# Hierarchical Complexity: A Formal Theory

Michael Lamport Commons
Program in Psychiatry and the Law Harvard Medical School
Massachusetts Mental Health Center
234 Huron Avenue, Cambridge, MA 02138
commons@tiac.net

Alexander Pekker
Department of Mathematics, Harvard University
1 Oxford Street, Cambridge, MA 02138
apekker@math.harvard.edu

August 15, 2004

**Abstract**

Theories of complexity have generally not addressed hierarchical complexity. However, within developmental psychology, notion of hierarchical complexity have come into being in the last twenty years. We outline the formal axioms for a model of hierarchical complexity which assigns an order of hierarchical complexity to every task regardless of domain. The orders correspond to natural numbers, thus ensuring that the orders are separated by gaps of equal size. The model naturally leads to the existence of performance stages, thereby formalizing many implicit properties of stage theories.
*Keywords:* Complexity, hierarchical complexity, formal theory, distributivity, stage, Rasch, Saltus.

## 1 Introduction

Distributivity is the property of addition and multiplication on the real numbers that ensures that $a \times (b + c) = (a \times b) + (a \times c)$. Of course distributivity also plays a fundamental role in more general contexts, such as the complex numbers and the definition of rings in modern algebra. The distributive law serves as a motivation for a newer form of complexity, called *hierarchical complexity,* which we aim to describe in this paper.

In particular, the distributive law suggests that the task of evaluating $a \times (b + c)$ is more complex than the task of evaluating $(a + b) + c$ or even the two-part task of first evaluating $a + b$ and then evaluating $c \times d$. The evaluation of $(a + b) + c$ is no more complex than addition, performed either as $(a + b) + c$ or $a+(b+c)$; the organization of the two actions of addition is arbitrary. Similarly, in the two-part task, evaluating $a + b$ and then $c \times d$ yields the same result as first evaluating $c \times d$ and then $a + b$. Both of these are *chain* actions. On the other hand, the evaluation of $a \times (b + c)$ requires a non-arbitrary organization of addition and multiplication, or, equivalently, the distributive law, and is therefore more complex than addition or multiplication. In modern algebra, the non-arbitrary *coordination* of addition and multiplication leads to the definition of rings, and the expressions in ring theory are usually more complex than the expressions in group theory (which involve only one operation).

We refer to addition and multiplication as actions, a term that is commonly used by developmental psychologists to refer to events that produce outcomes or, equivalently, accomplish certain tasks. The study of tasks appears in psychophysics (a branch of stimulus control theory in psychology) (Green & Swets, 1966; Luce, 1963) and in artificial intelligence (Goel & Chandrasekaran, 1992), and in general, actions may be attributed to organisms, computers, or others. Existent actions may be combined to produce new, more complex actions (Binder, 2000). Our goal is to describe how to measure the complexity of an action and to relate it to the complexity of other actions.

In the literature, two types of complexity have been identified (Commons, Trudeau, et al, 1998): horizontal (traditional) and vertical (hierarchical). (For a review of these definitions, see, e.g., Wolfram 2002 and Kauffman, 1993.) Roughly speaking, in traditional complexity, the complexity of an action is determined by the number of times a specific subaction is repeated. In hierarchical complexity, the complexity of an action is determined by the non-arbitrary way in which the subactions are organized, not how many subactions there are. In particular, provided all other things are held constant, the order of hierarchical complexity of an action is one greater than the order of hierarchical complexity of its subactions, provided they are organized in a non-arbitrary way.

To illustrate one difference between traditional and hierarchical complexity, consider the action $A$ of evaluating $1+2$ and the action $B$ of evaluating $(1+2)+3$. The traditional complexity of $A$ is smaller than the traditional complexity of $B$ since the action of addition is executed less often in $A$ than in $B$; on the other hand, since $A$ differs from $B$ only in how many times addition is executed, but not in the organization of the addition, $A$ and $B$ have the same hierarchical complexity. This example shows that the two types of complexity are independent and incommensurate.

Some examples of tasks and their orders in the model of hierarchical complexity are shown in Table 1. Most of our examples will be at orders 7, 8, 9, and 10, i.e., from the primary to the formal orders.

We state the main definitions and model axioms and illustrate them with examples in Sections 2 and 3. In essence, we formalize the idea of Piaget (e.g., Inhelder & Piaget, 1958) and Piaget's intellectual descendents (e.g., Campbell, 1991; Campbell & Bickhard, 1986; Tomasello & Farrar, 1986) that a higher-order action is defined in terms of lower-order actions in a non-arbitrary way by means of permutations. This formal construction allows us to separate the order of an action from participant performance, yielding a clear notion of stage of performance in Section 4. Finally, in Section 5, we define the measure of the an action as the minimum number of simple actions needed to accomplish the action; for an action of order $n$, its measure is $2\hat{}n$. We test the predictions of the model via Rasch and Saltus analysis of two task series in Sections 6 and 7, and we summarize our findings, particularly in relation to the stage of performance, in Section 8.

Some applications of this axiomatic model include stacked neural networks (Commons & White, 2004), programmed instruction in the discussion of prerequisites (Holland & Skinner 1961), and precision teaching in the discussion of combinations being built out of elements (e.g., Commons & Richards, 2002; Kubina & Morrison, 2000). Although the model itself has been previously described (Commons, Trudeau, et al, 1998), the formal, axiomatic version is presented here for the first time.

## 2 Actions

We begin by defining the fundamental terms. In a given system, there exist certain tasks that are to be accomplished. These tasks are accomplished via task-actions. Formally, a *task-action*, often abbreviated simply as an *action*, is defined inductively. There exists an unique *simple action* $\tilde{A}$, which is the simplest action possible in a system. (This is in agreement with Luce's choice theory (Luce, 1959).) Every other action $A$ consists of at least two (and possibly infinite many) previously defined actions and a *rule* for organizing those previously defined actions. Thus, every nonsimple action $A$ is an ordered pair $A = (\{A_1, \ldots\}, R\}$, where the first component is a multiset of at least two previously defined actions $A_i$ composing $A$ and $R$ is the rule for organizing those actions.

