retrieval (Gentner, Rattermann, and Forbus 1993; Holyoak and Koh 1987; Ross 1989; Hummel and Holyoak 1997; Gentner and Forbus 2011). Rog starts knowing that the *Halving* operation will be present in analog 2, and uses similarity-based retrieval to also include the related *Doubling* operation. Since there is no clear addition or subtraction present in the BBT, the *Adding one* and *Removing one* predicates are not retrieved.

The second possibility is that the unintuitive mapping of *Adding one* to *Doubling* is a simple accident—perhaps a consequence of Rog's still-developing analogical reasoning ability. If so, this suggests that the full version of this model, if tweaked to have the analogical reasoning ability of an adult, should be less likely to form the same mappings as Rog. However, since this likely has to do with the area of the model dealing

[4]However, reversibility, being a second-order relationship, cannot be currently represented in LISA.

Table 1: *Possible Mappings for Rog's Analogy*

Analog 1	Analog 2
Teams of players (a and b)	Sides of the scale (a and b)
Concept of balance (fairness)	Concept of balance (physical)
Adversarial relationship between teams	Opposite relationship between scale sides
Concept of contribution (number of marbles contributes to fairness)	Concept of contribution (both weight and distance contribute to balance)
Adding one	Doubling
Removing one	Halving

with retrieval (which is still in development), we do not address this possibility here.

We can now describe a model that can simulate the setup of this inference, the carrying out of this analogical inference, and the subsequent deductive reasoning steps leading to Rog's observed behavior. Formally, this model can be seen as finding a proof of the following:

Given **K**, if **C**, then **E**

where:

- **K** is knowledge about the world. **K** involves knowledge about source analogs the subject may have acquired elsewhere (such as the rules of the game of marbles), knowledge about the BBT, and and any other relevant knowledge. **K** is also divided into analogs or domains, one of which is L_{target}, the subset consisting of knowledge directly related to the BBT.
- **B** is knowledge describing the BBT. This may be a subset of **K**.
- **E** (for "effect") is the goal state, consisting of propositions describing the state of the successfully completed BBT. Presumably this is provided by the experimenter.
- **C** is the goal, a set of propositions that are the result of a successful application of our algorithm. **C** and **E** form a cause-effect grouping, which means that if all of the propositions in **C** are satisfied, then that will cause the propositions in **E** to become satisfied as well.

We can assume that we start with **K** and **G**. The problem then becomes finding **C** (What needs to be true to reach the goal state **E**?). Before explaining the details of the algorithm, we first provide a high-level description as follows:

1. We are given **K** and **E**. Collect **B** and L_{target}, which is a subset of **K** consisting of knowledge related to the BBT, and an *incomplete cause-effect grouping* (ICEG) constructed from **E** (described in detail shortly).
2. Determine the source analog (more on this shortly).
3. Construct L_{source}, which is a subset of **K** involving knowledge related to the source analog.
4. Perform analogical inference using LISA, which is a mapping from L_{source} to L_{target} that produces new propositional information. This process, if successful, returns a hypothesis **H** in the form of a cause-effect grouping containing the conjunction of all propositions in **E** as its effect, and the conjunction of all propositions in **C** as its cause; e.g.: If side a has half the weight of side b, and side a has twice the distance of side b, then the scale should balance.
5. Perform a consistency check, searching for contradictions in $\mathbf{K} \cup \mathbf{H}$. If one is found, then the algorithm has failed.

Although the purposes of each of these steps should be relatively straightforward, the implementation details require some elaboration. Certain assumptions are made which simplify the implementation of the above steps. **K** is already divided into analogs, each of which cover a restricted domain of knowledge. One of these domains is the BBT itself, and the corresponding analog consists of knowledge Rog presumably obtains through his interactions with the balance beam up until the point at which the ADR process we model here begins. Step two is also trivialized; we assume that some psychological process caused Rog to decide that his familiarity with the rules of marbles would be useful in understanding the behavior of the BBT.

