The Development of Size Sequencing Skills: An Empirical and Computational Analysis

Maggie McGonigle-Chalmers & Iain Kusel

Monographs of the Society for Research in Child Development
Lynn S. Liben, Editor

VOL. 84 | NO. 4 | 2019 | SERIAL NO. 335

SRCD
Society for Research in Child Development

INFORMATION FOR AUTHORS

The *Monographs of the Society for Research in Child Development* is a quarterly journal which publishes conceptually rich and empirically distinguished work in support of the SRCD mission—advancing developmental science and promoting its use to improve human lives. To be accepted for publication, manuscripts must be judged to serve one or more core goals of SRCD's Strategic Plan, which include (1) the advancement of cutting-edge and integrative developmental science research; (2) the fostering of racial, cultural, economic, national, and contextual diversity in developmental science; and (3) the application of developmental science to policies and practices that improve human well-being.

As explained more fully in the document *Editorial Statement* (https://www.srcd.org/publications/monographs/monographs-editorial-statement), such goals may be achieved in varied ways. For example, a monograph might integrate results across waves of a longitudinal study; consolidate and interpret research literatures through meta-analyses; examine developmental phenomena across diverse and intersecting ethnic, geographic, economic, political, or historical contexts; describe and demonstrate new developmental tools for data acquisition, visualization, analysis, sharing, or replication; or report work contributing to conceptualizing, designing, implementing, and evaluating national and global programs and policies (e.g., related to parenting, education, or health).

The journal has an extensive reach via direct distribution to 5,000+ SRCD members and to readers via access provided through 12,000+ institutions worldwide. Beginning with the first issue under the Editorship of Lynn Liben (84.2, June 2019), this reach will be enhanced by a companion web- and media-based forum *Monograph Matters (MM)*—designed to expand *Monographs*' contributions to scholarship, teaching, and outreach. *MM* may be reached via monographmatters.srcd.org. *MM* includes materials such as invited commentaries, video highlights of monograph content, video tutorials on methods, teaching resources, and discussions between authors and end-users (e.g., parents, teachers, and policymakers).

Each *Monograph* should be a cohesive contribution, written by an individual author or by a collaborative team. It should be engaging and accessible not only to other experts in the specialized topic of the *Monograph* but also to the general SRCD audience. Potential authors are strongly advised to begin by sending a brief description of the proposed *Monograph* directly to the Editor, liben@psu.edu. If the editor judges the idea to be suitable for the journal, she will invite and then comment on a more detailed prospectus. These initial steps are completely optional; all editorial decisions will be based on peer reviews of manuscripts as submitted. Complete manuscripts (text, tables, figures, references, etc.) should be about 150–200 double-spaced manuscript pages, and should be submitted electronically via http://mc.manuscriptcentral.com/mono. Details about manuscript preparation, submission, author assurances, and review are provided in the journal's Submission Guidelines, https://www.srcd.org/Monographs-Submission-Guidelines.

Inquiries may be addressed to the Editor:

Lynn S. Liben
Department of Psychology
The Pennsylvania State University
University Park, PA 16802
liben@psu.edu

The Development of Size Sequencing Skills: An Empirical and Computational Analysis

Maggie McGonigle-Chalmers
University of Edinburgh

Iain Kusel
University of Edinburgh

Lynn S. Liben
Series Editor

Monographs of the Society for Research in Child Development
Vol. 84, No. 4, 2019, Serial No. 335

WILEY

Boston, Massachusetts *Oxford, United Kingdom*

The Development of Size Sequencing Skills: An Empirical and Computational Analysis

Contents

The Development of Size Sequencing Skills: An Empirical and Computational Analysis

Maggie McGonigle-Chalmers[1] *and Iain Kusel*[1]

Abstract We explore a long-observed phenomenon in children's cognitive development known as size seriation. It is not until children are around 7 years of age that they spontaneously use a strict ascending or descending order of magnitude to organize sets of objects differing in size. Incomplete and inaccurate ordering shown by younger children has been thought to be related to their incomplete grasp of the mathematical concept of a unit. Piaget first brought attention to children's difficulties in solving ordering and size-matching tests, but his tasks and explanations have been progressively neglected due to major theoretical shifts in scholarship on developmental cognition. A cogent alternative to his account has never emerged, leaving size seriation and related abilities as an unexplained case of discontinuity in mental growth. In this monograph, we use a new training methodology, together with computational modeling of the data to offer a new explanation of size seriation development and the emergence of related skills.

We describe a connected set of touchscreen tasks that measure the abilities of 5- and 7-year-old children to (a) learn a linear size sequence of five or seven items and (b) identify unique (unit) values within those same sets, such as second biggest and middle-sized. Older children required little or no training to succeed in the sequencing tasks, whereas younger children evinced trial-and-error performance. Marked age differences were found on ordinal identification tasks using matching-to-sample and other methods. Confirming Piaget's findings, these tasks generated learning data with which to develop a computational model of the change.

Using variables to represent working and long-term memory (WM and LTM), the computational model represents the information processing of the younger child in terms of a perception-action feedback loop, resulting in a

[1]Department of Psychology, University of Edinburgh

Corresponding author: Maggie McGonigle-Chalmers, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK, email: Maggie.McGonigle@ed.ac.uk

Citation Information: McGonigle-Chalmers, M. & Kusel, I. (2019).The Development of Size Sequencing Skills: An Empirical and Computational Analysis, 84(4).

DOI: 10.1111/mono.12411

heuristic for achieving a correct sequence. To explain why older children do not require training on the size task, it was hypothesized that an increase in WM to a certain threshold level provides the information-processing capacity to allow the participant to start to detect a minimum interval between each item in the selection. The probabilistic heuristic is thus thought to be replaced during a transitional stage by a serial algorithm that guarantees success. The minimum interval discovery has the effect of controlling search for the next item in a principled monotonic direction. Through a minor additional processing step, this algorithm permits relatively easy identification of ordinal values.

The model was tested by simulating the perceptual learning and action selection processes thought to be taking place during trial-and-error sequencing. Error distributions were generated across each item in the sequence and these were found to correspond to the error patterns shown by 5-year-olds. The algorithm that is thought to emerge from successful learning was also tested. It simulated high levels of success on seriation and also on ordinal identification tasks, as shown by 7-year-olds.

An unexpected finding from the empirical studies was that, unlike adults, the 7-year-old children showed marked difficulty when they had to compute ordinal size values in tasks that did not permit the use of the serial algorithm. For example, when required to learn a non-monotonic sequence where the ordinal values were in a fixed random order such as "second biggest, middle-sized, smallest, second smallest, biggest," each item has to be found without reference to the "smallest difference" rule used by the algorithm. The difficulty evinced by 7-year-olds was consistent with the idea that the information in LTM is integrally tied to the search procedure itself as a search-and-stop based on a cumulative tally, as distinct from being accessed from a more permanent and atemporal store of stand-alone ordinal values in LTM. The implications of this possible constraint in understanding are discussed in terms of further developmental changes.

We conclude that the seriation behavior shown by children at around 7 years represents a qualitative shift in their understanding but not in the sense that Piaget first proposed. We see the emergent algorithm as an information-reducing device, representing a default strategy for how humans come to deal with potentially complex sets of relations. We argue this with regard to counting behaviors in children and also with regard to how linear monotonic devices for resolving certain logical tasks endure into adulthood.

Insofar as the monograph reprises any aspect of the Piagetian account, it is in his highlighting of an important cognitive discontinuity in logicomathematical understanding at around the age of 7, and his quest for understanding the transactions with the physical world that lead to it.

I. Discontinuity in Discrete Set Understanding: An Enduring Issue

The Development of Discrete Set Understanding

Piaget's theory has had one of the most enduring influences on the way we think about children's grasp of the basic underpinnings of all number systems by claiming that around the age 6 or 7 years, children appear to gain sudden insight into the logicomathematical properties of discrete sets composed of simple tangible elements. Whether they are beads, blocks, or counters, and whether presented in a row, heap, or collection, these elements have intrinsic logicomathematical interconnections simply by virtue of their membership within a countable set. The interconnections that define such objects relative to one another and to the total set are as follows.

Cardinal Value. The term *cardinal value* refers to the total number in a set, commonly arrived at by counting. However, the cardinal value stands both for the final item of a count (e.g., the number six) but also represents the numerosity of that set (six). Cardinality applies to any set of discrete objects in a countable collection.

Ordinal Value. This term applies to objects (such as sticks or blocks) that vary along a single dimension such as size or weight. The ordinal value of a given item is its unique position along the differentiating dimension (e.g., the third biggest or heaviest). The ordinal value is its permanent, logicomathematical—as opposed to its spatio-temporal—position within the set (e.g., third from the left). Its ordinal value within the set does not change even when its location within the set is changed.

Unit. The common denominator to both ordinality and cardinality is the concept of the unit. Arriving at a specific ordinal value by seriating (ordering from least to greatest or the reverse), or at a cardinal value by counting, necessarily represents increments using the smallest divisible interval.

Piaget's theory was that these properties and their interrelationships are gradually constructed through direct interaction with the physical world, and that, in a very real sense, they do not exist in the mind of the preschool and early school-aged child. His interest in children's mental development was therefore primarily an exercise in *epistemology*, or in revealing the nature of knowledge, rather than an exercise in developmental psychology per se. By exploring how children interact with sets and collections of objects during the preschool to early school years, he set about illustrating how the principles of logic and mathematics are acquired through a process of active individual discovery.

Once these logicomathematical principles have been acquired, the child's behaviors with regard to discrete sets are no longer exploratory but conform to logical *operations*. The child at this stage (at around 7 years) is described accordingly as *operational*. Understanding the properties as defined above is therefore seen by Piaget to be an outcome of extensive experience. The key to following his theory of logicomathematical development is to suspend the adult conceptions of number and unit value and to try to share his vision of how these concepts come into being.

A second important key to Piaget's theory is that it embraces two distinct aspects, and both employ the concept of an *operation*. One is the practical and experiential means by which discoveries are made regarding rules governing the physical world—a *functionalist* explanation sometimes described as *transactionalism*. An operation in this context would be any action performed by the child whose transformative effect could be directly observed, such as moving an object back and forth from A to B, or opening and closing a container. Another is the stage-like progression of logical understanding arising later from these fundamental sensorimotor discoveries. Here, Piaget references logicomathematical terms as listed above, but in the sense of showing that concepts such as ordinality and cardinality are not (yet) used to reason about the physical world by children less than 6 years old, who are thus labeled *preoperational*. An end-state defined approach, this aspect of his account is described as *structuralist*, in which an "operation" now also refers to a logical transformation whose crucial property is *reversibility* (e.g., $A > B$ implies $B < A$). We consider in due course concerns as to whether the transactionalist and structuralist aspects of his account are in fact compatible, and specifically how and whether operations at the level of action can lead to the mental operations that form the core of his structural account.

These concerns notwithstanding, an enduring legacy to developmental psychology from Piaget's research was the exercise of documenting and explaining how children deal with discrete sets during the preoperational phase. From a young age, children are invited to attend to the numerosity of sets, relations between and among items such as "bigger than" and "smaller than," to sort and order items accordingly, and so on. These are natural and enculturated activities but are subject to slow development and improvement. By turning these activities into a set of dedicated experimental tasks, and by recording behavior and interview responses from preschool and early school-aged children, Piaget set about detailing how these activities are the basis on which ordinality, cardinality, unitization, and the very concept of number develops.

A key element determining the design of the tasks was the idea that children should discover that discrete sets are decomposable into unique indivisible units that are logically immune to how they are displayed in visual space. We see later how this was tested in conservation and one-to-one correspondence tasks. To understand Piaget's tasks and to follow his thinking and conclusions, it is helpful to consult original or translated volumes in which Piaget, often in collaboration with Inhelder, describes his research first-hand (Inhelder & Piaget, 1964; Piaget, 1952a; Piaget & Inhelder, 1971, 1974). Chapman (1988) provides a thorough exposition and critique of Piaget's structuralist and functionalist approach.

A History Leading to Progressive Neglect

In the usual manner of theoretical changes and developments in all scientific disciplines, Piaget's work on logicomathematical development has seen many shifts of focus and emphasis in the writings of later researchers. One of the topics that has suffered from this shift has been Piaget's claim that at around the age of 6 or 7, children gain a sudden insight into the logicomathematical properties of discrete sets, and in so doing, exhibit an important discontinuity in their cognitive development. Without Piaget's claim being repudiated directly, it has attracted diminishing interest. As criticisms of Piaget's theoretical position started to emerge, his theory was slowly replaced with alternative visions of developmental growth very few of which accepted his concept of operational development. There were several strands to the growing critique. In brief, and in roughly chronological order, these include a theoretical rebuff of Piaget's logical take on the mastery of skills that were thought to be more convincingly explained in terms of contemporary cognitive psychology, an empirical challenge to Piaget's claim that younger children lack logical insight, a rejection of the idea of domain-general advances at particular ages and stages, and, most radically of all, the growing popularity of theories offering innate explanations of competences that Piaget regarded as emerging gradually from prolonged interaction with the world. Together, these forces drew attention away from the critical age of 6 or 7 and away from the insights that Piaget thought would emerge in children at about that age. The widening gap between the classic Piagetian vision and more contemporary standpoints was further advanced through the rise of various information-processing theories that considered developmental tasks in terms of their complexity and informational load on memory rather than their logical properties. More recently, a further shift and change of direction has arisen from *Executive Functioning* (EF) approaches to developmental cognition, originating as much from neuropsychology as from classical developmental psychology, in which memory and processing factors are considered in terms of functional changes in the prefrontal cortex (PFC) of the brain.

With this shift toward the study of growth in psychological rather than logical processes came the substitution of classic tests of logic with tests of WM, planning, and response inhibition (RI). As we describe below, these paradigmatic and theoretical shifts have led to considerable new insights into cognitive growth and its neuropsychological underpinnings. The premise of this monograph, however, is that there has been a corresponding neglect of what were once considered as the crucial discoveries made by children about the age of 6–7 that forever alter how they view and deal with the physical world.

The Structure of This Chapter

The broad topic we address in this monograph is children's developing understanding of discrete properties of sets of multiple items. We focus in particular on sets of items that differ only by size. We describe how children appear to move from a weak and indeterminate concept of a set, to one which is fully unitized, and thereby permits principled ordering and the identification of ordinal values. First, we describe the key tasks that were used in developing Piaget's thesis. Second, we consider how Piaget's findings bear on the issue of discontinuity in development and, finally, how he dealt with this in terms of his transactionalist account. With the topic raised and investigated by him thus described in as objective a manner as possible, we move on to the various theories, approaches, and critiques that, we shall argue, left this issue isolated and unresolved.

Ordinality, Cardinality, and One-to-One Correspondence: The Tasks

The first principle that we might expect to appear during early childhood is cardinality. After all, children are encouraged to count almost as soon as they can speak. For Piaget, however, a cardinal value should be more than a verbal label and should come to be understood as enduring property of an item within a set, indifferent to how it was counted or the perceptual layout on which the count was performed. Accordingly, it should be understood as equivalent to the same cardinal value of an item from another set, again irrespective of perceptual factors. This criterion of understanding was the rationale for Piaget's famous number conservation tasks (Piaget, 1952a; Piaget & Szeminska, 1941) where, typically, the child is presented with two sets of objects with same number in each, displayed in horizontal rows one above the other. The test involves asking children how many there are in a set before and after one or other set is transformed by spacing the items out. The standard test question put to the child each time is, "are there more here (pointing) or more here (second set) or are they just the same?" The classic finding (and we shall come later to refutations) was that children are likely to give a different answer after the transformation. Piaget's explanation for this, and other tasks, usually involves making distinctions across three main stages, the last being operational. His stage analysis took account not only the accuracy of the answer, but also the child's behaviors and verbal justifications for their judgments. At stage I, children consider the array as a global whole, that is, they do not decompose the array into units and the answer is thus likely to be based on overall appearance (i.e., length of the array). By stage II, children (of five or six) will observe the one-to-one correspondence of items across sets if they are arranged in direct alignment with one another, but will then use the global comparison to decide "which has more" if the sets are moved out of alignment (though the conflict between their two answers may be obvious to them). By stage III (the operational child), the correct answer will be given despite the transformation. It appeared to Piaget that the conflict the child experiences at the preceding stage becomes resolved through the realization that one-to-one correspondence can by reinstated by moving the items up again. The act of decomposing the set to achieve correspondence thus appears to "free" the child from the influence of the immediate perceptual appearance.