There are two categories of rules: chain rules and coordination rules. In a nonsimple action $A = (\{A_1, \ldots\}, R)$, a *chain rule* $R$ is simply a sequential execution of the actions $A_i$ in some order, but the order of the executions *does not matter*. That is, regardless of the order in which the subactions are executed,the result of $A$ is achieved. A *coordination rule*, on the other hand, requires the execution of the actions $A_i$ in some specific, non-arbitrary order, so that the order *does* matter.

We now formalize these notions. Suppose first that $A$ consists of finitely many subactions, i.e., $A = (\{A_1, A_2, \ldots, A_n\}, R)$. Given a permutation $\sigma = (i_1, i_2, \ldots, i_n)$ of the numbers $1, 2, \ldots n$, the execution of the $A_i$ according to $\sigma$ is simply

$$A_{i_1} A_{i_2} \ldots A_{i_n}.$$

In this notation, the rule $R$ is a chain rule if the outcome of $A$ is the same for all $n!$ permutations of the numbers $1, 2, \ldots, n$. That is, the outcome of the order of actions

$$A_{i_1} A_{i_2} \ldots A_{i_n}$$

is the same for all permutations $(i_1, i_2, \ldots, i_n)$ of $1, 2, \ldots, n$. The rule $R$ is a coordination rule if this is not the case; i.e., if there exists at least one permutation $\tau = (j_1, j_2, \ldots, j_n)$ of the numbers $1, 2, \ldots, n$ so that the execution of the actions $A_i$ according to $\tau$, i.e.,

$$A_{j_1} A_{j_2} \ldots A_{j_n},$$

is *not* the the same as the outcome of the action $A$. Hence, the the outcome of $A_i$ is given by at least one, but *not all*, permutations of the $A_i$. We extend similarly to the cases where $A$ consists of infinitely many actions.

We summarize these definitions as the first three action axioms; we will refine them in the following section.

(A1) There exists a simple action $\tilde{A}$.

(A2) Every action $A$ is either simple (so $A = \tilde{A}$) or composed of at least two previously defined actions $\{A_1, \ldots\}$ and a rule $R$ for organizing those actions (so $A = (\{A_1, \ldots\}, R)$).

(A3) Each rule is either a chain or a coordination.

To motivate the definition of hierarchical complexity in the next section, we will rely on the following example.

**Example 1**. Let $+$ and $\times$ denote the traditional addition and multiplication on the real numbers, and let $\oplus$ and $\otimes$ denote the traditional addition and multiplication of variables (having values, say, in the real numbers). Then, consider the following four actions.

(a) $A = (\{+, \times\}, R_A)$ consisting of $1 + 2$ (i.e., adding the numbers 1 and 2) followed by $3 \times 4$ (i.e., multiplying the numbers 3 and 4). Clearly, the order in which the two subactions are executed does not matter: adding 1 and 2 and then multiplying 3 and 4 yields the same results, namely 3 and 12, as multiplying 3 and 4 and then adding 1 and 2. Thus, $A$ is a chain action.

(b) $B = (\{+, \otimes\}, R_B)$ consisting of $1 + 2$ followed by $x \otimes y$. Again, the order in which the two subactions are executed does not matter: adding 1 and 2 and then multiplying $x$ and $y$ yields the same results, namely 3 and $xy$, as multiplying $x$ and $y$ and then adding 1 and 2. Thus, $B$ is also a chain action.

(c) $C = (\{+, \times\}, R_C)$ consisting of the expression $2 \times (3 + 4)$. This is not a chain, for the order of the subactions matters: if we multiply 2 and 3 first and then add 4, we get 10, not 14, which is the answer dictated by rule $R_C$ (i.e., adding 3 and 4 first and multiplying the result by 2). Thus $C$ is a coordination, not a chain.

(d) $D = (\{\oplus, \otimes\}, R_D)$ consisting of the expression $x \otimes (1 \oplus 2)$. Notice that since the expression involves real numbers *and* variables, we must necessarily use $\oplus$ and $\otimes$ and not simply $+$ and $\times$. In particular, because the distributive law dictates that

$$x \otimes (1 \oplus 2) = (x \otimes 1) \oplus (x \otimes 2),$$

we cannot replace $\oplus$ by $+$. This observation will be important in the next section. As in the previous case, it is clear that $D$ is a coordination action.

(e) $E = (\{\oplus, \otimes\}, R_E)$ consisting of the expression $x \otimes (y \oplus z)$. This is exactly the same as (c) but at a more abstract level, and is, therefore, a coordination rule.

# 3 Hierarchical Complexity

To each action $A$ we wish to associate a notion of that action's hierarchical complexity, $h(A)$. Since actions are defined inductively, so is the function $h$, known as the *order of the hierarchical complexity.* For a simple action $A$, we set $h(A) = 0$. For a non-simple action, $A = (\{A_1, \ldots\}, R)$, we have to consider several cases. To get an intuitive idea, we analyze the complexity of the actions in Example 1.

**Example 1 (Continued)**. Let $m$ be the hierarchical complexity of $+$ and $\times$, the traditional addition and multiplication on the real numbers, and let $n$ be the hierarchical complexity of the operations $\oplus$ and $\otimes$, the traditional addition and multiplication of variables. Intuitively we understand that $m < n$.

(a) Since action $A$ is a chain, with the order in which the subactions are executed irrelevant, executing $A$ does not require any skill beyond the execution of each of the subactions individually. Consequently, we expect $h(A) = \max(h(+), h(\times)) = m$.