Step four involves LISA performing analogical inference using self-supervised learning, which requires L_{source} and L_{target} to conform to LISA's structured connectionist representation. However, since the steps immediately preceding and following step four require a form of knowledge representation which is appropriate for deductive inference (which we will conveniently refer to here as "deductive mode," or DM, as opposed to "LISA mode," or LM), we must use a bidirectional, knowledge-preserving transformation. This is simple for both propositions and nested propositions (e.g., *Knows(John, Likes(Jack, Television)))*). However, there are several types of knowledge which deserve special consideration:

Grouped Propositions Because higher-order groupings of propositions, such as those representing causal relationships, are treated differently from nested propositions in LISA (Hummel and Landy 2009), they must be somehow represented differently from nested propositions in DM. We use the specially named predicate $Grouping_Causes(p, q)$ to represent a cause-effect grouping between propositions p and q, and can create similarly named predicates if any other groupings are necessary.

Incomplete Cause-Effect Groupings We assume that Rog's search is motivated by the question, possibly asked by the experimenter: What do you have

to do to make the scale balance? This question is represented with an *incomplete cause-effect grouping* (ICEG), which consists of a lone "Effect" grouping and a set of minimal propositions (see analog 2 in Figure 2). These propositions are the conditions of satisfaction of a cause-effect grouping that provides a successful answer to the question asked by the experimenter; alternately, what propositions must be true to cause the scale to be balanced? In this case, the relevant propositions are trivially satisfied: If the scale is balanced, then side a is balanced with side b, and side b is balanced with side a. Note here that two propositions are needed to express this relation, because given any two-argument predicate A, for objects p, q, it is not necessarily true that $A(p, q) \rightarrow A(q, p)$. A successful run involves both the completing of the cause-effect grouping which subsumes the ICEG, and the recruiting of any necessary propositions to fill the corresponding "Cause" grouping.

Semantic Knowledge LISA achieves a simultaneously localist and distributed representation by linking semantic units to instances of objects and predicate roles wherever they appear. For example, the object *Dog* might be connected to the semantic units *hairy*, *four-legged*, and *animal*, among others. A predicate such as $Loves(p, q)$ has two roles: that played by the first argument (the object which is the *lover*), and that played by the second argument (the object which is *loved*). Each of these roles is connected to a set of semantic units, and those same sets are connected to the roles more or less consistently in every analog where a proposition containing the *Loves* predicate appears. We say 'more or less' because as a result of contextual shading, a given object or role may have slightly different contexts in different contexts.

In other words, two objects or roles are separate tokens but of the same type if and only if the semantic units to which they connect are the same. In order to preserve this property, every object or role type has a unique identifying semantic unit containing the same name. Therefore, every instance of the *Dog* object is connected to a semantic unit named *Dog*, and every instance of the *Rover* object is connected to the *Rover* semantic unit. However, even though Rover is a dog, *Rover* is not connected to the *Dog* semantic unit; rather, the relationship between the two objects is represented by the proposition $Is_A(Rover, Dog)$.[5]

These unique identifying semantic units are not explicitly represented in DM, but are created for each object in **D** when switching over to LM. They are then removed when converting back to DM. The other semantic connections, however, are preserved in DM, and represented by the special predicate $Semantic_Prop$, which associates each object in **K** with another specially named semantic unit. Although the semantic units are technically represented in DM as objects, they are not treated as such and are not part of **K** for the purposes of this discussion.

[5]Note that this is not a requirement; rather it is the approach we take in this paper, which is a departure from Hummel and Holyoak (2003a; 2003b).

First and Higher-order Knowledge It is indeed likely that analogical reasoning occurs over knowledge which requires representation in a logic more expressive than the propositional calculus we have been using thus far. We have already mentioned this, but to further explicate the point, consider an example inspired by Shakespeare's *Romeo and Juliet*:

1. All Montagues hate Juliet.
2. Tybalt is a Capulet.
3. Therefore, all Capulets hate Romeo. (from 1)
4. Therefore, Tybalt hates Romeo. (from 3 and 2)

In this simple example, step three comes from step one by way of analogical inference (although the domains for both steps are the same, and it is a false analogy for which a counterexample can easily be found by anyone familiar with the play: Juliet herself is a Capulet who does not hate Romeo). Yet note that representation for the sentences in both steps requires at least first-order expressivity, in order to perform the universal elimination and subsequent modus ponens that leads to step four.

Although while in DM, statements in first and higher-order logics may be allowed, at this time LISA does not have the ability to represent knowledge beyond the propositional calculus. As a result, we simply do not attempt to convert such statements when switching to LM. We hope to re-visit this problem in future work.