Piaget was concerned to show that the stages describing the development of cardinality can also explain how children come to understand ordinality at around the same age. Using sets composed of objects that could be ordered along a dimension of change such as size, Piaget (1952a) constructed ordinal and serial correspondence tasks. There are several versions of these tasks, but one will serve as an example here. The child is given a set of 10 dolls of different sizes and 10 toy walking sticks also of different sizes and these are put in order and in spatial correspondence until the child understands that each stick can be paired with a particular doll. One of the test questions involves putting the sets out of order and asking the child to find the stick that "goes with" a particular doll. Once again, there are three stages leading to the development of the correct answer. The solution requires spontaneous *seriation* or ordering of the objects, and at the first stage children fail altogether. At the second stage, there may be trial-and-error ordering of objects, but when the sets are disarranged, they are unable to identify the ordinal value of a stick (e.g., "third biggest") in order to match it to the corresponding doll. This is accompanied by a failure to correctly insert a single missing item into a disordered set, again suggesting a failure to separate the ordinal value from the global whole. At the third stage, the act of spontaneously seriating the sticks allows one-to-one correspondence tasks to be solved. As with conservation, this is purportedly because children apprehend that their own actions on the series—such as "setting out the series in either direction" (p. 155) are reversible. This reversibility in action leads to understanding the true nature of an ordinal relation—namely, that an asymmetric relation is simultaneously expressible in a "bigger than" ($A > B$) and a "smaller than" ($B < A$) direction. This, in turn, enables the identification of a specific unique value and can be used in the correspondence tasks to find the correct stick. The very concept of a unit emerges from these discoveries, according to Piaget.

Seriation and Serial Understanding

Piaget's tasks illuminated the way in which three interconnected aspects of operational understanding (ordinality, cardinality, and unitization) might become manifest. Although tested in conservation and one-to-one correspondence tasks, an important emergent behavior enabling the latter was *spontaneous seriation,* that is, methodically ordering a given set of items in order of size. Seriation was also tested by Piaget in its own right as a standalone task in which the children simply have to select sticks from a disordered array and put them in order of size. The questions it raised are the subject of this monograph. Because we shall be introducing a paradigm in Chapter II that is serial in nature but different from the classic seriation task, we define next the difference between Piaget's size seriation tasks and the concept of size sequencing (which are not entirely synonymous).

Size Seriation

We use the term size seriation to refer to Piaget's physical construction task in which a single set of objects differing in length or size have to be selected from a pool of jumbled elements and placed in a monotonic order (from biggest to smallest or the reverse). The task typically involves the presentation of a set of eight to ten 2-dimensional objects such as drawn rectangles (Piaget, 1952a) or 3-dimensional objects such as vertical rods (Piaget & Szeminska, 1941) and the child is invited to copy an already ordered set or simply asked to "make a staircase" with them. Operational success at around the age of 7 is the ability to do this immediately and without trial and error. A corollary task is based on asking the child to insert a missing stick into its correct place, which, according to Piaget, demonstrates that the child understands that the item is simultaneously greater/less than all the prior items in the ordering, and less/greater than all the subsequent ones (though it could in principle be based on local comparisons with the immediately adjacent sticks).

Size Sequencing

Size sequencing refers to the ability to adhere to a serial monotonic order even if there is no requirement to select or place objects. The size-ordering task testing this ability will be described in Chapter II and we, therefore, return to it in greater detail in due course. In the meantime, we note that selecting objects in a principled order of size does not have to be demonstrated by a select-and-place task.

We now turn to why the area in question still deserves close consideration in its own right and not simply as a facet of a historical theory.

Discontinuity: An Enduring Issue

Later in this chapter, we describe the ways in which Piaget's stage-based account came to be refuted. But before doing so, it is important to try to separate his explanation from the behavioral phenomena he so painstakingly documented. Considered as objectively as possible, the achievements of the operational child at around the age of 7 can indeed be viewed as representing an important discontinuity in cognitive development. Whether it is to understand the additive and subtractive relations in numbered sets (e.g., that $3 + 2 = 5$ and $5 - 3 = 2$); the cardinal and ordinal equivalence of items with same unit value; or the necessity of assigning a unique unit position to each item when ordering multiple items; the power of being able to unitize, quantify, and spontaneously order discrete sets cannot be underestimated.

The importance of self-discovery in the domain of number can be overlooked given the large amount of cultural exposure to the count alphabet from the earliest years to the symbolic manipulation of numbers as provided in elementary math instruction. But when it comes to their practical application in human activities like sharing, quantifying, and measuring, no amount of third-party tuition alone would be effective without deep understanding by the individual of certain properties of discrete sets, and in particular the uniqueness of a unit. In the absence of that understanding, myriad other routes to the judgment of quantity and amount are clearly available as evinced by the accounts of preoperational behavior by Piaget, and also the subsequent replications and variations of his methods (Sigel & Hooper, 1968). We know, for example, that there can be counting without principled application to the members in a set (Fuson, 1988), and that conservation tasks may invite answers based on perceptual variables only (Bryant, 1972; Gollin, Moody, & Schadler, 1974). Conservation and one-to-one correspondence tests can display a failure to use counting as an appropriate adjudicator of equivalence or difference (Cowan & Daniels, 1989; Desrochers, 2008). Seriation tasks can be met with anything from a rough sorting of items into big and little ones to incorrect or partially ordered sets (Kingma, 1983a, 1983b; Kingma, 1984a, 1984b; Tomlinson-Keasey, Eisert, Kahle, Hardy-Brown, & Keasey, 1979).

There is, in short, one unique correct solution that exists for any and all evaluations of discrete sets in these tasks and that is the deployment of a system for unitizing, involving an implicit ascending monotonic structure starting from one. In that sense, there is no transitional skill; attempts through trial and error, approximations or partial solutions are essentially failures to solve the problem presented and imply a lack of apprehension of the properties of the set as a whole. This does not deny their relevance to what emerges later. Trial and error, partial or provoked success, sometimes procured simply by dint of reducing the size of set to be considered (Gelman, 1972; Lawson, Baron, & Siegel, 1974) surely provide the basic platforms for the unique discovered solutions on which all children appear to converge. But they are, in principle, distinct from what emerges later—not only by chronology but also by their variability and heterogeneity. This variability of solution across ages, individuals, and contexts is a property that, by definition, does not apply once precise enumeration or principled monotonicity of ordering is available. Once discovered, the solution should be immediately and spontaneously applicable without learning or corrective feedback from the task environment. Its application should be impervious to perceptual transformation and set expansion. It is a one-size-fits-all solution to the evaluation of discrete sets of similar and dissimilar elements. It is in this objective sense that the robust success displayed in number conservation, one-to-one correspondence and size seriation after the age of 6 can be described as qualitatively different from what has gone before. And in the sense that the system properties are not to be found in any of the precursor solutions, it can be seen to be a candidate case of developmental discontinuity. As Liben (2008a) puts it with regard to much earlier discontinuities:

> *What emerges is thus different from the sum of the parts; integration yields novel systemic properties that are not characteristic of any of the parts. There is still continuity insofar as later forms subsume earlier forms, but there is also discontinuity or qualitative change.* (p. 1603)

Making the case for a behavioral discontinuity, however, does not directly translate into a case for discontinuity in mental operations or processes. We consider this next in relation to Piaget's theory.

Piaget's Transactionalism as an Explanation of Change

As many have argued, the issue of whether the stages of cognitive development are continuous or discontinuous can sometimes become a question of semantics, and are sometimes even argued to be both (Fischer, 1980). However, it is the drivers of change that require clarification in these terms, whether it is to help inform the neuropsychology of the developing brain (Rubenstein & Rakic, 2013), or to specify the experiential factors that may promote new expertise (Kolb, 1983). We have indicated earlier why the correct solutions of the operational child seem discontinuous with earlier ones. The deeper, causal question is whether the mental processes underwriting later solutions are themselves substantively and distinctively different from those underwriting earlier ones. With a new solution now available, an earlier (more intuitive) one would become a choice, rather than the only default option. For example, adults may often engage in eye-balling the extent or density of a collection to obtain a number estimate, when precision isn't required. This is not to deny the discontinuity but rather to acknowledge that diverse processes can coexist in the developed mind (Minsky, 1988).

For Piaget, the major cognitive shifts that change the child forever are described in his main *phase* transitions, which themselves come about through transactions with the physical world. As all scholars of Piaget are aware, the key transactionalist concept for Piaget is *interiorization of action*. This explains how logical structures are constructed by the human mind through a progressive abstraction and symbolization of knowledge that is first experienced at the sensorimotor level. The term *constructivism* is therefore sometimes used to capture the functionalist aspects of his theory regarding the passage from practical to symbolic and then logical understanding. However, there are two distinct ways in which Piaget expounded his functionalist position, depending on the era in which he was writing and the domain of development he was addressing.

Functionalism and Adaptation

A great deal of Piaget's work was influenced by his background in zoology and an interest in self-regularizing mechanisms such as ingestion and thermoregulation. He introduced these notions into his accounts of sensorimotor development (covering roughly the first 2 years of life) in particular, viewing every stage as a new form of adaptation to the environment, each plateauing for a while—a state of *equilibrium* until the cognitive system was ready to take in more information. These he explained in great detail in relation to the growing sensorimotor capabilities of infants (see e.g., Piaget, 1952b). The term he used to describe these adaptations were *schemata,* which refers to the stable set of behaviors for acting on the world. A behavioral schema typically involves both *assimilation*, referring to the incorporation of new aspects of the external environment (such as extending a visual tracking behavior beyond the boundary edge of a cot), and *accommodation*, which refers to how particular constraints from the environment can shape the particular schema, such as the manual adjustments made when gripping objects of a particular size or weight. A constant interplay between assimilation and accommodation builds a stable schema for a particular behavior. One highly researched area that illustrates these ideas was Piaget's account of how infants move from simple visual tracking behaviors to searching for hidden objects under visible, and then invisible displacement. This became known as the development of the object concept (Bower, Broughton, & Moore, 1971). A succession of substages was used by Piaget to describe how simple equilibrated schemata become gradually more elaborate and thus how knowledge of the physical world grows as a consequence. In the case of the object concept, schemata for following a moving object with eyes only will gradually incorporate reaching and grasping behaviors that ultimately extend to reaching and lifting an occluder from a covered object. For Piaget's constructivist explanation, it is important to understand that knowledge is one and the same with the current state of the infant's ability to behave in the world during the sensorimotor phase. Thus, when an infant is eventually able to retrieve a hidden object, only then can it be said to have the concept of an object that endures when it is out of sight.

This form of transactionalism leads directly to the first phase transition between the sensorimotor and preoperational phases, which is now marked by symbolic representation of earlier adaptations, and knowledge is no longer strictly synonymous with sensorimotor behavior. Thus, imitation and play recapitulate the actual "operations" that the child had previously enacted with real objects, such as moving items from A to B, opening and closing containers, and so on (Piaget, 1972). Assimilation can take place at this symbolic level such as using an object to represent something else during pretend play. Accommodation would arise when the real world directly shapes a behavior, as in imitating actions such as hair-brushing. According to Piaget, these new behaviors represent the interiorization of action and the first stage of symbolic thought, but he is at pains to point out that it is not "simply a translation but a restructuring with a lag which takes considerable time" (p. 18).

Preoperational Development

The "considerable time" alluded to by Piaget in the quotation above covers the preoperational period from around 2–6 or 7 years, leading to the set of mathematical discoveries we are considering here. But in this area Piaget draws on a different (and earlier) concept of transaction with the world and it is notably different from the functionalism of assimilation, accommodation, and equilibrium—which are not mentioned at all, for example in *The Child's Conception of Number* (Piaget, 1952a). First of all, certain aspects of the physical environment involved in early attempts at conservation, seriation, and correspondence tasks are not incorporated in what emerges later in the sense of a more extended or elaborate schema. The use of spatial layout (length or density) as a determiner of numerosity is precisely **not** part of a later resolution based on operational understanding. Instead, this long period of preoperational thought is described by Piaget almost as a contest between what the child witnesses in his exchanges with world and how those exchanges eventually come to be symbolized. This does not mean Piaget failed to use a transactionalist explanation at this stage in his writings, however. He does indeed describe in clear action-based terms how "operations" on objects can become the logical operations describing discrete sets. However, the properties of the world that the child discovers during the preoperational phase are not about how objects behave in space and time, but how they are logically interconnected.

From Concrete to Formal Operations

The end-state of the preoperational period is the acquisition of simple reversible structures that are represented symbolically. But the defining property of the child's operational understanding at around the age of 7 is that these structures are not capable of being represented in the abstract, that is, in the absence of the concrete situations giving rise to them, and this is why Piaget describes the period from around 7–11 years as *concrete operational*. The next main phase transition, leading to the end-state of cognitive development and starting at around 11 years is the fully symbolic representation of necessary logical and reversible interconnections at the *formal operational* stage. Here Piaget invoked the concept of a set of logical *mother structures* devised by the influential Bourbaki group of mathematicians of the 1930s to describe necessary aspects of advanced logical and mathematical thinking. These emerge, according to Piaget in the form of a closed network of logically reversible operations composed of (relational) inversion, negation, reciprocity and correlation. In total, this comprises a set of commutative logical relations that permit formal hypothetico-deductive reasoning across the domains of space, time, causality and mathematics. This description dominated Piaget's accounts of development in later childhood and adolescence (Inhelder & Piaget, 1958). A further symbolic advance was thought to be necessary for formal operational thought and was sometimes referred to as *reflective abstraction*—a controversial and sometimes inconsistent concept (Chapman, 1988) that Piaget used to explain how operations at any level of functioning become restructured at a new more abstract level of thought. We reconsider the applicability of this concept in Chapter V.

Taking the latter two Piagetian phases together, then, there is an important discontinuity at around the age of 7 on which all subsequent development seems to depend. That is, that operations are not just reversible actions performed on objects but are also become a form of internalized knowledge. The relationship between logical thought and the contributing behaviors can be controversial, and cogent arguments can be made that structures are only a formalism, and should not be thought as literally causal to the child s behavior (Chapman, 1988; Liben, 1987). The explanation of how interiorized action ultimately relates to these formalisms are certainly more elusive (Sigel, 1968; Smith, 1993). Nevertheless, Piaget undoubtedly believed that structures arise from acting in the world and we attempt to describe this next.

Piaget's Transactionalist Explanation of Discrete Set Understanding

Returning to the 7-year-old, we can now ask, What are the possible transactions that lead the child over this first hurdle of becoming concrete operational? As with the sensorimotor discoveries, the key to this achievement is reversibility of action. As he puts it in *The Child's Conception of Number*, it "is the purpose of this book to prove [*that*] an operation is indeed a reversible action" (Piaget, 1952a, p. 81). The discoveries at the concrete operational stage are very much more advanced than, for example, understanding that an object that has been displaced from A to B can be returned to its starting point by reversing the act of displacement. Nevertheless, there is a strong similarity of concept across these two distinct phases. The connection between an action and the reversible operation deriving from it is perhaps most transparent when Piaget discusses number conservation. At the substage immediately prior to concrete operational success (stage II) 5/6-year-old children typically discover that opening and closing up the row into its original length is a fully reversible operation and will form their judgments based on noticing this. The reactions of children "who space out or close up the elements in order to restore the equivalences, are forerunners of the construction of true operations" (Piaget, 1952a, p. 81). The true operation in this context is a reversibility based on imagining a physical transformation in space. Its counterpart in size seriation is based on the reversibility of the ordering itself. Here reversibility applies to asymmetric relations such as greater or less than, such that, for example, if $A > B$ then $B < A$. The reversible actions leading to the operational transition at stage III (spontaneous seriation without error) is apparently based on seeing "the possibility of setting out the series in either direction" (Piaget, 1952a, p. 155).