(b) Similarly, $B$ is a chain rule, but executing $B$ requires being able to multiply at the abstract level (which is more complex than adding at the primary level), and so $h(B) = \max((h(+), h(\otimes)) = h(\otimes) = n$. Notice that unlike action $A$, action $B$ consists of subactions of different complexities.

(c) Observe now that action $C$ coordinates two subactions of the same order, namely $m$. Since the order in which the two subactions are executed is *nonarbitrary*, the hierarchical complexity of this action is *higher* than the complexity of its subactions: $h(C) > \max(h(+), h(\times)) = m$.

(d) As we remarked in Example 1, it may seem at first that action $D$ coordinates two actions of different orders, $+$ of lower order and $\otimes$ of higher order. However, due to the distributive law, it actually coordinates two actions of the same order, i.e., $n$. In particular, we observe that a coordinating action, at least in arithmetic, necessarily coordinates subactions of equal order. As in the previous case, we see that $h(D) > \max(h(\oplus), h(\otimes)) = n$.

(e) Lastly, as in (c), it is clear that $h(E) > \max(h(\oplus), h(\otimes)) = n$.

This analysis illustrates that the only way to raise hierarchical complexity is by coordinating actions of lower complexity. Moreover, from part (d) of Example 1, we obtain the fundamental condition that *coordination requires the subactions to be of the same order*. In light of Example 1, we now state the hierarchical complexity axioms which incorporate the action axioms (A1)-(A3). As before we denote the complexity of an action $A$ by $h(A)$.

**Hierarchical Complexity Axioms**

(HC1) There exists a simple action $\tilde{A}$, and $h(\tilde{A}) = 0$.

(HC2) Every nonsimple action $A = (\{A_1, \ldots\}, R)$ is either a chain of at least two previously defined actions of arbitrary orders of hierarchical complexity or a coordination of at least two previously defined actions *all of which have the same order of hierarchical complexity.*

(HC3) For a nonsimple action $A = (\{A_1, \ldots\}, R)$, $h(A) = \max_i h(A_i)$ if $A$ is a chain, and $h(A) = h(A_1) + 1$ if $A$ is a coordination.

Notice that by Axiom (HC2), a coordination action $A = (\{A_1, \ldots\}, R)$ necessarily coordinates subactions of equal orders of hierarchical complexity (i.e., $h(A_1) = h(A_2) = \ldots$), and so the order of hierarchical complexity of $A$ is one higher than the order of hierarchical complexity of all its subactions. (In particular, in the last equation in Axiom (HC3) we may replace $A_1$ by *any* subaction of $A$ and still obtain the same result.)

As a consequence of these axioms, we see that if we let $\mathcal{A}$ denote the collection of all actions in a given system, then the hierarchical complexity is a function $h : \mathcal{A} \to \mathbb{N}$, where $\mathbb{N} = \{0, 1, \ldots\}$ is the set of natural numbers (and zero) under the usual ordering. From the properties of the natural numbers, we immediately obtain the following four essential properties of hierarchical complexity.

**Consequences of Hierarchical Complexity Axioms**

(HC4) (Discreteness) The order of hierarchical complexity of any action is a nonnegative integer. In particular, there are gaps between orders.

(HC5) (Equal Spacing) The orders of hierarchical complexity are equally spaced, i.e., all gaps between orders are of size one.

(HC6) (Existence) If there exists an action of order $n$ and an action of order $n+2$, then there necessarily exists an action of order $n+1$.

(HC7) (Comparison) For any two actions $A$ and $B$, exactly one of the following holds: $h(A) > h(B)$, $h(A) = h(B)$, $h(A) < h(B)$. That is, the orders of hierarchical complexity of any two actions can be compared.

(HC8) (Transitivity) For any three actions $A, B$, and $C$, if $h(A) > h(B)$ and $h(B) > h(C)$, then $h(A) > h(C)$.

In light of Table 1 that describes the orders of hierarchical complexity for, among others, arithmetic tasks, we can assign the exact natural numbers corresponding to the orders of tasks in Example 1.

**Example 1 (Continued)**. According to Table 1, both $+$ and $\times$ have order 7, i.e., primary, while $\oplus$ and $\otimes$ have order 9, i.e., abstract.

(a) Since $A$ is a chain, $h(A) = \max(h(+), h(\times)) = 7$, i.e., also primary.

(b) Since $B$ is a chain, $h(B) = \max((h(+), h(\otimes)) = 9$, i.e., also abstract.

(c) Since $C$ is a coordination, $h(C) = \max(h(+), h(\times)) + 1 = 8$, i.e., concrete.

(d) Since $D$ is a coordination, $h(D) = \max(h(\oplus), h(\otimes)) = 10$, i.e., formal.

(e) Again, since $E$ is a coordination, $h(E) = \max(h(\oplus), h(\otimes)) + 1 = 10$, i.e., formal.

# 4 Stages

The notion of stages is fundamental in the description of human, organismic, and machine evolution. Previously it has been defined in some ad hoc ways; here we describe it formally in terms of the model of hierarchical complexity. Given a collection of actions $\mathcal{A}$ and a participant $S$ performing $\mathcal{A}$, the *stage of performance of $S$ on $\mathcal{A}$* is the highest order of the actions in $\mathcal{A}$ completed successfully, i.e., it is

$$\text{stage}(S, \mathcal{A}) = \max\{h(A) \mid A \in \mathcal{A} \text{ and } A \text{ completed successfully by } S\}.$$

Thus, the hierarchical complexity axioms (HC4) - (HC6) immediately imply the following three properties of stages:

**Properties of Stages:**

(S1) (Discreteness) The stage of performance is a nonnegative integer. In particular, stages are discontinuous with gaps.