Inferred Units In self-supervised learning, LISA can recruit new groups, propositions, predicates, or objects in the target analog (we will alternately say these new units were *inferred*). When converting from LM to DM, we make use only of inferred propositions and groups, as predicates and objects not referred to by any proposition have no effect on the space of deductive conclusions that can be reached. Inferred propositions can either use predicates that already exist in the target analog (in which case they use the correct role bindings), or new predicates which have been recruited. In the latter case, LISA assigns the new predicates names based on the names of the predicates in the source analog to which they map.

Inferred group propositions are treated similarly, except they are given special predicate names such as $Grouping_Causes$, as mentioned earlier.

Finally, step five subsumes the bulk of the deductive reasoning. A successful run of step four (pictured in Figure 2) will produce the hypothesis **H**:

$$Grouping_Causes(Double(distance, side_a) \wedge Halve(weight, side_a), Balanced(side_a, side_b) \wedge Balanced(side_b, side_a))$$

Or, restated, "the distance on side a being doubled and the weight on side a being halved causes the scale to

balance." The resulting process first tries to use this new proposition to deductively reach a contradiction, in which case the process is regarded to have disproven the hypothesis analogically inferred in step four. For example, Rog may have the pre-existing (and false) belief that if the propositions in **C** are true, then it causes the negation of **E** to be true. This would clearly contradict **H**.

If a contradiction is not found or a pre-set time limit expires, then **H** can be tested by satisfying the propositions in **C** and observing to see if the propositions in **E** become true as an effect.

Simulation Results

As a first step in testing the model's viability, we decided to test step four. This step involves finding an analogical mapping which produces **H**, which is arguably the most crucial step of the algorithm. Our test case used the analogs pictured in Figure 2, which are described in full in the appendix.

The initial results were straightforward—using certain patterns of proposition firings frequently produced the correct inferred items, which encouraged some further experimentation. The ability for LISA to find the right mapping requires that the right information be placed in the relevant analogs, which itself depends completely on the accuracy of the data selection/filtering process (steps one through three). It is important, then, to know precisely the extent of this dependence: Would step four find relevant answers even if L_{source} and L_{target} were slightly different?

In order to explore this, step four was repeated with certain propositions or predicates removed or added. Each case was repeated 200 times: 100 trials with normal settings (n), and 100 with unlimited working memory (uwm). For all trials, the phase set in the appendix was used.

A trial was regarded to be successful if the following criteria were all met:

1. The proposition $Double(distance, side_a)$ was inferred in analog two.
2. A cause grouping **C** was formed in analog two, containing at least of the newly inferred proposition $Double(distance, side_a)$, and the already existing proposition $Halve(weight, side_a)$.
3. **C** did not contain any proposition that recommended impossible or in other ways contradictory actions. For example, if the proposition $Adversaries(side_a, side_b)$ is a part of **C**, since it is not impossible or contradictory (it is simply a restatement of another instance of the same proposition, which was already coded in and happens to be true), **C** is still regarded to have been successfully inferred. However, if the proposition $Balance(distance, side_a)$ is inferred, it is regarded to be a nonsensical proposition and the trial is regarded to have failed.

Table 2: *Trial Results. A trial was judged as correct if it met all criteria. Propositions which were removed or added are listed below the table.*

Change Made	% correct (n)	% correct (uwm)
A1: Removed proposition 1 from analog 2	35	32
A2: Removed proposition 2 from analog 1	0	0
A3: Removed proposition 3 from analog 1	8	4
A4: Removed propositions 3, 4 from analog 1	0	0
A5: Removed proposition 5 from analog 2	38	44
A6: Removed propositions 5, 6 from analog 2	66	71
A7: Removed propositions 3, 4 from analog 1, and 5, 6 from analog 2	55	54
C: No changes made	57	61

1. $Contributing_Factor(weight)$
2. $Contributing_Factor(marbles)$
3. $Adversaries(team_a, team_b)$
4. $Adversaries(team_b, team_a)$
5. $Adversaries(side_a, side_b)$
6. $Adversaries(side_b, side_a)$

4. A cause-effect grouping with **C** as the cause and **E** as the effect is created in analog two.

Table 2 presents the most interesting of these trials. Table 3 shows the results of a similar experiment which, instead of modifying the propositional knowledge, modified the predicates available to each analog.