It is relatively easy to grasp the necessary precursor behaviors that the child could employ even through random play and experimentation with the types of material in question that would lead to the discovery of reversibility. It is the transition from reversibility in action to reversibility of thought that many find elusive (Voneche & Vidal, 1985, as cited in Smith, 1993, p. 42). How exactly does it happen? The repeated assertion from Piaget is that it must involve a "freeing from perception." Prior to this:

> *cardinal correspondence is not lasting, and does not as yet entail permanent and necessary equivalence, firstly because it is not sufficiently dissociated from qualitative correspondence, and secondly because it is still dependent on perception. The same is true of ordination, which is not sufficiently differentiated from qualitative seriation, which is also still intuitive, that is the order is understood only in so far as the total series is actually perceived.* (Piaget, 1952a, p. 154)

Piaget's expression "freeing from perception" gives an insight into how he views the connection between perceptual activity and development. It certainly appears to him that perception is somehow an anathema to thought at least at the operative level. There is no question that Piaget employs a circular argument on this point. If children succeed on his operational tasks, they behave as if they understand that each transformation can be compensated by another and that they apparently include each perceptual situation in the "system of all the possible situations" (Piaget, 1952a, p. 83). When they haven't yet made that transition is because they are still tied to one or other feature of the immediate present (such as the length of a row of objects). Although perception and action are the bedrock for logical discovery, according to Piaget, it is clear that he sees the perceptual plane as an ultimate constraint on logical thinking because it ties the child to an immediate present rather than a set of imagined possibilities and represented relational coordinations. Because Piaget defines success a priori in terms of a coordinate logic that cannot exist in a directly perceivable form, it has to follow that direct perception must be abandoned as a solution to the task.

From Figurative to Operative Understanding

The freeing from perception found expression in Piagetian distinction between *figurative* and *operative* thought—a more generalized way of expressing his idea (just described) that children come to use the fact of a transformation as predominating over the immediate perceptual relation. It is almost impossible to understand Piaget's distinction without accepting that, in the case of operative thought, he is talking about some sort of simultaneous awareness of the different states in which an object could be perceived, that an item that is "bigger than" can also and at the same time be "smaller than." (This is why he includes single item insertion into the essentially unidirectional task of seriation.) What he often describes as lasting numerical correspondence, therefore, involves a form of representation that is inevitably *atemporal* in nature. In terms of the development of seriation, he describes it thus: "a growing coordination between successive actions which eventually overcomes the one-directionality inherent in a succession and takes the form of a shuttling from the present to the past which very soon begins to impinge on the future" (Inhelder & Piaget, 1964, p. 287). The final and complete transition from figurative to operative thought is not the act of performing the reversible actions, nor even being able to represent them symbolically—it is in understanding the actual objects in a new way because the *possibility* of such actions now predominates in how the objects are perceived.

> *Indeed, it should be well understood that an operation is not the representation of a transformation, it is in itself, an object transformation, but one that can be done symbolically. Thus an operation remains an action and is reduced neither to a figure or a symbol.* (Piaget, 1972, p. 76)

If this seems unnecessarily obscure, it should be noted that in describing the transition from operative from figurative thought, Piaget was not prepared to hand over the role of this new sort of abstraction to language. Piaget is notoriously dismissive as language being anything other than the verbal expression of something already understood at a deeper level. That is, the terms "bigger" and "smaller" will start to acquire the reversible character of the underlying thought process, but they do not causally convey that reversibility of thought in themselves. Insofar as Piaget mentions language at all in these contexts, a characteristic comment is: "we should be wrong to confine our search for the origins of these operations to the symbols and concepts of language" (Inhelder & Piaget, 1964, p. 282). He elaborated this point further in this statement:

> *In short, adequate verbal transmission of information relative to the operatory structures is assimilated only on the levels where these structures are elaborated on the basis of actions themselves or operations as interiorized actions, and if language favors this interiorization, it neither creates nor transmits ready-made these structures by an exclusively linguistic means.* (Piaget, 1972, p. 119)

Piaget concedes a much greater role for language and symbol-level representation in general at the formal operational stage. For the purposes of the area in question (the logic of discrete sets), what he means by a concrete operation is the mental representation of a reversible behavior performed on the objects in those sets.

Falsification of Piaget's Structuralist Approach

Despite it being presented as an empirically validated theory, a recurring question is whether Piaget's structuralism is open to falsification (Smith, 1993). The processes of internalization are private and not open to scrutiny, and changes in the way the child views the world cannot be directly verified. However cogently one argues for the causal influence arising from reversibility of action, the allegedly emerging structures are themselves *closed* (containing relations than can exhaustively map onto one another). It is, in fact, not at all clear how one might falsify the claim that the mathematical relations linking items in a unitized set are reversible, as this is a self-evident truth. The real question is the applicability of this truth to the processes of logical development. Precisely because Piaget avoids specifying operations in terms of serially ordered steps leading to a logical conclusion, it is hard to know, furthermore, how the idea of reversibility would translate into more contemporary information-processing terms. The rise of such approaches (which we describe in more detail below) originated from thinking of cognition as a time-ordered process of taking information in from the environment, storing and recalling it, and, where a solution is required, acting on it in a series of steps (Atkinson & Shiffrin, 1968). With new ideas regarding serial and step-wise processes, later to be instantiated in computer metaphors (Newell & Simon, 1972), the Piagetian account could be seen in retrospect to have raised a possible conflict between logical processes and logical structures. In-depth analyses of this problem and possible resolutions of the conflict can be found in Campbell and Bickhard (1986) and Smith (1993). Smith argues that preoperational solutions evince a form of modal (possible worlds) logic that is compatible with Piaget's structural descriptions of later logical solutions, whereas Campbell and Bickhard argue that the concepts of logical necessity arising from Piaget's account can be explained in terms of an information-processing approach that they describe as a *knowing levels* analysis.

Despite the concerns of others regarding structure versus processes, Piaget himself was clearly comfortable with the idea that the cognitive processes leading to operational discoveries can be appropriately described in structural terms. This apparent paradox was pursued in a series of conversations with Piaget (Bringuier, 1980). The excerpts here followed an assertion from Piaget regarding the necessity understood by an operational child regarding the transitive relationship $A = C$, if $A = B$ and $B = C$, based on the structure $A = B = C$. Piaget states, "Necessity is the criterion of the structure's closure, the achievement of a structure." Bringuier then asks: "does this mean there's a structure only when the child begins to do operations, for example?" Piaget replies, "Before operations—if you accept our definition of them as internalized actions—there are already structures of action" (all quotations, p. 41). A little later Bringuier observes, "listening to you I get the impression that the child suddenly changes intellectually, as if there are sudden mutations" (p. 45), to which Piaget replies: "no, the transformation is slow. What is sudden is the final comprehension when the structure is completed." He follows through with this telling phrase "and of course it presupposes a whole preliminary labor, underneath of which the child had no consciousness." If we map this back to the behaviors leading to the insights governing conservation and seriation success as described above, this labor would be, for example, setting out a series in both directions, or opening and closing up a row of counters. In emphasizing the reversibility inherent in these actions, even if it is not consciously grasped, Piaget gives a structural account of a functional process. Whether or not this is a coherent position from a philosophical point of view perhaps misses the point that Piaget's own description of the behavioral developments remains open to scrutiny. Should those same outcomes be explicable in some form of functional or processing terms without recourse to the structural concepts of reversibility, the explanatory usefulness of his structuralism itself has to be completely reconsidered. What is important is not to lose sight of the outcomes themselves as they may denote a significant turning point in the mental life of the child. However, a loss of interest and focus is unfortunately what appeared to happen in the case of discrete set size understanding.

The Outstanding Questions

There are many issues tied up in Piaget's view of the concrete operational understanding of discrete sets including the methodology behind the tasks, the database, and measurement, as well as the theoretical account and the possibility of alternative explanations. There are also larger metatheoretical issues such as the role of perception and action in a numeric task involving concrete entities, and whether being "freed from perception" is a valid concept in this context. More general still are issues regarding his position that key aspects of mathematical understanding arise directly from private individual experience and insight (as opposed to being imposed through enculturation). These issues form the basis from the account we offer here across the following four chapters. In the meantime, we must conclude that these reservations notwithstanding, a very strong case for some kind of discontinuity of function and process leading to new knowledge was made by Piaget in specific regard to the understanding of the properties of discrete sets. Next, we trace what followed Piaget's thinking, why it provoked refutation, and why the issue of stage-like shifts in the domain of discrete set understanding has all but disappeared from the literature.

The Dissolution of Interest

The Logic Rebuff

Piaget's attempt to explain not only the growth of knowledge but the very nature of knowledge itself through the private discoveries of the maturing child was famously called *genetic epistemology*. Its remit was to explain how the closed structures of logic and mathematics that we tend to consider as public and trans-individual could arise through private discovery by every individual and it was the inspiration that launched much of the study of developmental cognition in the 20th century. Some scholars were unconvinced of the logical consistency of Piaget's quest, however, and, as mentioned earlier, his attempt to combine a structurally defined end-state for cognitive growth with a functional explanation of how that could be reached. The very concept of a genetic epistemology was met with skepticism by some philosophers on the grounds that epistemology is the exclusive province of philosophy (Beilin, 1999). Rather than defend the introduction of logicism into his position, Piaget tended to defend the introduction of psychological constructivism into the area of logic:

> *We can formulate our problem in the following terms: by what means does the human mind go from a state of less sufficient knowledge to a state of higher knowledge? The decision of what is lower or less adequate knowledge, and what is higher knowledge, has, of course, formal and normative aspects. It is not up to psychologists to determine whether or not a certain state of knowledge is superior to another state. That decision is one for logicians or for specialists within a given realm of science. For instance, in the area of physics, it is up to physicists to decide whether or not a given theory shows some progress over another theory. Our problem, from the point of view of psychology and from the point of view of genetic epistemology, is to explain how the transition is made from a lower level of knowledge to a level that is judged to be higher. The nature of these transitions is a factual question.* (Piaget, 1968, p. 5)

Although apparently acquitting himself from the accusation of being logicist within his empirical approach, there is still no doubt that, for Piaget, it is only by recourse to concepts such as logical reversibility that he can explain how operational structures come into being. As described above in terms of the simultaneity of understanding asymmetric relations (such as A > B and B < C), this is essentially a passage from temporal to atemporal knowledge (Smith, 1993, p. 183). And so, the question of whether Piaget's genetic epistemology can be said to have succeeded depends not only validating the key causal interactions with the world, but also on accepting the premise that the transitions and end-states represent the acquisition of generally applicable reversible structures. There have been some strenuous efforts to validate this premise (e.g., Leiser & Gillierion, 1990; Lourenço & Machado, 1996) but they became increasingly rare. Others simply repudiated the need for a structural account altogether (Brainerd, 1978; Cohen, 1983). Some went on to use similar concepts (Halford, 1993) but based on rather different (predicate/argument) logic. Most simply dropped them for reasons to do with experimental fashion and trend as we review below. Yet the interest in when children became logical was still of natural interest to psychologists and educators (Brainerd, 1978; Donaldson, 1978; Halford, 1989). What survived rather better than the search for structure itself, therefore, was the study of logical reasoning and deduction, very much reprised later in research known as Theory of Mind (ToM) (Wimmer & Perner, 1983) and deontic (permission rules) reasoning (Harris & Nunez, 1996). Indeed, of all the tasks Piaget used to study the seriation of size relations, it was transitivity (to which we return later) that continued to produce considerable debate in terms of whether a mental deduction had taken place (see Breslow, 1981 for a review). In short, the

Piagetian tasks that endured the longest in subsequent research were those that seemed to supply direct evidence for active reasoning. A related contention—but one that was often disputed—was that children needed to justify their answers in Piagetian tasks in order to show that they had reasoned deductively (Chapman & McBride, 1992). For example, much of the later number conservation research was concerned to include the justifications given for declaring equality after transformation (Papalia & Hooper, 1971).

Seriation, on the other hand, was evidence of structure in the acting; in its spontaneous execution, immunity to set expansion or interval difference, and the ability to insert a new element into series. But the structure was presumed here, the mental deductions required not entirely clear, and there was no obvious verbal justification that could provide evidence that children were combining asymmetric relations. If any form of logical structuralism was to survive from Piaget's account, it seemed to require more direct evidence of logical reasoning, and it would seem that the absence of any such evidence in the case of size seriation claimed this task as a victim.

The Empirical Rebuff: Neo-Piagetianism and the Fracturing of the Program

Despite some in-principle objections to Piaget's approach, the wake of research endeavor inspired by his work has been vast and far-reaching. However, subsequent empirical enquiry and investigation gradually evolved over time in ways that lost sight of at least some of the original objectives of his program. It began with a concept of *early competence* that became associated with the neo-Piagetian approach of the 1970s. This was not so much refutation of Piaget's stages as a challenge to his ages, and recruited experimental evidence that the competences sought by Piaget can be shown to be expressible by younger children if the task and language used are reframed to be more in line with the child's everyday experiences (Donaldson, 1978; Greeno, Riley, & Gelman, 1984). Similarly, learning approaches (some of which derived from earlier behaviorist accounts) had also long been challenging the idea that stages of ability were inevitably age-dependent came about by experimenting with specific training aimed at accelerating development in specific domains (Braine, 1964; Sigel, Roeper, & Hooper, 1966). On the whole, this proved disappointing with regard to actually accelerating ages of achievement (Kuhn, 1992; Sigel, 1968). What it did achieve, however, was a challenge to Piaget's notion of a tight synchrony across related tasks (Kuhn, 1992). Although Lourenço and Machado (1996) defend Piaget's position from the view that such synchrony was an essential aspect of his theory, the challenge of domain generality was a pervasive feature of neo-Piagetian accounts (Karmiloff-Smith, 1992; Wellman & Inagaki, 1997).

A consequence of the early competence argument was that the empirical modifications and revisions generated by the neo-Piagetian tradition started to concentrate on a re-evaluation of the stages prior to concrete operations, rather than the success achieved by children at around that age. When children of the appropriate age have been included there is in fact a fairly consistent replication of group success by the age of 7 on tasks such as number conservation (Dodwell, 1968; Elkind, 1968) and seriation (Kingma, 1983a, 1984b; Little, 1972). But the more pertinent point is that the concern with earlier competences produced more and more studies using children only of preoperational age.

As replicability and modification of Piaget's tasks became a major preoccupation with developmentalists in the 1960s and 1970s and the growing view that Piaget's account was too domain-general, came a fracturing of the tasks that Piaget had devised essentially as connected packages. For example, results from number conservation, one-to-one correspondence, and seriation tasks were constantly cross-referenced in Piaget's account but became increasingly treated as stand-alone tests by other researchers. The distribution of research effort on these specific aspects of numeric knowledge became extremely uneven, with a large concentration on conservation tasks, rather less on ordinal and cardinal correspondence, and very few indeed that isolated single set seriation after the 1970s. Some of this can be traced to particular papers on early competence, which sometimes prompted decades of follow-up studies on a single task. Notable examples are the transitivity debate (we return to this briefly below and in more detail in the next chapter), conservation "accidents," in which children accepted the equivalence of two rows of objects if the spatial transformation was made by a clumsy teddy (Dockrell, Campbell, & Neilson, 1980; Eames, Shorrocks, & Tomlinson, 1990; McGarrigle & Donaldson, 1974), and class-inclusion success achieved by altering the form of the test question (McGarrigle, Grieve, & Hughes, 1978). A curious aspect was the fact that, despite the growing consensus that Piaget had been too domain-general in his approach, it became more commonplace to take a single example of apparent early success as evidence that children were logical after all (Bryant & Trabasso, 1971; Pears & Bryant, 1990). No such claim was made for early competence in size seriation, however, and although it is strongly linked in concept and in origin to transitivity, research on Piaget's single seriation task was not forthcoming.