(S2) (Equal Spacing) Stages are equally spaced, i.e., all gaps between are of size one.

(S3) (Existence) Skipping stages is impossible.

This is in agreement with previous definitions (Commons, Trudeau, et al, 1998; Commons & Miller, 2001.). We will return to the notion of stage in the experimental results in Sections 6 and 7. Table 2 lists the stages described by the model of hierarchical complexity.

# 5 Measure of Hierarchical Complexity

We define the *measure of complexity at order* $n$, denoted by $\varphi_n$, as the minimum number of simple actions required to complete an action of order $n$. By axioms (HC2) and (HC3), an action of order $n$ organizes at least two actions of order $n-1$, each of which in turn organizes at least two actions of order $n-2$, and so forth, until we reach the lowest-order, simple actions. Consequently, given the inductive definition of the hierarchical complexity orders, it is not surprising that $\varphi_n = 2^n$. Formally, a zero-order action, consists of at least one simple action, so $\varphi_0 = 1 = 2^0$. For the inductive case, suppose $\varphi_{n-1} = 2^{n-1}$. Since by axioms (HC2) and (HC3), an action of order $n$ is either a coordination of at least two actions of order $n-1$ or a chain which includes an action of order $n$ (and hence eventually is composed of at least two actions of order $n-1$), we have $\varphi_n = 2\varphi_{n-1} = 2^n$, by induction.

# 6 Rasch and Saltus Models and Hierarchical Complexity

The hierarchical complexity model makes four predictions that should be evident in real world data. First, in interviews that probe for stage of performance, the scoring of the stage derived from the model of hierarchical complexity should provide the clearest and most reliable account among all scoring systems. Second, the empirically scaled orders of complexity of tasks should match the analytically predicted sequence of orders of complexity of these tasks. Third, the empirically scaled orders of complexity of tasks of the same type and content should be related by a simple unidimensional linear transformation. Fourth, the empirically scaled orders of tasks should produce gaps due to the natural number scale of hierarchical complexity. The first prediction has been verified in (Dawson, 2002), and so we focus on the last three.

We use Rasch analysis (Rasch, 1966; Rasch, 1980) and the related Saltus analysis to test these predictions. (The relationship between the Rasch model and conjoint measurement is discussed in Brogden, 1977; Fischer, 1968; Keats, 1967; and Keats, 1971; for more on the Rasch model as an application of conjoint measurement to empirical data, see Young, 1972; Luce & Tukey, 1964; and Perline, Wright, and Wainer, 1979.) Suppose we have a collection of tasks with hierarchical orders of complexity $d_j$ $(1 \leq j \leq J)$ and a collection of participants with proclivities to answer correctly $b_i$ $(1 \leq i \leq I)$; the parameters $d_j$ and $b_i$ are determined analytically. The Rasch model predicts that participant $i$ completes task $j$ correctly with probability

$$P(X_{ij} = 1) = \frac{\exp(b_i - d_j)}{1 + \exp(b_i - d_j)}.$$

Evidently, the probability that participant $i$ fails to complete task $j$ correctly is

$$P(X_{ij} = 0) = 1 - P(X_{ij} = 1) = \frac{1}{1 + \exp(b_i - d_j)}.$$

The Rasch model assumes that (1) the performances are drawn from a single population with a common set of proficiencies and (2) the proficiencies of the participants and the orders of the tasks are distributed *continuously* on a single scale. Clearly, the second assumption is violated by the natural number nature of the hierarchical complexity model, for it predicts gaps both in participant proficiencies and task orders. Moreover, the participant data violates the first assumption. Nevertheless, Rasch analysis can be used to obtain useful evidence in support of hierarchical complexity; we expect the Rasch model to produce gaps in measured task orders; that is, we should find clusters of tasks of the same hierarchical complexity and a few scaled tasks between them. Still, we should investigate how Rasch analysis can mislead us and, how significant quantitatively the violation of the continuity assumption is?

To address these issues of discontinuity, Saltus analysis was developed (Wilson, 1984; Wilson, 1989). Here, in addition to the parameters $b_i$ and $d_j$, further parameters are introduced. Each participant is mapped to one or more of $H$ participant groups, and each task is mapped into exactly one of $K$ task classes. These groups correspond to different levels of proficiency and orders of complexity. This therefore yields additional parameters: the binary parameter $\phi_{ih}$ indicating whether person $i$ belongs to group $h$ and a continuous parameter $\tau_{hk}$ indicating the relationship between participant group $h$ and task class $k$. The parameters $\phi_{ih}$ and $\tau_{hk}$ are unobserved and must be determined from the data. The Saltus model then predicts that participant $i$ completes task $j$ correctly, with task $j$ belonging to class $k$, with probability

$$P(X_{ij} = 1) = \frac{\exp(b_i - d_j + \sum_{h=1}^{H} \phi_{ih}\tau_{hk})}{1 + \exp(b_i - d_j + \sum_{h=1}^{H} \phi_{ih}\tau_{hk})}.$$

As in the Rasch analysis, the probability that participant $i$ fails to complete task $j$ correctly is

$$P(X_{ij} = 0) = 1 - P(X_{ij} = 1) = \frac{1}{1 + \exp(b_i - d_j + \sum_{h=1}^{H} \phi_{ih}\tau_{hk})}.$$

Observe that if there is exactly one group of participants $(H = 1)$ and one class of tasks $(K = 1)$, then the Saltus model is exactly the same as the Rasch model.

Thus the Saltus model may be thought of as an aggregate of Rasch models for each group and class.