Assuming that small variations in the filtering steps of our model's algorithm affect the analogs used in step 4, this series of trials was done to examine whether these small variations cause large effects in the success of the overall model's run. The control case for these trials was the set of ideal (manually created) analogs listed in the appendix. The success rates for the control case are shown in the last row of Tables 2 and 3.

Because of the small sample size, it is not worthwhile to place emphasis on small differences in the presented data. That being said, it is interesting to note that removing certain propositions had much more prominent effects on the success rates. Trial **A1** had an appreciable success rate, but trial **A2** failed every single time; not knowing that weight contributes to balance was less harmful to success than not knowing that the number of marbles played by each team contributes to fairness.

A similar result is observed in the difference between trials **A3** / **A4**, and **A5** / **A6**. This difference suggests one of two possibilities: either the correct mapping and inference depend on knowledge of an adversarial rela-

Table 3: *Trial Results. A trial was judged as correct if it met all criteria. Predicates which were removed or added are listed below the table.*

Change Made	% correct (n)	% correct (uwm)
B1: Added predicates 1, 2 to analog 1	51	59
B2: Added predicate 1 to analog 1	61	64
B3: Added predicate 2 to analog 1	64	54
B4: Added predicates 3, 4 to analog 2	0	0
B5: Added predicate 3 to analog 2	0	0
B6: Added predicate 4 to analog 2	63	62
B7: Added predicates 1, 2 to analog 1, and 3, 4 to analog 2	0	0
C: No changes made	57	61

1. *Halve*
2. *Double*
3. *Add_One*
4. *Remove_One*

tionship between the two teams of the game of marbles, which doesn't apply to knowledge of an adversarial relationship between the two sides of the balance beam; or this particular mismatch has some effect on the mapping algorithm that causes the process to fail.

To investigate this further, trial **A7** was carried out. Its results suggested the latter possibility, but raises further questions. Why does the mapping fail so consistently when the *Adversaries* propositions are missing from analog one, but not when they are missing from analog two? We do not have a clear answer to this question, but it is likely an idiosyncrasy of this particular set of analogs. In any case, this oddity illustrates the difficulty in determining which propositions should be included in each analog.

Moving on to table 3, we observe that the trials with non-zero success rates do not differ from the control case (trial **C**) to a statistically significant figure. The three trials which have zero success rates all have one thing in common, which the other trials lack—they all involve the addition of the *Add_One* predicate to analog 2. Intuitively, this makes sense, as the $Double(distance, side_a)$ proposition we want to be recruited into analog 2 must map to the proposition $Add_One(marble, team_a)$ in analog one; and nothing has more semantic units in common with a predicate than that predicate itself. Indeed, the results for trial **B5** show a lot of cases in which the proper cause-effect groups were made, but the proposition inferred was $Add_One(distance, side_a)$—not the correct output, but one that is part of a valid and testable hypothesis. It is interesting to consider that future work on the analogical mapping step may rectify this issue; allowing second-order relations, including the missing second-order knowledge we have already discussed in this paper, could lead to the correct output.

Finally, in true Piagetian fashion, it would be remiss of us not to discuss the errors themselves.[6] In successful cases, the *marble* object in analog one mapped to the *distance* object in analog two. However, a frequent error was to infer the existence of a new object, $marble'$, in analog 2. This allowed the (rather frequent) recruiting of the erroneous inferred proposition $Double(marble', side_a)$. In a more intuitive fashion, this proposition might be restated as the following: In the domain of the balance beam, there exists a concept ($marble'$), which is analogous to marbles in the game of marbles.

The results all seem to point to one glaring fact: this model's ability to reach the success condition will depend heavily on the filtering steps. It is for this reason that steps one through three will be the primary emphasis of future work on this model, which we elaborate on in the next section.

General Discussion and Future Work

Gentner and Forbus (2011) discuss a common problem with analogical models, which they call the "tailorability concern." Designers of analogical systems too often (sometimes without knowing it) carefully over-customize the input data for their example cases, in order to produce successful outcomes. This tends to limit the scope of possible problems that the system can successfully solve autonomously.