The New Theoretical Challengers

While some were engaged in challenging the empirical story in certain specific regard, others were taking fresh stock of the developmental landscape as a whole. With a growing skepticism of Piaget's position came new accounts of development each of which added a new perspective and often a new research focus. Most radically, perhaps, was the rise of nativism in the 1990s, which took the early competence arguments to their logical extreme, with many investigators concluding that certain core competences were innate and could be shown to be expressible by young infants using suitable experimental methods. Although this led to a vibrant new emphasis on cognition in infancy, including number-related domains (Gelman, 1972; Spelke, 1976, 1979; Wynn, 2008; Wynn, Bloom, & Chiang, 2002), it was also responsible for a further decline in research on school-aged children and one only need to look at the disproportionate space given to the first 2 years of childhood in psychology textbooks around that time to see this shift in emphasis.

With this change in focus, the very issue of discontinuity altered. As originally conceived by Piaget the measured discontinuity was based on response measures that were themselves unchanging. These typically involved a verbally expressed judgment of greater or less, of equivalence or difference, the selection by word or action of an individual item, or the actions of ordering and rearrangement of physical items. Interest in this type of discontinuity was overtaken when focused on the transition between early infancy and toddlerhood, the behavioral measures themselves became the main subject of debate. Argument and discussion commonly revolved around whether a behavior measured in terms of habituation, looking time or some other nonverbal index is continuous with a related competence expressed months or even years later in terms of a more overt motor or verbal response (Kagan, 2008). Although highly relevant to the issue of the innate origins of developing competences (Liben, 2008b), this does not directly address the issue we are contending with here. Specifically, in this monograph, we start with the contention that children undergo a radical alteration in response to exactly the same task requirements in the early school years. What remains to be clarified is the most accurate and psychologically plausible way to account for these alterations in behavior and judgment, should they prove to be validated. There were many scholars waiting to offer such clarification in the form of other alternatives to Piaget's account as we now review. We should note in advance, however, that although most of the approaches we consider next have implications for the development of numeric understanding, few authors directly raise the issue of seriation and related skills (for foregoing reasons).

Alternative Levels Approaches

Many new theories maintained the concept of stage-like levels either within or across domains but interpreted these in new ways. If we were to identify a running theme it would be that progress was considered as moving from limited to more extensive abilities rather than from failure to success. If Piaget described his "pre-" stages in terms of how children were failing to meet operational criteria, these new approaches were more likely now to stress what could be achieved at each growth stage (with corresponding constraints). One approach was to consider the degree of cognitive abstraction achieved at each level; another was the level of complexity. In the first camp, a theory often described as a domain-specific version of Piaget's account came from Karmiloff-Smith (1992) in her theory of Representational Redescription (R-R). Drawing on the concept of increasing representational abstraction Karmiloff-Smith's account reprises a theme long associated with developmental theory (Bruner, 1964; Kendler, 1975; Werner, 1948), which is that knowledge can be cast into various levels of understanding. In Karmiloff-Smith's case, the core idea is that children start to redescribe early successful but unconscious representational format, that she calls implicit, into new more explicit formats that can be accessible to consciousness. In her chapter "The Child as Mathematician," Karmiloff-Smith (1992) considers number conservation and seeks to reconcile its apparently late development with earlier competence in number understanding and counting (Antell & Keating, 1983; Gelman, 1972). In R-R theory, a toddler who can count but doesn't realize that the count is a stable description of the numerosity of the set is at the following implicit stage: "the toddler… is unable to focus on the individual component of the counting procedure that yields the array's cardinality" (p. 104). The toddler can count when asked how many objects are in the array but "the knowledge embedded in the procedure is not yet manipulable as separate components" (p. 104). This is called R-R, level 1. At the next level, the procedures become redescribed and Karmiloff-Smith goes onto argue that the first thing to become more accessible will be the end parts of the procedures but "ultimately all the component parts of the counting sequence … become accessible to cognitive manipulation" (p. 104). To explain the process of redescription, Karmiloff-Smith appeals to internal changes rather than further exchanges with which the child is interacting: "Clearly nothing in the external environment will directly inform the child" (p. 109).

Although R-R theory is grounded in procedures, this abstracted knowledge is not in itself procedural but is an internalized, redescribed and more compressed version which at its most explicit level (E-3) can become accessible to verbal report. The theory takes a limited innate procedure and abstracts it into a strong representation without further exchange with the environment. It, therefore, serves as a notable transition from Piagetian transactionalism both in the move to a more nativist-grounded approach, but also in the emphasis on the causal role of language in allowing new knowledge to become explicitly represented.

Other and more radical departures followed, largely by introducing more contemporary information-processing concepts (as well as cognitive abstraction) to explain the development of higher cognition. These utilized the idea of levels of cognitive complexity achievable at different ages. An example is that of Halford and colleagues (Halford, 1993; Halford, Andrews, Wilson, & Phillips, 2012). Halford, Wilson, and Phillips (1998) offer a global theory of developmental stages and the growth from nonstructured, to functionally structured, and then to symbolically structured representations. An elaboration of how the final stage is reached is based on the idea of a complexity ranking proceeding from level 1 to level 6. This analysis was applied in detail in the domains of transitivity, class inclusion, hierarchical classification, cardinality, sentence comprehension, and hypothesis testing. Drawing on the ability to coordinate asymmetric relations into a larger structure, transitivity is the most relevant here and we consider it now.

Transitive reasoning, arguably one of the most researched cognitive tests for school-aged children, was first devised by Piaget from an IQ test designed by Binet (Piaget, 1928). In Piaget's version children were given the linguistic relational premises, "Edith is fairer than Suzanne; Edith is darker than Lilli" and asked, "who is the fairest/darkest?" Being of a purely linguistic nature, this version of the task is really a test of formal operational reasoning and was solved by Piaget's participants at around the age of 11. Piaget himself converted the task into something with a more concrete foundation (Piaget & Szeminska, 1941). The test materials were colored sticks with barely perceptual differences. The child was shown that stick A was longer than stick B, when placed next to it, and, likewise, that stick B was longer than stick C, also when paired up directly. A correct answer to the question regarding the relative lengths of sticks A and C could be deduced by an inference of the form: $A > B$; $B > C$; therefore $A > C$ and was reliably given by children of around 7 and older. Replications and variations of this particular test of concrete operationality resulted in numerous claims of earlier success, however, and it became a highly researched area in its own right. We shall return to the reasons why decades of research on transitive inference has led to inconclusive results regarding the age of acquisition in Chapter II. But of interest here, in terms of a theoretical alternative to Piaget, is how Halford and colleagues defined stages of growth according to a structural complexity metric that they developed. Using something akin to predicate structure to define the complexity of a task, transitive reasoning was seen as an example of integrating binary relations into a more complex ternary structure. This is similar to Piaget's idea of coordinating asymmetric relations $A > B$ and $B > C$ into $A > B > C$, except that Halford describes the relations in the form of what is known as *predicate logic* where the relationship between two terms is denoted with the letter R and the two terms or *arguments* as algebraic characters. Thus any relationship between A and B (including the relationship greater than) would be written as aRb. Halford described this form of binary relational understanding as Rank 3 level. To be able to combine a binary relation with another such bRc would require a ternary relation aRbRc; that is, Rank 4 understanding on his model, representing greater structural complexity than Rank 3. The similarity with Piaget's structural account is that it is only by considering both premises simultaneously that each element A, B, and C can be assigned a unique place.

What produces the advance to Rank 4 understanding? This is where the approach of Halford and colleagues depart most critically from the Piagetian one. For Piaget the more complex structure is arrived at by insight based on practical experience. Piaget was notoriously unwilling to concede a strong role for maturation as an independent factor in advancing the child toward operational insight. Halford and colleagues show both by argument and by formal modeling that greater structural complexity requires greater processing capacity. As the latter increases with age, maturation of the brain has to be acknowledged and indeed Halford et al. (2012) even offer an age (5 years) at which ternary relations might be achievable. Halford's approach is not simply deferring explanation to an unobservable maturation of processing capacity, however. Halford (1993) proposed that children can be helped toward new levels of complexity if they can make an analogy with a structure that is familiar to them. For example, the spatial relationship between above, middle and below provides a structural analogy for the size relationship between A, B, and C in the transitivity problem.

Halford's theory illustrates the new considerations that were redefining the goal of developmental cognition. First developmental levels were now becoming described in terms of task (e.g., structural complexity) on the one hand and agent (child's processing capacity) on the other, rather than the dynamic transaction between child and task where each coevolves through physical experience. Although this raised the criticism of failing to show how the system self-modifies (Klahr, 1992), it presaged a growing recognition that the maturational resources of the child could no longer be ignored and would have to be defined in terms of some kind of WM and/or LTM capacity. However, the ensuing search for new tasks that could be formally defined in terms of levels of complexity (Zelazo, Muller, Frye, & Marcovitch, 2003) drew attention even further from the classic set of tasks provided by Piaget and the Genevan school.

From Stages to Waves

A further departure from Piaget's stage model was the blurring of the hard edges around phase and stage transitions giving rise to approaches that saw them as more like overlapping waves of growth. Siegler, Strauss, and Levin (1981) expounded this in terms of a scale of (four) types of rule that fitted the preoperational and operational responses of children to classic Piagetian tasks. Although similar to the complexity analysis of Halford and colleagues, Siegler's point was that lower and higher-level rules could operate within the same child and that variability across and within the domains was a more normative feature of growth than sudden age-related shifts. Chen and Siegler (2000) claimed that such an approach could detect broad similarities across stages of toddlerhood in terms of common strategic learning and thus resolve the issue of discontinuous stages altogether.

The concepts of complexity with variability was also central to the theory of Fischer (1980) who, like Siegler, developed an entirely new developmental theory, eschewing the classic Piagetian terminology for stages and substages and replacing it with the concept of *skill*. Fischer introduced a definitive hierarchy of skills (from level to level) that could apply across the total landscape of sensorimotor, social and cognitive development, where each new skill was more complex than the previous one(s) from which it was built. Fischer, however, retained the importance of direct experience in promoting change. He argued this with regard to the understanding of conservation, where individual skills, such as predicting a change in width and also in length when an object is manipulated, can become coordinated through practice resulting in a correct solution to the conservation task (Fischer, 1980).

Despite framing developmental advance in very different terms from Piaget, Fischer's approach conserved Piaget's concern with the small processes of change occurring as the child interacts with the environment, and Bidell and Fischer (1992) called for measurements that are more sensitive to change:

> *The use of a scale sensitive to small developmental steps helps to move research in cognitive development beyond the dilemma created by stage theory by reframing the debate in terms of processes instead of categories. Instead of asking whether or not children have "really" reached the concrete operational stage … researchers can ask more analytic questions: What are the particular sequences of reorganizations children go through in this domain? How do children move from one step to another?* (p. 66)

The skills approach of Fischer reprised the Piagetian position that practical knowledge leads to convergence on a single unique solution at least within certain domains. However, the promising idea that that cognitive advance would be better understood by considering specific processes in context became swamped by the growing idea that it resided (mainly) in the general and maturing capacity to process information, as we now review.

Maturational Approaches

As memory research became one of the most rapidly expanding areas of cognitive psychology through the pioneering work of Baddeley and colleagues (Baddeley, 1981, 1984; Hitch & Baddeley, 1976), and as the computer metaphor began to dominate cognitive science (Newell & Simon, 1972), it was inevitable that memory capacity and processing efficiency would become a predominant consideration in understanding cognitive growth. Much of this new movement was founded on the theory and investigations of Pascual-Leone who posited a linear increase in memory capacity, which he called M space, from around 5–11 years accounting for transitions from preoperational to concrete operational stages (Pascual-Leone, 1970, 1976). Adopting such an approach, Chapman and Lindenberger (1989) for example, used the idea of fixed units of attentional capacity that they call the "organismic conditions" to provide necessary if not sufficient conditions for solving concrete operational tasks such as transitivity of length.

We have already noted this influence on the work of Halford. Another example of how this was further developed is the highly influential position of Case (see e.g., Case, 1985). As detailed in a long and ambitious report, Okamoto and Case (1996) reprise the Piagetian notion of a general coordinated core to cognitive growth that they call Central Conceptual Structures (CCS). Including Piagetian domains of enquiry (space and number) CCS also extends to social interaction. They describe its relevance to scientific reasoning such as the balance beam task, the understanding of cultural artifacts and tests of computation.

Once again, however, a shift of focus toward the external demands of the task and not just the internal state of the child's computing power was shaping developmental cognition. Like Halford and colleagues, and also Chapman and Lindenberger (1989), Okamato and Case predict ages and stages of passing tests according to a hierarchical model of difficulty. Of particular relevance here is the domain of number. A simple unidimensional level of number understanding would be the number line tests where children are asked which number comes before or after another, or which is the bigger of two numbers. A bi-dimensional level would be, for example, what number comes four numbers before 60, that is, where the child must use one number line to represent the position of two numbers and another to compute the difference between them. Case, Okamoto, Henderson, and McKeough (1993) simulated developmental levels where textual analysis arithmetic word problems were translated into a flow diagram. The difference in levels as simulated was based on the premise that children come to have "a more explicit representation of the problem that is involved (change, compare, combine)" (p. 43).

It is difficult to situate the understanding of discrete sets as envisioned by Piaget directly within the research program of Case and colleagues and not just because of the widely different theoretical approach. The domain of number in Case's account is exactly that—a facility with the base 10 system for numerals as culturally defined. Siegler asks in his commentary "does the ability to answer (such) questions reflect general quantitative reasoning capabilities? It might just reflect when certain knowledge and procedures are taught in school?" (Siegler, 1996, p. 270). By using tasks that reflected the culturally indoctrinated aspects of numbers rather than the more abstract structures of unitized order tested by Piaget could be why Okamoto and Case argue that "we do not see the structures themselves as being closed systems" (p. 290). In this regard, it is hard to place the ordinal and cardinal understanding implied by number conservation and size seriation. More crucially, however, the work of Case and colleagues illustrates how far the discontinuity theme of Piagetian psychology had dissembled in the new information-processing and maturational accounts. The move from incorrect to correct (and logically justified) responses to the same task that defined changes in the child were now overtaken by formal task analyses that fitted the maturing child to the type of task and degree of complexity they could understand or solve (see also Zelazo et al., 2003). Tasks that represented a hierarchy of in-principle difficulty were replacing tasks that were designed to reflect a changing solution within the child.

Other Relevant Research Areas

Numeracy Research

If Piaget's ideas about the logicomathematical properties of discrete sets had become diverted into quite different themes by the 1980s and 1990s, not so the attention to the development of number knowledge in and of itself. With its roots not only in developmental psychology but also in internal psychophysics and education, a huge research literature has addressed the issue of how children acquire a sense of number (see LeFevre, 2016, for a review).

Piaget viewed the understanding of number as very fluid and variable entity until the period of concrete operations and specifically until such time as cardination, ordination, and unitization were all equally part of that understanding: "number is organized, stage after stage, in close connection with the gradual elaboration of systems of inclusions … and systems of asymmetrical relations" (Piaget, 1952a, p. viii). The concept of number and the logical deployment of counting to enumerate, or establish equivalence, must wait for logicomathematical development in general. To what extent has numeracy research challenged that view?