The parameters in the Saltus model provide a means of measuring the significance of the gaps. We use it here to investigate whether the expected response patterns occur with enough regularity to support the claim that there should be jump discontinuities between successive stage of hierarchical integration. Thus we take $K$ to be the number of orders of hierarchical complexity of tasks to be studied; within each class of tasks of the same hierarchical complexity we assume that the Rasch model holds. This means that the tasks are measured unidimensionally within each class, but the classes differ in their order of complexity. Consequently, a different latent dimension of ability is being measured in each class.

There are several ways of determining $H$, $K$, and the parameters $\phi_{ih}$ and $\tau_{hk}$. The first would be to fit the mixed Rasch models with different numbers of classes and to choose the best-fitting one (e.g., with the Akaike Information Criterion (AIC), the Best Information Criterion (BIC), or the Consistent Akaike Information Criterion (CAIC)). These criteria relate the likelihood of a model to the number of parameters in the model, preferring models with fewer parameters. Another way is to assign participants to latent classes not deterministically but with varying probabilities; the higher the mean probability, the more unequivocal the assignment of a participant to a given group. (For more details, see Mislevy and Wilson, 1996 and Rost, 2001.)

We should note that proficiencies in real-life experiments depend on more than just the hierarchical complexity of the tasks; for example, familiarity, horizontal complexity, context, and bias also play a role. For example, a participant may report that a lower-stage task in an unfamiliar domain is harder than a higher-stage task in a familiar domain. Also, when explanations of performance are part of the action performed, the minimal set of actions may include remembering the actions, as well as naming and reflecting upon the actions (King, Kitchener, Wood & Davison, 1989; Kitchener & King, 1990; Tappan, 1990); these additional non-stage requirements make an action more difficult (resulting, e.g., in a higher age at which a participant completes the tasks successfully), but do not raise the hierarchical complexity of the action. In the studies reported here, we control for some of these variables, such as support: providing support by giving examples, giving practice of more extensive training, or removing support by requiring participants to discover the problems, the questions, and the phenomena themselves.

# 7 Rasch and Saltus Analysis of Two Different Tasks

We analyzed the data for two experiments: the Balanced Beam Series and the Causality (Laundry) Series. The Balanced Beam Series (adapted from Inhelder & Piaget, 1958 and Commons, Goodheart, & Bresette, 1995), consists of sets of balance beam tasks in increasing order of hierarchical complexity. In each case, a beam is described as a fulcrum for which different size weights may be attached at different distances on each side. Participant responses were assessed using multiple choice answers.

The Laundry (Causality) Series (Commons, Miller, & Kuhn, 1982) is based on an isolation of variables problem called the pendulum problem (Inhelder & Piaget, 1958). The problems presented the participants with a different kind of a stain on a cloth and many different ingredients that were described as either removing the stain or not removing the stain. The task consisted of predicting which combination of ingredients would remove the stain. Although each configuration of variables was repeated once, this was not apparent until the problem had been solved.

Both exercises consisted of tasks at the concrete, abstract, formal, and systematic orders of complexity (i.e., orders 8, 9, 10, and 11), and participants included 5th and 6th graders and adults. Both problems were paper and pencil exercises. The tasks form a series because every higher order task has the lower order task embedded in it. For each of the tasks Quest software (Adams & Khoo, 1993) generated a separate Rasch model.

The Balance Beam Series results support our prediction from the model of hierarchical complexity that the Balanced Beam Series may be viewed as a measurement in a single dimension of performance even if additional parameters might improve the fit. The tasks that were posited to be less complex were indeed easier for the participants (Commons et al, 1997). The tight linear relationship between difficulty and hierarchical complexity was observed, thus supporting the second and third predictions of the model. The logarithm of the scaled item difficulty (called *threshold* by Rasch analysts) is graphed against the order of hierarchical complexity in Figure 1. Recall from Section 5 that the measure of hierarchical complexity order $n$ is $2^n$, so we should expect a linear plot in Figure 1; indeed, this is observed. The regression equation with 16 participants yields excellent agreement: the correlation coefficient is $r(16) = 0.9244$ ($r^2 = 0.8545$); the $F$-value is $F(1, 16) = 93.96$; and the $p$-value is $p < 0.00005$.

As the model of hierarchical complexity predicts, the items form a series of clusters along the dimension corresponding to their order of hierarchical complexity. The order of hierarchical complexity, which reflects the order in which items are learned, predicted the item difficulty correctly with notable exception at the formal and systematic levels.

A Saltus analysis was used to address the third prediction, whether the natural number of hierarchical complexity did produce gaps in task difficulty. In the balance beam study, a two-level Saltus model was used to examine the gap between the compact/abstract and formal/systematic classes of tasks. The two-level model was a better predictor of performance than the simple Rasch analysis, supporting the existence of gaps: $\chi^2(4) = 71.91$ and $p < 0.01$.

The results from the Laundry (Causality) Series were very similar (Goodheart & Dawson, 1996; Goodheart et al, 1997), providing additional support for the notion that task difficulty is measured along a single dimension. The regression equation for difficulty versus hierarchical complexity also yields excellent agreement with correlation coefficient $r(16) = 0.918$ ($r^2 = 0.843$); the $F$-value is $F(1, 22) = 118.42$; and the $p$-value is $p < 0.00005$. The regression plot may be seen in Figure 2.

# 8 Conclusions

In this paper we have presented a formal model of hierarchical complexity that leads to a quantal notion of stage. The key feature of the model is that the orders of hierarchical complexity of actions correspond to the natural numbers. As a result, we obtain the following predictions for stages:

1. Sequentiality of stages is perfect; skipping stages is not possible. This has been shown here and elsewhere (Dawson, Commons, Wilson, in preparation).

2. Because orders of hierarchical complexity are equally spaced, groups of tasks at different orders of complexity should cluster in well-defined and equally spaced groups. This was observed using Rasch analysis here and elsewhere (Dawson, Commons, Wilson, in preparation).