We very much sympathize with this concern, and given our careful selection of input data throughout this paper, acknowledge the inapplicability of this model in its current state to general problem solving. Gentner and Forbus offer at least one solution: "[U]se pre-existing databases and automatic (or semi-automatic) parsing and semantic representation of the input" (Gentner and Forbus 2011). This is a path which we hope to take as work continues on this model.

For example, integration with a larger cognitive model that allows for the simultaneous representation of rich semantic data, and symbolic, logical knowledge (such as CLARION (Sun 2002)) might offer us a way to automatically parse and manage large bases of information. The similarity-based and other retrieval mechanisms within such an architecture could then be incorporated into the filtering steps of our algorithm. Architectures which already have natural language processing functionality (such as Polyscheme (Cassimatis 2006)) can expedite the process of importing data from

[6] Part of Piaget's genius was his ability to not only notice when children made errors in reasoning, but to closely scrutinize both the types and causes of these errors.(Chapman 1988)

various sources; it would certainly be very interesting to see if the sorts of natural-language justifications given by such a system correspond to those observed in the classic Piagetian literature.

Appendix : Analogs Used

Analog A

Predicates The following listing is formatted as follows: The predicate name, followed by the semantic unit(s). For each semantic unit, two copies are made: one which connects to each role of the predicate. For example, *Remove_One* has two semantic units connected to its first role: *remove_one*1, and *makes_smaller*1. Its second role connects to the semantic units: *remove_one*2, and *makes_smaller*2. The only exception is *Contributing_Factor*, which only has one role.

- *Remove_One* - remove_one, makes_smaller
- *Add_One* - add_one, makes_bigger
- *Adversaries* - adversaries
- *Balanced* - balanced1, balanced2
- *Contributing_Factor* - cf1, cf2

Objects The objects are listed along with their semantic units.

- *team_a* - ta1, ta2
- *team_b* - tb1, tb2
- *player* - p1, p2
- *marble* - m1, m2

Propositions The proposition's identifier is listed, followed by the actual proposition.

- P1 - $Remove_One(player, team_a)$
- P2 - $Add_One(marble, team_a)$
- P3 - $Balanced(team_a, team_b)$
- P4 - $Balanced(team_b, team_a)$
- P5 - $Adversaries(team_a, team_b)$
- P6 - $Adversaries(team_b, team_a)$
- P10 - $Contributing_Factor(marble)$
- P11 - $Contributing_Factor(player)$

Groups The group identifier is listed, then the propositions and/or groups which are members, followed by the semantics of that group (in parentheses).

- G1 - P1, P2 (cause)
- G2 - P3, P4 (effect)
- G3 - G1, G2 (cause-relation)

Analog B

Predicates

- *Halve* - halve, makes_smaller
- *Double* - double, makes_bigger
- *Adversaries* - adversaries
- *Balanced* - balanced1, balanced2
- *Contributing_Factor* - cf1, cf2

Objects The objects are listed along with their semantic units.

- *side_a* - sa1, sa2
- *side_b* - sb1, sb2
- *weight* - w1, w2
- *distance* - d1, d2

Propositions The proposition's identifier is listed, followed by the actual proposition.

- P1 - $Halve(weight, side_a)$
- P3 - $Balanced(side_a, side_b)$
- P4 - $Balanced(side_b, side_a)$
- P5 - $Adversaries(side_b, side_a)$
- P6 - $Adversaries(side_a, side_b)$
- P10 - $Contributing_Factor(weight)$
- P11 - $Contributing_Factor(distance)$

Groups

- G2 - P3, P4 (effect)

Phase Set

The phase set, or the order in which the propositions fire, was as follows:

1. With analog 1 as source and 2 as target, fire 10 random propositions.
2. With analog 2 as source and 1 as target, fire 10 random propositions.
3. Begin self-supervised learning. LISA is now authorized to recruit new units if necessary.
4. With analog 1 as source and 2 as target, fire 15 random propositions.
5. With analog 2 as source and 1 as target, fire 15 random propositions.
6. With analog 1 as source and 2 as target, fire 15 random propositions.

The above sequence was repeated 100 times per trial.

References

Bringsjord, S., and Licato, J. forthcoming. Psychometric artificial general intelligence: The piaget-macguyver room. In *Theoretical Foundations of Artificial General Intelligence*. Atlantis Press.