There are several difficulties in answering this question. One is that it falls victim to the comment by Siegler (with regard to Okamoto and Case, 1996)—that to some extent what is being measured is what has been taught explicitly at school. Children are taught that numbers follow on an equal interval progression. They are taught the principle of cardination and it is reinforced in elementary arithmetic. Nevertheless, as most have observed, the evidence from mental number line research suggests that young children do not regard numbers as democratic members of a set where intervals are necessarily equivalent. Typically, these tasks require children to place a number on a scale; for example, from 1 to 100 or on an open-ended end scale from zero. Various techniques are used but usually involve giving children an actual physical line or scale where they mark the position of numbers or select points that they think represent them. There is fair agreement that children move from a logarithmic assessment where numbers become more compressed the larger they are, essentially rendering the represented interval between, say, 9 and 10 as considerably larger than the interval between, say, 99 and 100. The consensus seems to be that children move toward a more linear evaluation in the early school years (Berteletti, Lucangeli, & Zorzi, 2012; Siegler & Opfer, 2003), although some have challenged the generality of this effect (Asmuth, Morson, & Rips, 2018).

Mental number-line research is often cited as supporting the influential view of Dehaene (1997) as detailed in his book *The Number Sense*. Dehaene points to the considerable evidence that there are innate intuitions about relative amount in terms of discrete values like one, two, and three, as shown by the many studies on numeracy comparisons in young infants (Barth, Baron, Spelke, & Carey, 2009; Coubart, Izard, Spelke, Marie, & Streri, 2014; Lourenço & Longo, 2010, 2011). This is sometimes referred to as the *Approximate Number System* (ANS). However, although babies can detect smaller or greater numerosities, he believes that they are "unaware of the natural ordering of numbers" and speculates that elementary arithmetic would see "the detectors for 1, 2 and 3 light up in a reproducible order in their mind" (p. 63). Dehaene does not believe, however, that such precision is carried forward into the later apprehension of large numbers and cites extensive evidence that the human perceptual system uses spatial grouping and other properties to make approximations to greater or lesser amounts. A key argument in his position is that learning the cultural symbols for number never eliminates the inherent fuzziness in the basic intuitions about number:

> *language eases the computation and communication of precise numerical quantities. However, the availability of precise number notations does obliterate the continuous and approximate representation of quantities with which we are endowed.* (p.86)

Others have disputed Dehaene's essentially nativist and continuous view of number understanding. Research by Le Corre and Carey (2007) involved a variety of tasks requiring children to estimate relative and absolute numbers (of stickers). The authors report a change from early number "knowing" (of numbers up to four) and the acquisition of later number-related skills. They show that a shift occurs by the age of 5 when children have acquired the basic counting principle outlined by Gelman and Gallistel (1978): If a numeral "*n*" refers to cardinal value *n* and "*p*" immediately follows "*n*" in the count list, then "*p*" refers to $n + 1$. Although they see the acquisition of this principle as an important discontinuity, they also suggest that it is emergent from early number knowing based on what they call enriched parallel individuation:

> *The idea is that the child makes an analogy between two very different ordering relations: sequential order in the count list (e.g., "two" after "one" and "three" after "two"), and sets related by addition of a single individual* ($\{i_x\}$, $\{i_x\ i_y\}$, $\{i_x\ i_y\ i_z\}$)*.This analogy then supports the induction that each numeral refers to a set that can be put into 1–1 correspondence with a set of a given cardinality, with cardinalities individuated by additional individuals. It also supports the induction that for each numeral on the list that refers to a set of cardinality n, the next numeral on the list refers to a set with cardinality* $n + 1$. (p. 432)

This is an important claim regarding discontinuity, but it raises the question as to why children do not appear to apply this principle in situations such as conservation where amount is not being explicitly elicited. This question is particularly pertinent in that the critical shift we are concerned with here relates to relatively small numbers that are well within the counting range of most 5-year-olds and well differentiated in number line terms. If the representation of numbers below 10 are reasonably well differentiated/individuated why are they not applied spontaneously to number conservation, correspondence and seriation tasks?

A final concern is that much of the numeracy literature is within-domain, that is it looks at correlations and predictors between and among different aspects of number understanding, such as the relationship between knowing the sequential relations between numbers and the number line (Xu & LeFevre, 2016), or mapping number symbols onto their referents (Leibovich & Ansari, 2016).

Numeracy research, in short, does not contradict Piaget's claim that number is not fully understood or spontaneously implemented below the age of 7, but neither can it support it directly, largely due to the very strong disconnect between the two types of research. We know that in many cultures, children are invited to count, compare or enumerate at home, in the nursery and then at school where numerical abilities become subject to formal education and testing. The nature of much of this knowledge is directly transmitted by adults and can even be learned by rote and it will go on to be part of a formal math curriculum. Research dedicated to the understanding of number and numeracy throughout this period is bound to be to some extent compromised by the influence of such teaching. Research on the predictive value of variation in ANS on later number development (Halberda, Mazzocco, & Feigenson, 2008) has been seen to be equivocal for exactly his reason: "individual differences in ANS acuity might give rise to individual differences in math ability. Alternatively, individual differences in the quantity or quality of engagement in formal mathematics might increase ANS acuity" (p. 666).

By comparison, cultural exposure to the act of ordering per se, but in the absence of numbers (as in the Goldilocks story), is also sure to exist both in school and at home, and nurseries often have Montessori blocks or their equivalent as play materials. It seems, however, not to persist in most curricula as an explicitly taught skill in school, except for the teaching of measurement and calibration. These skills are themselves, however, number-related abilities and have been shown to reflect the understanding of the number-line (Cohen & Sarnecka, 2014). What remains to be clarified in any research are the causal relationships between growing numeracy skills and understanding the logicomathematical properties of small discrete sets that are not themselves composed of numbers.

EF Approaches

Perhaps the most apparently radical shift since Piagetian inspired research is the concept of EF (Diamond, 2013; Miyake & Shah, 1999; Pennington, Bennetto, McAleer, & Roberts, 1996; Zelazo et al., 2003). We say "apparently" because EF approaches continued to explore the role of WM in cognitive functioning and development as established by Pascual-Leone and Chase. But the origins of EF arose as much from neuropsychology (especially from studies with patients with head injuries) and clinical psychology as it did from experimental lab-based psychology. Whereas the concept of a *central executive*, or a general monitoring system, had arisen in traditional memory research to account for things like attentional control during memorizing tasks (Baddeley, 1992), EF has much wider origins (Borkowski & Burke, 1996). The overarching concept of EF is goal-oriented behavior. Based on the strong foundation that significant changes take place in frontal lobe functioning up to and including adolescence (Davidson, Amso, Anderson, & Diamond, 2006), neuropsychologists have identified components of goal-oriented behavior that are controlled by different parts of the PFC and that can be identified in dedicated tasks that are now available in computerized test batteries (Robbins et al., 1994). The separable components suggested by numerous brain imaging studies as well as by task analysis, include *WM*, *RI*, and *Planning*. Originally a great deal of EF research has considered these components as isolable. Planning is regularly tested by the class of "Towers" type of task, where the participant has to envisage how to displace a stack of three objects to another location using the most efficient sequence of moves (Welsh, 1991). WM is most frequently tested by classic serial recall (Pennington et al., 1996), and inhibition by the Wisconsin Card Sort Tasks (WCST) (Mullane & Corkum, 2007), which tests the ability to switch a sorting criterion from (e.g.) the size to the color of objects displayed on cards. Attention, however, is now often drawn to the fact that these components can overlap within a task (Shing, Lindenberger, Diamond, Li, & Davidson, 2010). For example, the WCST taps in into WM resources (Barendse et al., 2013; Kercood, Grskovicb, Bandac, & Begesked, 2014), as well the ability to inhibit inappropriate motor and prepotent responses (Geurts, Corbett, & Solomon, 2009; Ozonoff, 1995) in ways that are difficult to separate. Similarly, Ross, Hanouskova, Giarla, Calhoun, and Tucker (2007) have argued that planning in search tasks is actually best considered in terms of WM and weaknesses in both WM and RI and as these are usually the underlying factors marked out as responsible for poorly formed plans and we consider these in a little more detail.

Working Memory

The concept of WM in contemporary research has strong theoretical and experimental allegiance to the pioneering work of Baddeley and colleagues, which is to a large extent concerned with remembering content tested using, for example, list recall or digit span (Baddeley, 1981, 1984). The downstream effect of this within the EF approach is to test WM using such tests in order to see if it explains weaknesses in the performance of clinical groups on goal-directed tasks (Pennington & Ozonoff, 1996). The highly complex literature deriving from this tradition, however, has moved beyond the basic processes of rehearsal and retrieval associated with list learning and has embraced memory mechanisms from every angle. An integrated review of research approaches (Miyake & Shah, 1999) asked the contributors to address eight key questions covering everything from basic mechanisms and representations to more permanent long-term knowledge and consciousness. The widely ranging responses even within this one volume testifies to the complexity of this area. In terms of memory within a developmental perspective, the issue of "what changes" was (still) debated in terms of learning versus maturation, and what is actually meant by WM was far from consensual. As Miyake and Shah (1999) put it:

> *Although it is agreed that there is substantial individual or age-related differences in the amount of information one can keep track of simultaneously, the specific factors assumed to underlie the variation vary from proposal to proposal including the total amount of activation resources available to the system.* (p. 11)

The role of attentional mechanisms, capacity, processing, and speed all figured to a greater or lesser degree in different accounts, however, and it is no surprise that some studies have tried to separate these factors experimentally (Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005). All of these memory factors are likely to have a strong bearing on the development of logicomathematical skills as some have argued (Toll, Van der Ven, Kroesbergen, & Van Luit, 2011). However, applicability to our current topic is limited in the fact that the consideration of memory as a retrieval and maintenance mechanism is much more evident than thinking of it in direct relation to the acquisition of new cognitive skills. Exceptions to this tended to come from computational modeling approaches in the review by Miyake and Shah. Lovett, Reder, and Lebiere (1999) used ACT-R to show how WM interacts with LTM in achieving subgoals within algebra problems. Young and Lewis (1999) showed how memory within productions systems (the SOAR architecture) could deal with problem-solving and spatial navigation. Perhaps the most eloquent summary of the problem of defining memory followed a theoretical analysis proposing a solution called Interacting Cognitive Subsystems (Barnard, 1999) that deals with how propositions become represented from linguistic input. The author summarizes the general conclusions from most computational approaches to memory as follows:

> *The wider picture is not one of performance delimited by a simple list of specific capacities, but rather a picture in which the fundamental limitation lies in the capacity of the entire architecture to reconfigure dynamically and use all its representational and processing resources to best advantage.* (p. 327)

Understanding how the architecture of memory systems might be reconfigured in the way he suggests stands as the challenge to any EF approach to logicomathematical development if it is to embrace memory as part of its explanation of discontinuous change during development.

Response Inhibition

Interest in the ability to maintain goal-directed attention under a change of task requirements stemmed initially from the observation that patients with frontal lesions perform poorly on tasks requiring inhibitory control (Pennington & Ozonoff, 1996). Relevant to autistic symptomatology, RI has also been tested with both clinical and neurotypical groups in tasks such as the windows task (Russell, Mauthner, Sharpe, & Tidswell, 1991) or the go/no-go task (Simpson & Riggs, 2006). Although largely relevant to withholding a prepotent motor response established by repetition and practice, RI has also been shown to be applicable to higher cognition (Zelazo et al., 2003). Recent attention has been paid, in fact, to the question of stage-like change in relation to concrete operational tasks (Houdé & Borst, 2014, 2015; Houdé & Guichart, 2001). This work adopts the rule-based stance of Siegler, where change is seen to be a gradual replacement of one set of rules to another. Specifically, they point out that tasks such as number conservation require a clear shift from a useful heuristic (e.g., length equals number) to the deployment of an algorithm for correct solution. The concept of algorithm is applied here to what Piaget would call an operation, being an analytical strategy that "necessarily lead(s) to a correct (i.e., logical) solution in every situation" (Houdé & Borst, 2014).

Rather than see this shift as an automatic process of maturation or experience, the authors argue that it is the development of inhibitory control that accounts for it and, specifically, that the heuristic has to be inhibited in favor of the algorithm. Following on from a general argument along these lines by Dempster (1992), they cite evidence in favor of this based on functional magnetic resonance imaging (fMRI). Frontal activation related to success by schoolchildren on number conservation was correlated with an independent test of inhibition (the Stroop test). When comparing non-conserving 5/6-year-olds with conserving 10-year-olds Houde et al. (2011) also found greater involvement if the parietal regions related to inhibitory control in successful children and they conclude:

> *Enhanced prefrontal areas are needed to perform this Piaget task successfully (i.e., before and after objects are moved), as shown primarily by the activation of the bilateral inferior frontal gyri in successful children.* (p. 343)

Other studies by Borst and colleagues (Borst, Poirel, Pineau, Cassotti, & Houde, 2012) looked at the neuropsychological correlates of either increasing or decreasing the need for strategy inhibition using negative and intertask priming. Priming with a task inviting an incorrect strategy slowed or restricted task performance in children aged 9 years and older, whereas intertask priming on structurally related tasks (class inclusion and conservation) enhanced it.

The fMRI studies of Houde and colleagues make a strong case for the involvement of inhibition in Piagetian tasks and in that sense bring EF approaches in closer alignment with the classic literature, but it raises rather than answers questions about how the explanatory value of the inhibitory component of EF. What remains unclear is whether active suppression of an unsuccessful heuristic is the mechanism by which conservation success is acquired in the first place. On one interpretation, the algorithm would be available to younger children but suppressed by more salient strategies due to weaker inhibition. Yet there is no direct evidence that this is the case.

From this research, we must conclude that whereas RI that can affect the operation of an algorithm it does not explain how that algorithm came into being. To make that connection would require a deep operationalizing of the information processing and memorial requirements in the Piagetian task. We argue below that this can only be done by computational modeling.

The Enduring Question

For Piaget and for developmental researchers ever since, a pervasive and uniting endeavor has been the specification of what internal changes, reorganizations of a problem space, new insights, or learning experiences could effect changes in competence and shifts in understanding. For others the question has been whether such shifts are more apparent than real; a predictable outcome of linear development in information-processing power and memory capacity that opens up new problem-solving possibilities.

We have summarized some of the reasons why specific areas of enquiry relating to discontinuity in the numeric abilities studied by Piaget have dropped from contention. This can be broadly summarized as the eschewing of Piaget's logicism and replacing it with a variety of alternative psychological approaches, ranging from levels of complexity, maturational, skill, rule-based, and EF approaches all of which can overlap to a greater or lesser degree. The research paradigms, theories, models, and direct implications for brain development have been various and far-reaching. The core question that endures, however, is the one that Piaget famously encapsulated in his genetic epistemology. For Piaget, the knowledge structures were not just a property of the world waiting to be discovered—they are also constructions of the human mind and his quest was to describe the human investment in the physical world without which no such knowledge would exist. If something has been lost (or thrown out with the bathwater) is it not the understanding of this very particular transaction that every learning agent can participate in, and, if Piaget is right, not just to acquire knowledge but also in a sense to invent it? The enduring question is not whether Piaget's genetic epistemology was right or wrong in its structuralist and functionalist aspects, but rather whether in his methodical investigations of play and exploration with discrete objects he identified an important discontinuity that remains to be satisfactorily explained.