3. Because task orders have gaps, there exist no intermediate stages of performances, i.e., stages are discontinuous. This was shown by a Saltus model.

4. Participants generally perform in a consistent manner across tasks of the same complexity. Most performances are predominantly at their most frequent stage of performances.

The model of hierarchical complexity sets forth the core requirements for a theory of stages; many researchers posit more core requirements (e.g., Fischer, 1980), but none require fewer. Any model that fails to account for the hierarchical complexity of tasks in the definition of stage will by definition fail to yield results that are accurate or even significant as to the order of developmental complexity.

The establishment of an analytic measure of hierarchical complexity has many benefits for cognitive science, psychology and evolutionary studies. First, by classifying task complexity analytically, such a model produces measures that are independent of observation and of actual organism or machine performance, thus leading to a greater degree of accuracy and consistency in stage measurement. Second, the model defines a single sequence underlying all domains of development whether intragenerational (as in animal or human development) or intergenerational (as in neural networks) (e.g., Kohlberg & Armon, 1984).

The model of hierarchical complexity answers some of the most fundamental questions about complexity. By theoretically presenting a method for the analysis of tasks and deriving an actual task chain, the model demonstrates that such chains exist. By removing the notion of domain, i.e., the general area to which a task belongs, from the measurement of the hierarchy of tasks, the model shows that the sequences of orders and stages are invariable across all domains; in particular, task complexity remains unchanged regardless of how broadly or narrowly domains are defined. We caution, however, that an action's difficulty is related to but different from its complexity; while difficulty depends on the domain and other non-stage properties of the task, complexity does not.

Finally, using formal axioms, we obtain an analytic model of stage development, which formalizes key notions stated implicitly in most stage theories and predicts that stages exist as more than *ad hoc* descriptions of sequential changes in human actions. As such it offers clarity and consistency to the field of stage theory and to the study of development and evolution in general.

# Authors' Note

During a conversation among R. Duncan Luce, the first author, and the first author's son Lucas Commons-Miller, there was a discussion of why nonarbitrary organization of actions was at the heart of hierarchical complexity, and therefore stage theory and distributivity. Some curious issues of distributivity are addressed by Luce (2004). Our paper grows out of that discussion with the authors taking the blame for its faults and Luce and Lucas taking the credit for the inspiration, if not its strengths. Also Luce read many drafts of this paper and tried to correct many of its weaknesses.

Portions of this paper appear in Commons, Trudeau, et. al (1998), Commons and Miller (1998),. Minimal parts of this paper were based upon material from LaLlave and Commons (1996). They are reproduced here with permission of the respective authors and publishers. Some parts of this paper were presented at the Society for Research in Child Development, April 1987, the Third Beyond Formal Operations Symposium held at Harvard: Positive Development During Adolescence and Adulthood, June, 1987, and the 17th Annual Convention for the Association of Behavior Analysis, May, 1991. 

Address reprint requests to Michael Lamport Commons, Department of Psychiatry, Harvard Medical School, Massachusetts Mental Health Center, 234 Huron Avenue, Cambridge, MA 02138. Commons@tiac.net.

## References


Adams, R. J & Khoo, S.-T. (1993). *Quest: The interactive test analysis system.* Victoria, Australia Australian Council for Educational Research Ltd.

Binder, C. (2000). *Component/composite analysis and programming from the "real world."* Presented at the Association for Behavior Analysis, Washington D.C.

Brogden H. E. (1977). The Rasch model, the law of comparative judgment, and additive conjoint measurement. *Psychometrika*, 42, 631-634.

Campbell, R. L. (1991). Does class inclusion have mathematical prerequisites? C*ognitive Development*, 6(2), 169-194.

Campbell, R. L. & Bickhard, M. H. (1986). *Knowing levels and developmental stages. Human Development*, 16.

Commons, M. L., Goodheart, E. A., Bresette, L. M., with Bauer, N. F, Farrell, E. W., McCarthy, K. G., Danaher, D. L., Richards, F. A., Ellis, J. B., O'Brien, A. M., Rodriguez, J. A., & Schrader, D. (1995). Formal, systematic, and metasystematic operations with a balance-beam task series: A reply to Kallio's claim of no distinct systematic stage. *Adult Development*, 2(3), 193-199.

Commons, M. L, Goodheart, E. A., Dawson, T. L., Draney, K. (1997). *A Saltus analysis of behavioral developmental data from the balance beam task series.* Presented at the International Objective Measurement Workshop, Judd Hall, University of Chicago.

Commons, M. L., & Miller, P. A. (2001). A quantitative behavioral model of developmental stage based upon hierarchical complexity theory. *Behavior Analyst Today*, 2(3), 222-240.

Commons, M. L., Miller, P. M., & Kuhn, D. (1982). The relation between formal operational reasoning and academic course selection and performance among college freshmen and sophomores. *Journal of Applied Developmental Psychology*, 3, 1-10.

Commons, M. L. & Richards, F. A. (2002). Organizing components into combinations: How stage transition works. *Journal of Adult Development*, 9(2).

Commons, M. L. Trudeau, E. J., Stein, S. A., Richards, F. A., Krause, S. R. (1998). The existence of developmental stages as shown by the hierarchical complexity of tasks. *Developmental Review*, 8 (3). 237-278.

Commons, M. L. & White, M. S. (2004). A complete theory of tests for a theory of mind must consider hierarchical complexity and stage: A commentary on Anderson and Lebiere target article, The Newell Test for a theory of mind. *Behavioral and Brain Sciences.*

Dawson, T. L. (2002). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement.* 3(2), 146-189.

Dawson, T. L., Commons, M. L., Wilson, M. (In preparation). *The shape of development.*

Fischer, G. (1968). *Psychologische Test Theorie.* Bern: Huber.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477-531.