Cassimatis, N. 2006. A Cognitive Substrate for Achieving Human-Level Intelligence.

Chapman, M. 1988. *Constructive Evolution: Origins and Development of Piaget's Thought*. Cambridge Univ Press.

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.

Ferretti, R. P., and Butterfield, E. C. 1986. Are children's rule-assessment classifications invariant across instances of problem types? *Child Development* 57(6).

Forbus, K.; Gentner, D.; and Law, K. 1995. Mac/fac: A model of similarity-based retrieval. *Cognitive Science* 19:141–205.

Forbus, K.; Klenk, M.; and Hinrichs, T. 2009. Compaion cognitive systems: Design goals and lessons learned so far. *IEEE Intelligent Systems* 24(4):36–46.

Gentner, D., and Forbus, K. 2011. Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(3):266–276.

Gentner, D.; Rattermann, M. J.; and Forbus, K. 1993. The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25:524–575.

Gentner, D. 1983. Structure-Mapping: A theoretical framework for Analogy. *Cognitive Science* 7:155–170.

Gentner, D. 2003. Why We're So Smart. In Gentner, D., and Golden-Meadow, S., eds., *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press. 195–235.

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

Halford, G. S., and McCredden, J. E. 1998. Cognitive science questions for cognitive development: The concepts of learning, analogy, and capacity. *Learning and Instruction* 8(4):289–308.

Halford, G. S.; Andrews, G.; Dalton, C.; Boag, C.; and Zielinski, T. 2002. Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology* 81(4):417–445.

Hofstadter, D. R. 2001. Epilogue: Analogy as the core of cognition. In Gentner, D.; Holyoak, K. J.; and Kokinov, B. N., eds., *The Analogical Mind: Perspectives from Cognitive Science*. The MIT Press. chapter 15.

Holyoak, K. J., and Koh, K. 1987. Surface and structural similarity in analogical transfer. *Memory and Cognition* 15(4):332–340.

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

Hummel, J. E., and Holyoak, K. J. 1997. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* 104:427–466.

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 Kokinov, B.; Holyoak, K. J.; and Gentner, D., eds., *New Frontiers in Analogy Research: Proceedings of the Second International Conference on Analogy*.

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

Jansen, B. R. J., and van der Maas, H. L. J. 1997. Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review* 17:321–357.

McClelland, J. 1989. Parallel distributed processing: Implications for cognition and development. In Morris, R., ed., *Parallel distributed processing: Implications for psychology and neurobiology*. Oxford University Press.

McClelland, J. 1995. A connectionist perspective on knowledge and development. In Simon, T., and Halford, G. S., eds., *Developing cognitive competence: new approaches to process modeling*. Psychology Press.

Normandeau, S.; Larivée, S.; Roulin, J.-L.; and Longeot, F. 1989. The balance-scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology* 150(3):237–249.

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

Ross, B. H. 1989. Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(3):456–468.

Shultz, T. R.; Mareschal, D.; and Schmidt, W. C. 1994. Modeling cognitive development on balance scale phenomena. *Machine Learning* 16:57–86.

Shultz, T. R. 2003. *Computational Developmental Psychology*. Cambridge, Massachusetts: The MIT Press.

Siegler, R. S. 1976. Three aspects of cognitive development. *Cognitive Psychology* 8:481–520.

Siegler, R. S. 1981. Developmental sequences within

and between concepts. *Monographs of the Society for Research in Child Development* 46(2):1–74.

Sun, R. 2002. *Duality of the Mind: A Bottom Up Approach Toward Cognition*. Lawrence Erlbaum Associates, Mah- wah, NJ.

van der Maas, H. L. J., and Raijmakers, M. E. 2009. Transitions in cognitive development: prospects and limitations of a neural dynamic approach. In Spencer, J.; Thomas, M. S. C.; and McClelland, J., eds., *Toward a new grand theory of development: Connectionism and dynamical systems theory re-considered*. Oxford Univ. Press. 229–312.

van Rijn, H.; van Someren, M.; and van der Maas, H. L. J. 2003. Modeling developmental transitions on the balance scale task. *Cognitive Science* 27:227–257.

Wilkening, F., and Anderson, N. H. 1982. Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structure. *Psychological Bulletin* 92(1):215–237.