Size Sequencing and Ordinal Size Understanding: A Window on Change

As we have noted above, the solution to the above question is paradoxically not well served by considering number development. It is also not well served by judgment tasks that ultimately devolve to a binary decision (more/less same/different, etc.). If we are to sustain a contemporary approach to this, we need a goal-directed task that demands a behavioral response that is adequate to the phenomenon in question but not dependent on a cultural device. Ordering by a physical dimension such as size or weight offers this window. As we have reviewed, claims relating to (size) seriation and ordinal size understanding as first developed by Piaget have not so much suffered from refutation as from neglect, following the tidal wave of new theorizing. It is also possible that it has suffered from a mistrust of the perceptual domain—an issue that we take up in the next and final chapters. In short, an important and specific question still confronts developmental psychology, and that is how do children come to organize perceived relations according to the necessary rules governing discrete sets?

Size seriation remains our most transparent window on this issue. It is transparent in the sense that is inherently nonverbal, does not depend on verbal justification or report, reducing variability within and across age groups in terms of linguistic accessibility to the relations involved. It is transparent in that the criterion for success is explicit in action and not simply implicit in terms of a single judgment of more or less, same or different, or X is the biggest. It offers possibilities for manipulating information-processing variables that would affect WM such as set size. The corollary task of identifying a specific ordinal position through item insertion is another available and transparent measure and does not depend on the child being able to articulate the ordinal size in question. Although the application of counting is necessary for the latter, dealing with a set size such as 10 sticks should not be a task prerequisite beyond the capability of a school-aged child.

But the most transparent feature of seriation is the fact that it is founded upon an environmental given (perceptual relations) as available to the younger child as it is to the older. It does not require and is not confounded with, the cultural acquisition of number knowledge in the sense of what is explicitly taught regarding integers, number systems, and symbols at school. It is possible to be capable of seriating spontaneously without knowing that five is greater than four and so on. Conversely, the implications of being able to order any set of asymmetric relationships has obvious implications for how number knowing and the number line develop if we accept the prima facie case that ordinal understanding depends on the concept of a unit. This makes seriation a highly relevant test of the readiness of a child to really understand the properties of number. There are, however, certain methodological aspects of Piaget's task that suggests it is not as transparent as it might be regarding the core ability to sequence size relations and we shall take up this theme in the next chapter.

Modeling Size Seriation: A Window on Process

A recurring issue that has dogged all the developmental accounts and theories since Piaget's has been the extent to which there is a need to explain sudden transition. If continuous and linear change can effect apparent discrete changes in development by rapid acceleration, then the account of what is changing internally in the child is very different from one that assumes some new intervening level of knowledge. This distinction is less one of continuity versus discontinuity (as sudden success will always seem to be a discontinuous shift)—as the characterization of that shift either with or without emergence of a new intervening variable. This is a critical distinction for brain reorganization, and the transfer of knowledge across domains. Even if research effort into logicomathematical development had been sustained during subsequent decades following Piaget's work, it is extremely difficult, if not impossible to adjudicate on this issue from the perspective of empirical evidence alone.

The primary motive behind a computational modeling approach is to be able to operationalize what can and will change, given a learning agent, a specific learning environment, and to decide whether (or not) that change incorporates new knowledge leading to old strategies being replaced by new ones. A means of unearthing precise causal links in particular problem domains is if those domains are themselves represented artificially. In the real world, a learning environment embraces many schooled and cultural influences on cognitive development, over and above the private discoveries made by the child as to how the environment is structured. A second motive behind computational modeling, therefore, is to try to isolate the private learning experiences from directly schooled influences on cognitive growth. The standard solution is to build a computational model that represents the task space and essential aspects of a cognitive agent within the memory of a computer in which learning processes and representations are rendered explicit, unambiguous and unpolluted.

Production Systems, Connectionist, Bayesian and Dynamical Systems Approaches

Computational modeling in cognitive development comes from two distinct traditions in computer science, that of symbolic and subsymbolic information processing (Boden, 1996; Klahr, 1992; Miłkowski, 2013). Symbolic computation emphasized automated logical theorem-proving and more generally adult problem-solving in the 1950s, an approach which gave rise to production systems in the 1970s. A production system comprises a set of conditional rules within a knowledge base of the format condition-action (such as *if X then select Y, if X then avoid Y*). Facts are also contained in this knowledge base. These conditional rules are triggered by information inserted into WM, at which point an inference engine tries to match this information with the condition portions of the rules. If there is a match, the rule fires, the resulting action potentially triggering the condition portion of another rule, and so on in a cycle of inference. Production systems are well suited to representing human expertise (Boden, 1996; Yule, Fox, Glasspool, & Cooper, 2013) and are the representational basis of the two main cognitive architectures, ACT-R and SOAR, in use within cognitive modeling research (Anderson, 2007; Yule et al., 2013). They are less well suited to modeling representational change than connectionist systems. (See Chapter III for a further exposition of production systems within a seriation context.)

By contrast, subsymbolic computation emerged from a cybernetic tradition, which used differential and difference equations to model growth in complex systems (Holland & McFarland, 2001; Van Geert, 1994). Cybernetics gave rise to connectionist, dynamical, and embodied artificial life (simulated and robotic) systems that are now used to model cognitive development (Clark, 2008; Halford et al., 2012; McFarland & Bösser, 1993; Parisi & Schlesinger, 2002). Connectionist models are normally made of a mathematical structure termed a directed acyclic graph, in the format of a network of nodes joined by links, both of which can have real numbers assigned to them, the links being inhibitory or excitatory in nature. They produce outputs in the form of a vector of real numbers or integers in response to an input vector of real numbers or integers. Learning is represented by error minimization algorithms, such that responses to input vectors gradually approximate the target input–output vector pairs that serve as training examples that are presented to them. Development is represented by a principled increase of the number of hidden layer nodes in the graph, providing it with more power (Quinlan, 2003). Connectionist systems are thus well suited to modeling scenarios that involve development and learning. Due to their incorporation of two types of structural change (internode link value evolution and the increasing number of hidden layer nodes), they allow representational change across multiple timescales (Shultz, 2003). However, knowledge in connectionist models is distributed across many nodes in a graph, and not identifiable in the same way as a production rule, which makes interpretation difficult (see Chapter III for a further exposition of connectionist systems within a seriation context).

In terms of evaluating a computational model, comparison with a data set from the real world is necessary. These might include successful and unsuccessful actions made, and the time taken to make them, to be compared with the data a child produces in the laboratory (Mareschal & Thomas, 2007). The limitations of such data-matching exercises should be recognized, in that matching data from a simulated system X to a biological system Y can tell one at best that the mechanism producing the data in system X cannot be ruled out as a candidate mechanism for system Y; it does not necessarily mean that is the same mechanism (McClelland, 2009). Yet, the benefits of the modeling approach make this a testable Popperian hypothesis as opposed to guessing as to what may be going on inside an unknown biological "black box" (Braitenberg, 1984), which is why such information-processing approaches to development have become increasingly popular (Klahr, 1992; Schlesinger & McMurray, 2012). Production systems and connectionist models are of special relevance here, as working models of classic seriation exist in the cognitive-developmental literature for each of these modeling approaches.

We return to these special cases in Chapter III. We have summarized their roots within the modeling literature, and we now situate them alongside dynamical and Bayesian modeling approaches within cognitive development.

The Bayesian approach[1] to cognitive modeling allows learning to be represented, and furthermore alleviates the representational transparency concern, in that within Bayesian models, graphical structures are represented, labeled and manipulated in a much more explicit way (Lee, 2013). Bayesian inference can be used to find out the probability of a cause (usually not observable) of some observed data when there is a preexisting idea of what the cause is, but when there is relatively little data to support a hypothesis. This preexisting idea is termed the *prior* of the observed data. The term *likelihood* represents how likely the possible values of the cause are, given the observed data. However, after observing data, it is possible to update the prior distribution for the possible cause, taking the data into consideration, resulting in a *posterior* probability.

Dynamical systems developmental theorists share the connectionist and Bayesian emphasis on representing change (Van Geert, 1994).[2] They see cognitive systems as being made up of many parts, all of which change over developmental time as they interact with each other as well as the environment. Change in a dynamical system is described by sets of difference or differential equations that contain interdependent variables aligning to the parts of a cognitive system. These equations generate developmental growth curves representing, for example, how the length of sentences uttered by a child may change over time. The growth of such a competence can thus be mapped back to a set of interacting variables. However, this modeling approach is better suited to describing the shape of change, and less well suited to a mechanistic account of this change. The model we present in this monograph incorporates aspects of all four approaches, and so does not align exclusively to the symbolic or subsymbolic modeling traditions, and yet benefits from the strengths of them both.

A New Model of Sequential Size Understanding

The strengths of production systems (representing knowledge explicitly), connectionist and dynamical systems (representing learning and development), Bayesian (representing knowledge explicitly and learning simultaneously) and cognitive architectural approaches (well-understood components) can be leveraged in a principled way within cognitive models, regardless of the historically charged debate as to their relative merits (Boden, 1996; Dreyfus, 1992; Yule et al., 2013). The model we propose is inspired by a synthesis of their strengths, in which we make explicit the variables affecting the growth of sequential size understanding, and its progression to ordinal understanding. To this end, we propose an architecture split into task, WM, and LTM modules, and define a learning heuristic and an algorithm that process information within these modules. Our model does three things to help our understanding here. First, it infers the order of a discrete set of size items by trial and error, via a Bayesian ranking mechanism similar to that of Jensen, Muñoz, Alkan, Ferrera, and Terrace (2015) that creates a ranked representation in LTM. Second, it makes explicit the transition to spontaneous ordering of size-related sequences. Third, it explains and predicts the transition to ordinal identification competence. These two behavioral transitions (heuristic search to principled search, principled search to ordinal competence) happen due to discoveries made about the invariant properties of the size-related sequences, these discoveries themselves facilitated by architectural changes to WM and representational changes to LTM.

Conclusion

Research and theory development since Piaget's work has moved in many and diverse directions, but none has directly challenged the case for considering the development of discrete set understanding as an important example of discontinuity of cognitive functioning. By general implication arising from the criticisms and subsequent reduction in interest in Piaget's theory, his explanation of logicomathematical development could be taken to be wrong in full or in part. It seems time therefore either to accept his account or come up with an alternative. One part of this task should be to review the evidence base with as searching a methodology as possible, using children at around the critical preoperational transition.

Refreshing the evidence base alone is unlikely, however, to go beyond the speculative and circumstantial (which was also a problem with Piaget's own explanations). Data from children who represent the before-and-after of the alleged phase transition can only ever be suggestive of internal changes and reorganizations. This is why we offer an account that not only addresses a range of empirical findings relating to seriation development, but also a computational account that goes deeper than the behavior ever can. In the following chapters, we offer a renewed evidence base followed by a data-informed computational model of size sequencing and related abilities.

II. The Development of Sequential Size Understanding: Evidence for Developmental Discontinuity

Introduction

A monotonic sequence of items ordered by their relative magnitude is a logical construct that is essential for the full understanding of number and measurement. It comprises a linear progression in which each item has only one possible placement and it can only take one of two possible forms—from least to greatest or vice-versa. The ordinal value of each item (second, third, fourth largest, etc.) in such a sequence is only derivable in terms of this construct.

Much of what we know about when and how children come to grasp the serial and ordinal properties of a monotonic sequence comes from the classic Piagetian size seriation task devised by Piaget and Szeminska (1941) in which children are asked to construct an ordered set of sticks or blocks from a jumbled array. The conclusion from these investigations was that grasping the concept of an ordered size series is based on a sudden insight, as evinced by the emergence, at around the age of 7, in the ability to systematically construct an ordered set, item by item and without trial and error. Although this was an area of research that was not particularly sustained in post-Piagetian research (for reasons we reviewed in Chapter I), follow-up research reported thus far has not challenged this key finding from classic seriation. Window size seriation as measured by Piaget and colleagues would, therefore, appear to be a highly explicit and transparent on a discontinuity in cognitive development.

What has never been transparent, however, is how and why that change takes place. Unlike the number line and related aspects of numerical understanding, size seriation is not part of formal education and its development is not known to be dependent on cultural conventions that have to be explicitly taught. It does, however, necessarily depend on the more private skills of interrogation and processing of the perceptual relations of similarity and difference. Questions that are still unresolved today are how these skills explain the discontinuous emergence of spontaneous seriation and ordinal understanding, and whether the discontinuity of emergence is more apparent than real. Some time ago, we conducted a program of small connected experiments to explore these abilities in greater detail using size-sequencing tasks. Using explicit training on a touchscreen, we found a progression from trial-and-error learning in 5-year-olds to apparent sudden expertise in 7-year-olds, who did not need to be trained on either sequential or ordinal tasks. Part of this empirical assay was summarized earlier (see Chalmers &

McGonigle, 1997; McGonigle & Chalmers, 1998, 2001). The key conclusions were:

- The sudden emergence of expert sequential and ordinal size understanding is a genuine phenomenon of cognitive growth in middle childhood.
- Ordinal size comprehension is of an entirely different order of difficulty from the task of ordering size in a serial/temporal fashion.

The hypothesis put forward at the time was that the pressures on information management of large set sizes leads to the universal emergence of monotonic ordering as a solution to the combinatorial explosion of multiple relations. This fails, however, to get to the root of that emergence. Is it simply an accelerated linear process based on changes in WM, for example? Or is it indicative of an important and discontinuous representational shift that would have implications for related logicomathematical abilities? Why and how does ordinal size understanding emerge at around the same time? In short, the problem is how to make the internal processes transparent in ways that go beyond the data.

In the current monograph, we describe a detailed computational model that makes explicit the perceptual and learning processes that can account for the changes observed in these training tasks. It indicates how changes in EF such as LTM and WM are involved in supporting new emergent skills, which in turn alter the memory processes in new ways. The model is itself informed by the findings and error data collected during the sequential training. These findings are the subject of the current chapter.

We start by considering where the history of seriation research has led, and why it has left untouched the central issues of sequential and ordinal size understanding: on what it is based and why it takes so long to develop? In the second part of this chapter, we describe an empirical investigation of seriation development using touchscreen training tasks that form a database for a computational approach to answering these questions.

A Review of Seriation Research

Piaget and Seriation

In their length seriation task, Piaget and Szeminska (1941) presented children with 10 sticks that varied in length by 0.8 cm increments. They asked the children to arrange the sticks in order from shortest to longest by asking them to "make a staircase" or to copy an ordered set modeled by the experimenter. Four-year-old children generally failed altogether to make a series (stage IA) or constructed small subseries using only some of sticks (stage IB). By 6 years, over 50% of children succeeded in making the complete series through trial and error (stage II). Crucially, however, between 7 and 8 years (stage III), most children spontaneously made a principled series by systematic selection of each stick in order from longest to shortest, or vice-versa, without errors. This purportedly marks a phase change from preoperational to concrete operational thought as described in Chapter I. A key behavioral criterion defining the concrete operational phase in this context is that children can order any number of sticks such that they no longer show the set-size limitations evinced at the earlier stages of development (Inhelder & Piaget, 1964). In addition to principled ordering, Piaget found that operational children could correctly insert a new, single element directly into its correct place within a constructed series.

As described in Chapter I, Piaget explained the apparent discontinuity in behavior around 7 years of age in terms of a phase transition from figurative to concrete operational thought. Piaget considers perception to have a kind of structure for the child long before the child is capable of understanding structure analytically as in operational thinking. Accordingly, although he admits that "the perception of relations is elementary" (Inhelder & Piaget, 1964, p. 5), he is concerned to argue that perception is never a stand-alone factor in children's thinking but is always integral with sensorimotor schemata that the child performs on the perceptual field (see Inhelder & Piaget, 1964, pp. 5–16). During the preoperational stages of seriation development, the child's actions are likely to be guided by more global properties of the visual field such as the "good form" of items that happened to be already ordered by size. Although the overall configuration of such a series could have enabling effect on the choice of the next item, global perception occurs at the expense of a more analytic understanding, and prevents the child from knowing how to correct or rearrange a disordered series. The expression "good form" that Piaget uses in this context originates from the Gestalt School of the early 20th century to describe the perception of orderliness and coherence. Although Piaget uses the idea of global figurative perception over and again in describing the constraints of preoperational thinking, he distances himself from the Gestalt idea that good form is a psychological given arising from something akin to physical field forces (Kohler, 1925). For Piaget, perception is never primary in that sense but advances and evolves with knowledge itself. Global figurative perception is merely a stage along that path.