Goel, A., & Chandrasekaran, B. (1992). Case-Based Design: A Task Analysis in C. Tong and D. Sriram (editors), *Aritificial intelligence approaches to Egineering Design, Volume II: Innovative Design* (pp. 165-184). San Diego: Academic Press.

Goodheart, E. A. & Dawson, T. L. (1996). *A Rasch Analysis of developmental data from the laundry problem task series.* Presented at the 11th Annual Adult Development Symposium, Boston, MA.

Goodheart, E. A., Dawson, T. L., Draney, K., & Commons, M. L. (1997). *A Saltus analysis of developmental data from the laundry problem task series.* Presented at the International Objective Measurement Workshop, Judd Hall, University of Chicago.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Huntington, NY: Krieger.

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior: A program for self-instruction.* New York, McGraw-Hill.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the development of formal operational structures.* (A. Parsons & S. Milgrim, Trans.). New York: Basic Books. (Originally published 1955.)

Kauffman S. A. (1993). *The origins of order: self-organization and selection in evolution.* New York: Oxford University Press.

Keats, J. A. (1967). Test theory. *Annual Review of Psychology*, 18, 217-238.

Keats, J. A. (1971). *An Introduction to Quantitative Psychology.* Sydney: John Wiley.

Kohlberg, L. & Armon, C. (1984). Three types of stage models used in the study of adult development in M. L. Commons, F. A. Richards, & C. Armon (editors), *Beyond formal operations, Volume 1: Late adolescent and adult cognitive development* (pp. 357-381). New York: Praeger.

Kubina, R. M. & Morrison, R. S. (2000). Fluency in education. *Behavior & Social Issues*, 10(1-2), 83-99.

LaLlave J. & Commoms, M. L. (1996). Comparing psychodynamic, cognitive-developmental and relations-developmental views of psychological boundaries, in Commons, M. L., Derrick, J. and Goldberg, C. (editors), *Clinical approaches to adult development* (pp. 175-210). Norwood, NJ: Wiley.

Luce, R. D. (1959). On the possible psychological laws. *The Psychological Review*, 66(2) 81-95.

Luce, R. D. (1963). Detection and recognition, in Luce, R. and Bush, R. (editors), *Handbook of Mathematical Psychology* (pp. 103-190). New York: Wiley.

Luce, R.D. (2004). Increasing increment generalizations of rank-dependent theories. *Theory and Decision*, 55, 87-146.

Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology.* 1, 1-27.

Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, 61, 41-47.

Perline, R. Wright, B. D., & Wainer, H. (1979). The Rasch Model as Additive Conjoint Measurement. *Applied Psychological Measurement*, 3(2), 237-255

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57.

Rasch, G. (1980). *Probabilistic model for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Rost, J. (2001). The Growing Family of Rasch Models, in A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (editors). *Essays on item response theory.* (pp. 25-42). New York: Springer-Verlag.

Tomasello, M. & Farrar, M. J. (1986). Object permanence and relational words: A lexical training study. *Journal of Child Language*, 13(3), 495-505.

Wilson, M. (1984). *A psychometric model of hierarchical development* (unpublished doctoral dissertation). University of Chicago.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.

Wolfram, S. (2002). Notes for Chapter 10: Processes of perception and analysis in *A new kind of science.* Champaign, IL: Wolfram Media.

Young, F. W. (1972). A model for polynomial conjoint analysis algorithms in R. B. Shepard, A. K. Romney, S. B. Nerlove (editors.), Multidimensional scaling. *Theory and applications in the behavioral sciences.* New York: Seminar Press.

# Tables

| Order | Name | Example |
|---|---|---|
| 0 | Calculatory | Simple machine arithmetic on bits |
| 1 | Sensory and motor | Seeing circles, squares, etc. or touching them |
| 2 | Circular sensory-motor | Reaching and grasping a circle or a square |
| 3 | Sensory-motor | A class of circles or squares can be made |
| 4 | Nominal | A class of circles or squares can be named "Circles" or "Squares" |
| 5 | Sentential | The numbers 1, 2, 3, 4, 5 may be said in order |
| 6 | Preoperational | The numbers 1, 2, 3, 4, 5 may be counted and named (i.e., 5 is named "Five" or "Cinco") |
| 7 | Primary | Simple arithmetic operations: $$1 + 3 = 4 \quad \text{and} \quad 5 \times 4 = 20$$ |
| 8 | Concrete | Tasks involving order of simple arithmetic actions, including distributivity: $$5 \times (1 + 3) = (5 \times 1) + (5 \times 3) = 5 + 15 = 20$$ |
| 9 | Abstract | All the forms of the number 5 are equivalent to the same variables value $x = 5$; forming a class based on an abstract feature |
| 10 | Formal | The general left-hand distributive law: $$x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$$ |
| 11 | Systematic | The left-hand distributive law with addition and multiplication interchanged is not true |
| 12 | Metasystematic | Distributivity in propositional logic and elementary set theory are isomporhic: $$\begin{aligned} x \wedge (y \vee z) &= (x \wedge y) \vee (x \wedge z) \\ A \cap (B \cup C) &= (A \cap B) \cup (AcapC) \end{aligned}$$ |
| 13 | Paradigmatic | General distributive systems present in various areas of mathematics, including set theory, logic, algebra, probability theory, etc. |
| 14 | Crossparadigmatic | Integration of mathematics and other fields, such as physics (e.g., quantum mechanics, the standard model of elementary particles, special and general relativity) |

Table 1. A sequence of actions at different orders of hierarchical complexity

| Order | What They Do | How They Do It | Example |
|---|---|---|---|
| 0 | Exact – no generalization | Human-made program manipulates bits | None |
| 1 | Discrimination in a rote fashion, stimuli, generalization, move | Move limbs, lips, eyes, head, view objects and movement | Discriminative and conditioned stimuli |
| 2 | Form open-ended classes | Reach, touch, grab, shake objects, babble | Open-ended classes, phonemes |
| 3 | Form concepts | Respond to stimuli in a class successfully | Morphemes, concepts |
| 4 | Find relations among concepts, use names | Use names and other words as successful commands | Single words: ejaculatives and exclamations, verbs, nouns, number names, letter names |
| 5 | Imitate and acquire sequences, follow short sequential acts | Generalize match-dependent task actions; chain words | Pronouns: my, mine, I; yours, you; we, ours |
| 6 | Make simple deductions, follow lists of short sequential acts, tell stories | Count random events and objects; combine numbers and simple propositions | Connectives: as, when, then, why, before; products of simple operations |
| 7 | Simple logical deduction and empirical rules involving time sequence, simple arithmetic | Add, subtract, multiply divide, count, prove, execute series of tasks on own | Times, places,actors, arithmetic outcomes from calculations |
| 8 | Carry out full arithmetic form cliques, plan deals | Execute long division, follow complex social rules, coordinate perspective of oneself and others | Interrelations, social events, reasonable deals |
| 9 | Discriminate variables such as stereotypes, logical qualifications (none, some, all) | Form variables out of finite classes; make and quantify propositions | Variable time, place, act, actor, state, type; quantifiers; categorical assertions (e.g., "We all die") |
| 10 | Argue using empirical or logical evidence, logic is linear | Solve problems with one unknown using algebra, logic, and empiricism | Relatioships are formed out of variables; words: linear, logical; if-then, thus, therefore, because; correct scientific solutions |
| 11 | Construct multivariate systems and matrices | Coordinate more than one variable as input; consider relationships in contexts | Events and concepts in multivariate contexts; systems consisting of relations: legal, societal, economic |

| Order | What They Do | How They Do It | Example |
|---|---|---|---|
| 12 | Construct multisystems and metasystems out of disparate systems | Create supersystems out of systems; compare systems and perspectives; name properties of systems: isomoprhic, homeomorphic, complete, consistent | Supersystems and metasystems are formed out of systems of relationships |
| 13 | Fit metaystems together to form new paradigms | Synthesize metaystems | Paradigms are formed out of multiple metasystems |
| 14 | Fit paradigms together to form new fields | Form new fields by crossing paradigms | New fields are formed out of multiple paradigms |

Table 2. Stages described by the model of hierarchical complexity

# Figures

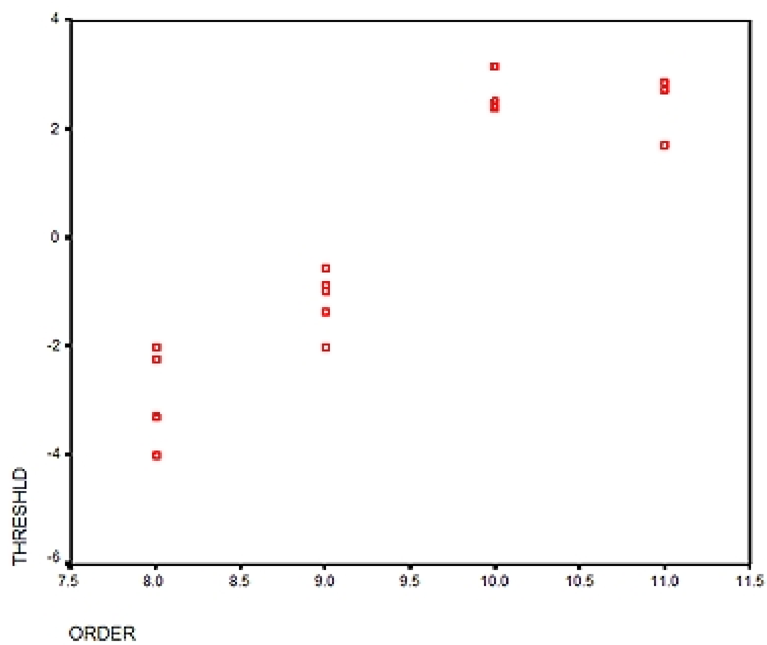


Figure 1. Threshold vs. task order for the Balanced Beam Series

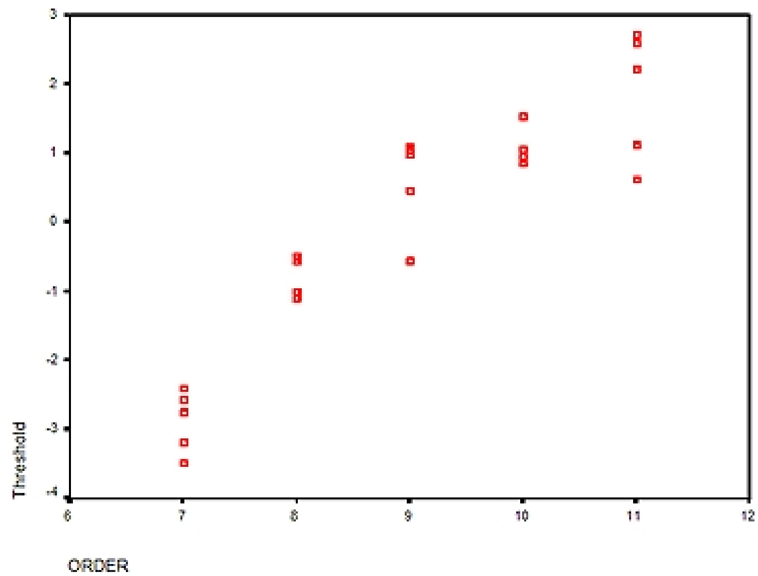


Figure 2. Threshold vs. task order for the Laundry (Causality) Series