The shift occurring around the age of 7 was argued to be brought about by an ability to overcome the influence of configurative perception and by being able to mentally represent the logical status of perceived relations irrespective of how they are presented visually. This shift to operational thinking was the sudden insight into the logicomathematical property of reciprocal relational reversibility inherent in all sets. What makes seriation "operational" is that it "deals with the transformation of asymmetric transitive relations" (Inhelder & Piaget, 1964, p. 11). Relational reciprocity allows a given item to be represented in terms of its status with respect to every other item in a set. For example, Inhelder and Piaget (1964) state that during the task of seriating the set ($A < B < C < D < E < F < G$), the older child is aware that "a given element, say E, is both longer than those already in the series ($E > D,C$), and shorter than the ones yet to follow ($E < F,G$)" (p. 257), and this awareness guides the selection and placement of each element.

The ability to seriate was thought by Piaget to require exactly the same mechanisms of relational coordination as required by his transitive reasoning task using sticks. Piaget, Inhelder & Szeminska (1960) presented children with three differently colored upright sticks with barely discriminable size differences between them. He presented them as separate pairs ($A > B$ and $B > C$) but asked them to infer the size relation between A and C without letting them compare sticks A and C directly. Children could not reliably give the correct answer on this test until the age of concrete operations. Piaget's argument was that the transitive inference, A must be greater than C, was achieved by mentally ordering the items into a series A > B > C. The mental seriation involved here was thought to depend on exactly the same processes as physical seriation in which the child recognizes the bidirectional nature of the asymmetric size relations A > B and B < A, allowing the construction of a middle position for item B in the series A > B > C. Thus, although in this task the child was not allowed to actually touch or replace the sticks, Piaget believed concrete transitive reasoning to emerge from the same sensorimotor origins as actual seriation.

As also described in Chapter I, linear structures evinced by seriation and transitivity also predicted other operational achievements that occur at about 6 or 7 years such as number conservation and one-to-one correspondence. These allow children to understand that cardinal and ordinal values are necessarily conserved despite changes in layout or the density of a set of discrete elements (Piaget, 1952a).

Replicability of Piaget's Findings

At the height of the popularity of Piaget's work, many of the tests of classic operational reasoning were explored within the broader context of experimental and educational psychology, and, in particular, involving tests of whether these abilities could be advanced through specific training (Coxford, 1964; Sigel et al., 1966). Much of the new work on seriation was carried out on large samples of children by Kingma (Kingma, 1983a, 1983b, 1984a). A typical study might involve several manipulations of Piaget's task including specific training, but it would also include a replication of the classic task in which children were invited to "make a staircase" with 10 sticks of different lengths. These replications showed a good agreement with Piaget's findings indicating a sharp rise in the likelihood of obtaining a correct series between the ages of 6 and 7 (Kingma, 1983a, 1984b). Although, not quite covering the operational transition, a training study carried out by Blevins-Knabe (1987b) demonstrated that 8-item seriation was subject to changes between the ages of 5 and 6.5. The majority of younger participants failed to seriate or insert correctly whilst the majority of older participants were able to seriate (with minor trial and error) and insert missing items into their correct place. These authors found no evidence, however, that training children to recognize correctly seriated constructions enhanced their seriation performance, leading them to conclude that seriation is stage-like and strongly age-related. We review training studies in more detail later on in this chapter, but suffice it to say here that even specific training has never led to claims that seriation difficulty in children younger than 6 is anything but genuine (see also Neapolitan, 1991).

Seriation in the Neo-Piagetian Era

Although seriation did not become victim to the claims of early competence that beset many other Piagetian tasks, as noted in Chapter I, this was less to do with direct refutation as neglect. One rare exception was a study by Koslowski (1980) who argued that trial-and-error seriation and the item-insertion behavior of preoperational aged children was nevertheless systematic (and thus operational). Her argument was based on the observation that item ordering by preoperational children generally observed a consistent direction of change, albeit based on crude, large-scale size distinctions. Accordingly, Koslowski argued for a more gradualist explanation than Piaget had for younger children, but her study did not include the transition to error-free seriation or insertion in older children.

A more radical claim, however, was made with regard to Piaget's test of transitive reasoning. Using a variation of Piaget task with sticks, Bryant and Trabasso (1971) reported a competence for making transitive choices in children as young as 4. The paradigm they employed arose specifically from two arguments. The first was that children may not remember the key pairwise relations that they first observed when asked about the test pair subsequently. The second was that Piaget's task did not control for false-positive responses. For example, sticks *A* and *C* could be verbally labeled as nominally "big" or "small" producing a correct response to the test question without requiring relational coordination (Bryant & Trabasso, 1971; Youniss & Murray, 1970). This second concern resulted in increasing the number of relations to four ($A > B$, $B < A$; $B > C$, $C < B$; $C > D$, $D < C$; $D > E$, $E < D$) on which children as young as four were explicitly trained. The critical test pair was the non-end-point pairing *B* versus *D* which could not be solved by simple labeling. To control for memory failure, the trained relations were tested along with the questions about the nontrained pairings. In the context of high levels of retention on the trained pairs, high levels of transitive choices were obtained children as young as four (Bryant & Trabasso, 1971).

Training introduced its own complications, however (Breslow, 1981; Riley & Trabasso, 1974; Thayer & Collyer, 1978). The convention of starting by training each pair in consecutive order, for example, allowed the possibility that transitive choices during test were based simply on the temporal order in which the relations were trained (*AB* through to *DE*). This was finally confirmed in training and testing with and without such clues by Kallio (1982). Spatial clues were also implicated through the use of a training box in the study by Bryant and Trabasso wherein the longest item was presented at one end and the smallest at the other. This was confirmed as another possible factor by Schnall and Gattis (1998).

The very extensive debate regarding transitive reasoning, essentially devolved to precise detail about how the information was trained, and it is perhaps no wonder that transitivity research has not yielded clear answers regarding the coordination of perceptual relations within this task. After many decades of research, the logical account (Bryant & Trabasso, 1971; Halford et al., 1998; Wright, Robertson, & Hadfield, 2011) sits alongside numerous others ranging from temporal associativity and stochastic probability accounts (Delius & Siemann, 1998; McGonigle & Chalmers, 1977) to fuzzy trace (Bouwmeester, Vermunt, & Sijtsma, 2007) and dual process theories (Wright et al., 2011).

There are two fundamental reasons why the transitivity paradigm is difficult to relate directly to seriation. One is that the role of perception has been all but eliminated from the tasks in order to control for nonlogical solutions such as the possibility that the children might remember the absolute sizes of items in the test (Braine, 1964; Smedslund, 1963). This concern rendered all subsequent tasks as quasi-symbolic, where the items, although shown to represent physical size differences, were usually token objects that were known—but not perceived—to be of different sizes and described linguistically only (Bryant & Trabasso, 1971; Kallio, 1982; Riley & Trabasso, 1974). If the idea was to remove distractions from children's natural attempt to use logic, research by Perner has cast doubt on this (Perner & Mansbridge, 1983). Perner, Steiner, and Staehilin (1981), for example, reintroduced visual feedback into the task and found that this manipulation led children as old as eight to produce responses that were different from those made by adults. In a task comparing 6-year-olds with adults, Perner and Mansbridge (1983) concluded that the spontaneous approach by young children is to consider pairs of sticks in terms of the nominal categories "long" and "short." But perhaps most crucially from the point of view of seriation abilities was the fact that children were rarely explicitly asked to order all the items on which they had been trained, and, when they were, children aged 6 and under showed the same difficulties as predicted from classic seriation (Chalmers & McGonigle, 1984; Trabasso & Riley, 1975).

As for ordinal understanding, children's reputed problems were queried by neo-Piagetians, but these challenges were based mainly on early numerical competences with small numbers (Gelman, 1972; Sarnecka & Wright, 2013), or the correspondence between count words and the items being counted (Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Gunderson, Spaepen, & Levine, 2015). In terms of relative size understanding in sets larger than two, there was remarkably little evidence reported during the neo-Piagetian era about when children can identify or insert items within larger sets on the basis of their ordinal size, that is smallest, middle-sized, second biggest, and so on. Exceptions are studies by Blevins-Knabe (1987a) and Siegel (1972), both of which showed that before the age of 6, children experienced considerable difficulty with identifying ordinal sizes even within small sets. Difficulties therefore with both sequential and ordinal size understanding by children under the age of 7 remain uncontested.

Improving the Evidence Base: Methodological Issues

Although the general empirical picture presented by Piaget has survived subsequent reinvestigations, it is not a suitable foundation for pursuing a more contemporary information-processing analysis for two main reasons. Using a one-off construction task is, first of all, not conducive to discovering how children learn from their own actions of visual interrogation and selection. The Piagetian account presumes that direct sensorimotor experience enables the transition from one level of understanding to the next, but what that experience actually confers is hard to tell. It is difficult to gauge within the classic paradigm because it is generally tested in the form of a one-off episode without any opportunity for the child to benefit from the specific experience of attempting to seriate. As mentioned above, the classic seriation task has been subject to specific training, but this has focused on whether children have proceeded to a new Piagetian level as measured by other tests, rather than to gain an insight into what is specifically learned from the task at hand (Bingham-Newman & Hooper, 1974; Kingma, 1987). In his modeling investigation (which we describe in Chapter III), Young (1976) attempted to induce success in his participants by varying the task, but in a highly ad-hoc fashion where blocks were added or taken away and/or where the set was under constant replenishment by the experimenter. What children learn from actually performing a size seriation task, or indeed whether it has to involve actual physical replacement at all, is still an unanswered question.

This raises a second and even more fundamental methodological issue generated by the classic task. Its apparent transparency notwithstanding, to what extent is the physical series constructed by the child a clear and direct reflection of their ability to sequence items monotonically, that is, to really follow a strict linear order? In selecting sticks from a set to be ordered, the child is constantly reducing the visible test pool, making it hard for the investigator to infer which particular elements in the total set are subject to greatest difficulty. In placing them into an array, the choices are then subject to the vagaries of whether or not the previous selections were correct. During the lifting and placing, moreover, there is no control over whether or not the child might make gratuitous comparisons with other sticks by pairing them up spatially and comparing directly. In short, size seriation is a competence that may emerge directly from learning to process visual information in the environment, but exactly how that happens has remained an issue locked up in a historical paradigm that requires revision.

In this second part of Chapter II, we describe the experiments that convinced us that there is indeed a major developmental shift in size sequencing ability that can be demonstrated even in small samples of children, using a simple training paradigm. It reflects and confirms the age of Piaget's operational transition but, in using a more in-depth methodology, it offers precise measurement, relevant task comparisons, and a suitable basis for computational modeling.

General Methodology

Our program of experiments used a training-based methodology conducted on a desktop computer and touchscreen. Feedback-informed training was used to assess whether or not the task was within the capability of the child, given reasonable task exposure. It also allowed measurement of difficulty based on the length of training required and type of error committed. The specific learning criteria and upper limits on task duration are described separately for each experiment below.

The testing was conducted by a single female senior researcher with access to a room dedicated to the experiment. Training was normally conducted on a daily basis during weekdays.

Experiment 1

Rationale

The research question for the first experiment was to see what, if any, difficulties would be encountered if children were trained to order differently sized squares on a computerized touchscreen and to test whether the difference between 5- and 7-year-olds' performance found with the classic task would be in evidence here. Without any precedent for this computerized version of seriation, a modest version of the seriation task with only five elements was used. To be certain of the extent to which the learning profiles were specific to sequencing size relations, and not simply a string of absolute size values, two interleaved sets of different size ranges were used throughout.

A remaining issue was whether or not any learning difficulty we might find could be attributed specifically to the child's difficulty in understanding size relations, as opposed to a more general difficulty in monitoring any sequence (size or otherwise) in WM. To obtain an independent assessment of WM in a serial learning context, a control comparison was included in which children were asked to learn an arbitrary sequence of different-colored but same-sized objects. The only basis for connecting the order of these different-colored items was the temporal order assigned randomly to the objects.

Method

Design

A nested design was used with age as a grouping factor, with two levels, 5- and 7-year-olds. Task was a within-participant factor, with two levels: monotonic size sequencing and color sequencing. Half the participants were assigned a size series in the biggest-to-smallest direction and half the reverse. Color series were assigned randomly to participants. For half the participants, color ordering preceded size sequencing; for the other half, the size-sequencing task came first.

Participants

Participants were 24 children drawn from a large private school in central Edinburgh with a predominately middle-class intake. Twelve of the children (seven girls and five boys) were drawn from the Primary 1 class and were aged between 5 years; 1 month (hereafter written as 5;1) and 5;9 (*Md* = 5;3), and twelve were Primary 3 children (five girls and seven boys), aged between 7;4 and 7;11 (*Md* = 7;7). The sample was a homogeneous one, consisting of children of professional parents with English as their first language. The children were predominantly white Scottish, and as such were representative of the local population. We did not assess the IQ of the pupils, but the entrance criteria would have precluded children with lower than average intelligence from taking part.

Tasks and Stimuli

The icons used in the monotonic size-sequencing tasks were uniformly colored squares with sides ranging from 10 to 50 mm, with an interval difference of 5 mm on each side. For the color-sequencing task, the squares were of equal sizes (30 mm) and varying colors deriving from a parent set of ten. Two sets of five stimuli were generated for use in the size sequences, with ranges of 10–30 mm, and 30–50 mm, respectively, both with a 5-mm interval difference. The stimuli were presented equally spaced in a horizontal array on a level base as depicted in Figure 1A. The arrangement of all stimuli was random and varied randomly from trial to trial.

Touchscreen Pretraining

All participants were given a familiarization episode that taught them how to use the touchscreen and what to expect from the experiment proper. This familiarization consisted of a simple two-stimulus practice task, which demonstrated the feedback value of a bleep (indicating a correct response) versus a buzz (indicating an error) as described in detail in the next section. The items (squares and circles) were to be touched in a specific order and the

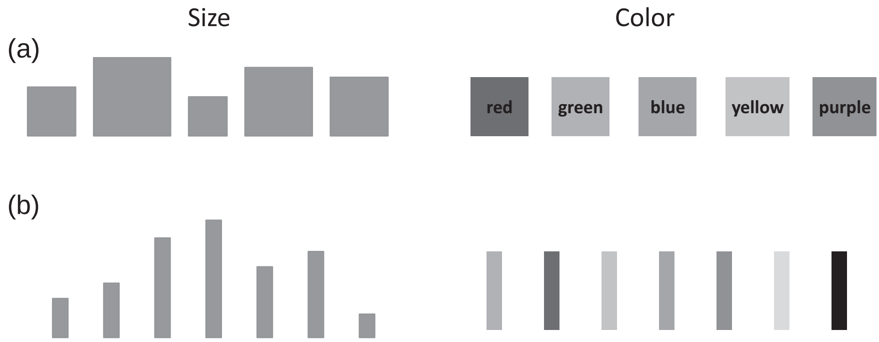

FIGURE 1.—Examples of the layout on a given trial for (A) the 5-item size and color-sequencing tasks (Experiment 1) and (B) the 7-item versions (Experiment 2). In the size condition, the items must be touched in a (fixed) increasing or decreasing order of size; in the color condition according to a fixed arbitrary sequence.

participant was required to reach a criterion of four out of five trials correct before proper training commenced.

Monotonic Size Sequence Learning

In these tasks, the participant was trained to touch differently sized squares on the screen in a decreasing or increasing order of size. Across trials, the spatial arrangement of stimuli within the horizontal layout varied randomly (see Figure 1A). The two size ranges were randomly interleaved across trials. When a stimulus was touched in its correct position with the sequence, it blinked and was followed by a bleep; when touched out of order, it blinked and was followed by a buzz. A second touch to an already-touched stimulus resulted in a blink only. The array remained on the screen until all items had been touched, a short tune indicated the end of the trial, and a new array followed after a 2-second interval. Training proceeded until a criterion of four out of five trials completely correct sequences had been met (two each from the two sets of sizes) within a limit of 40 trials in total.

Verbal report

After completing the task, participants were asked if they could describe the order in which they were to touch the squares and how many there were. They were not corrected or pressed for further information.

Arbitrary Color Sequence Learning

Every participant was assigned an arbitrary sequence of five items drawn from the parent set. For example, one such sequence might be red, green, blue, yellow, and purple (see Figure 1A), to be touched in that order irrespective of the spatial arrangement. Training on this sequence was reinforced exactly as it was for the size series just described. The learning criterion was four out of five completely correct sequences within the same trial limits as those used for the size series.

Verbal report

At the end of the task, participants were asked to describe the order of colors in which they had to touch the squares and how many there were.

General Procedures

The experiments were carried out on the school premises, in a quiet, darkened testing room that allowed the computer screen to be viewed under constant illumination conditions and testing to take place without distraction. Children were tested on a daily basis without interruption as far as possible. Test sessions normally lasted for about 20 min per participant.

In addition to the touch-by-touch feedback described above, and to encourage learning generally, both the color and size tasks utilized a game feature, taking the form of a graphic device showing a man ascending a five-rung ladder where each rung represented a completely correct sequence. The man descended by a rung, however, on an incorrect trial. When criterion was met, the man reached an apple on a tree and a fanfare was played.

All touch intervals were recorded in milliseconds on the computer.

Results

Analyses

All data were analyzed for age effects as well as the effect of task (size vs. color). Non-parametric tests were used where presumptions of normality of distribution were violated. Effect sizes were calculated using Cohen's d for F and t values, and r for Wilcoxon's Z, where $r = (Z/\sum N)$. Unless stated otherwise, N is the sample size of the group. Statistical comparisons using t tests and correlations (r) were two-tailed. Our general hypotheses were that 7-year-olds would out-perform 5-year-olds on both tasks and that they would show little or no error during the size task.

Choice

All children met the learning criteria. All but one of 7-year-olds were spontaneous in correctly sequencing the items, by starting their criterion run on the first or second trials (where an immediate switch was made to the correct end-point following an incorrect guess on the first trial). Only one 5-year-old did so. Performance was measured by the numbers of trials to the start of a criterion run (TSCR), as well as the number of error touches made by each child during training. Table 1 depicts the mean (and *SD*) for both groups and both tasks.

TABLE 1

Trials and Number of Error Touches (Mean and *SD*) for Both Age Groups During Size and color 5-Item Sequence Learning (Experiment 1)

	Size		Color	
Age	TSCR	Errors	TSCR	Errors
5 years	16.0 (15.4)	37.2 (40.5)	11.6 (9.1)	37.9 (33.0)
7 years	1.7 (0.5)	1.9 (2.0)	3.7 (2.1)	10.7 (6.6)

Note. TSCR = trials to start of criterion run.

A Shapiro–Wilk normality test found that the TSCR data were not normally distributed for either the size or the color task for 5-year-olds, $W = .83$, $p = .020$ in both cases, and so differences across age and tasks were evaluated using the Wilcoxon signed ranks test.

Age Comparisons

As hypothesized, 7-year-olds were significantly faster at reaching criterion than 5-year-olds, $Z = -2.94$, one-tailed, $p = .001$ for the size task, $r = .60$, a large effect, and also for color, $Z = -2.86$, one-tailed, $p = .002$, $r = .58$, a large effect.

Task Comparisons

On the hypothesis that the size sequence would be learned faster than an entirely arbitrary color sequence, 7-year-olds showed a significant effect, $Z = -1.65$, one-tailed, $p = .05$, $r = .33$, an intermediate effect, but 5-year-olds did not, $Z = -.82$, $p = .21$. Among the 5-year-olds there was no correlation across the two tasks in terms of rate of acquisition as measured either by TSCR, $r(10) = .15$, or error touches $r(10) = .14$ ($p > .50$ in both cases). Similarly, among the 7-year-olds, no correlation was found across tasks using TSCR, $r(10) = .21$, or error touches, $r(10) = .01$ ($p > .50$ in both cases).

Despite the significant overall age effect, both groups accelerated to completion. When criterion (4/5) was calculated for each item separately it was found that for 5-year-olds the median point at which all items in the sequence were at criterion levels was at the second item in the series for size, and at the third for color. For 7-year-olds, it was at the first and second, respectively. The respective learning curves are depicted in Figure 2A.

The types of error made by the children were divided into those due to missing items in a sequence (forwards) versus backtracking to an item already correctly touched (backwards). As can be seen from Table 2, the former considerably outnumbered the latter for both tasks and was significant for the 5-year-olds on paired-samples *t* tests (where each child was compared with him/herself) across tasks, $t(22) = 2.99$, $p = .012$; $t = 3.85$, $p = .002$ for size and

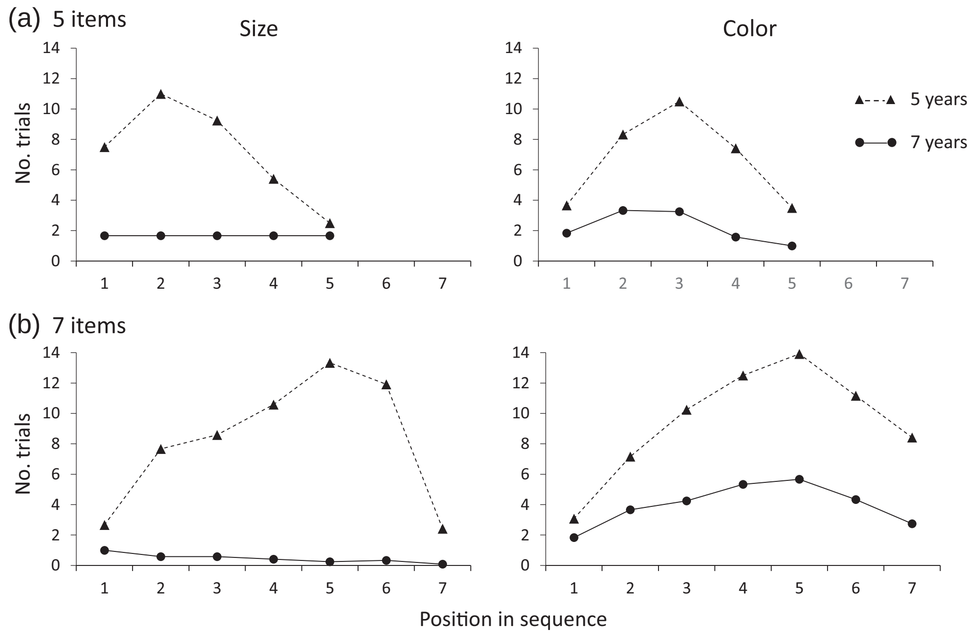

FIGURE 2.—Learning functions for (a) the 5-item size and color-sequencing tasks (Experiment 1) and (b) the 7-item versions (Experiment 2) depicting the average trials t7o the start of criterion run across each item position.

color, respectively. Both were large effects, $d = 0.86$ and 1.11. Error was too low for 7-year-olds to allow us to run a statistical test.

Reaction Times (RTs)

RTs from the correct (four) trials during the criterion run were used in the analyses. A standard procedure of removing outliers was employed (Borghi & Scorolli, 2009), which consisted of calculating the mean for each position in the sequence across the four entries for all participants. Any score higher than 2 *SD* from the mean was replaced by the mean for that position, before calculating the overall mean. This represented 5% of the total data set. The functions obtained are shown in Figure 3A. Controlling for lack of homogeneity of variance and normality of distributions, the Scheirer–Ray–Hare two-way non-parametric analysis of variance (ANOVA) was used

TABLE 2
MEAN (*SD*) FOR ERROR TYPE DURING EXPERIMENT 1

Age	Forwards	Backwards	Forwards	Backwards
5 years	32.5 (36.5)	4.8 (5.6)	31.0 (26.0)	6.9 (5.9)
7 years	1.9 (2.0)	0.0	10.0 (5.2)	1.3 (1.4)

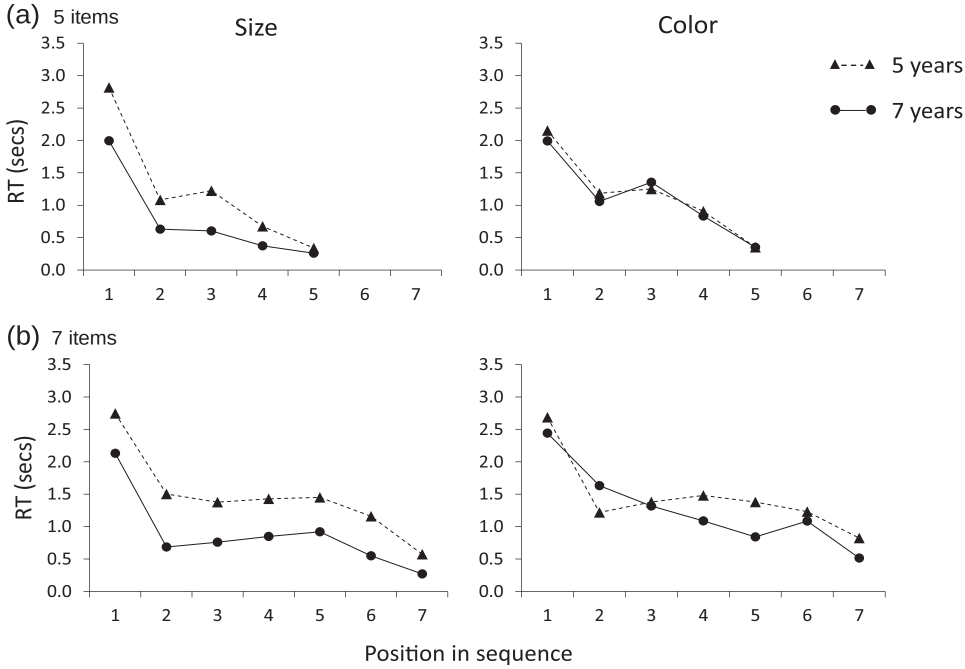

Figure 3.—Reaction time (RT) functions for (a) the 5-item size and color-sequencing tasks (Experiment 1) and (b) the 7-item versions (Experiment 2) depicting the mean RT (sec) for the last (4) successful trials across item position.

across age and task. The main effect was found for age for the size task, $F(1,116) = 23.27$, $p < .001$, $d = 2.05$, a large effect. No interaction was found between age and position, $F(1,116) < 1$. For the color task, there was no main effect of age, $F(1,116) < 1$, nor interaction between age and position, $F(1,116) < 1$. Comparing the effect of task within age groups, no main effect was found for the 5-year-olds, $F(1,116) < 1$, nor any interaction between task and position, $F(1,116) < 1$. For the 7-year-olds, there was a main effect of task with the size being significantly faster, $F(1,116) = 13.12$, $p < .001$, $d = 1.54$, a large effect. No interaction between task and position was found, $F(1,116) < 1$.

Verbal Report

Answers to questions about order, size, and color were recorded for both tasks and sorted into four categories. For each category, a score was given based on (a) degree of unambiguous unique specification for each item, (b) the number of different linguistic labels (whether ambiguous or not), (c) the number of items specified in the answer to the question "how many were there?" and (d) the number correctly specified in answer to the question about order. For example, for the size task, a participant's description of the type "smallest; nearly the smallest; middle; nearly biggest; biggest" or even "zero; little; little middle-sized; big middle-sized; big" would be scored as identifying every object in the set—a score of five in (a). Labels that were not

TABLE 3
MEAN SCORE FROM VERBAL REPORT OF 5-ITEM SEQUENCE (EXPERIMENT 1)

	Size				Color			
Age	a	b	c	d	a	b	c	d
5 years	3.17	4.08	4.90	3.00	5.00	5.00	4.60	4.45
7 years	3.83	5.08	5.46	3.83	5.00	5.00	5.25	5.00

Note. a = unique identifiers; b = separate identifiers; c = number estimate; d = order.

so specific, for example, "a bigger one and another bigger one" would be scored as specifically identifying only one object in (a) whereas "big and really big" was scored as identifying two in (a). Both of these would score two in category (b). Sequential information was admitted as a defining attribute for (b), for example, "the little one and the little one what was after the other little one" or even "bigger; bigger and bigger," where the child demonstrated with a hand gesture the increase in size. Number by answer (c) is self-explanatory. Order information (d) was given a numerical score in terms of degree of specification and correct order (e.g., for a biggest to smallest order: "big; little; little what was after other little, then middle-sized, then little", which would score three in (d) on the grounds that "middle-sized" was mentioned after "big" and was followed by a "little." Questions about the color were similarly scored, though colors in all categories had to be clearly demarcated even if they were not quite accurate (e.g., pink for purple).

The results are tabulated for both groups in Table 3. Although the data are essentially categorical and based to some extent on the experimenter's interpretation, we can arrive at a mean numeric score across participants to give a representative picture of the accuracy and specificity of the verbal descriptions. The ideal score for all categories is 5, but the mean can vary in either direction through overall underspecification or overspecification. Table 3 shows that linguistic specification of the items in the set is more accurate for the color task than for size for both groups. It can also be seen, however, that even 7-year-olds failed to denote the size values with a uniquely identifying description. Five-year-olds were as likely to use relational terms such as "bigger" or "smallest" as they were to used absolute descriptions like "big" and "little," but no 5-year-old used ordinal terms such as "second" or "third" in the size tasks. Three 7-year-olds did use ordinal terms (but all three overestimated the set size by one). Only one participant, a 7-year-old child, used the expression "biggest to smallest" when answering the question about order. Estimates for number were variable for both groups. For the size task, six of the 5-year-olds were correct, the answers for another four children ranged from "three" to "six" and two could not answer the question. For color, only three children were correct, most answered "four" and one could not answer the question. Seven out of the twelve 7-year-olds were correct on the number estimates for size, the rest ranged from "six" to "six or seven." For color, eight children were correct, the rest ranged from "four" to "seven" in their estimates.

Discussion

This study generated findings that are consistent with the Piagetian literature using the classic seriation task in that almost all of the 7-year-old children ordered the size objects with negligible error. Making an allowance for the fact that they had to guess which end to start from on the first trial, 7-year-olds in this study thus met Piaget's criterion for spontaneous seriation. Only one 5-year-old fell within this learning range. As a group, 5-year-olds conformed to Piaget's description of trial-and-error rather than spontaneous seriation. Only 7-year-olds, furthermore, showed a significant advantage for size over color sequencing in terms of choice measures. RTs also showed an age effect in favor of 7-year-olds in terms of overall speed of responding, as well as a selective advantage for the speed with which the size (as opposed to the color) sequence could be executed. The RT results show that even when they have learned to sequence sizes, 5-year-olds are not only less skilled in their performance than their older peers but that they are no faster in executing a learned size sequence than an entirely arbitrary one.

The study demonstrated, however, that monotonic size sequencing could be trained in children as young as five, albeit for a set containing only five items. The length of training for each individual was not correlated across the two tasks for the 5-year-olds, suggesting a different role for WM in the two tasks. The distribution of forwards versus backwards errors suggests a relatively small contribution in either task for having to remember what had been touched. The errors were mainly due to wrong forward selections which is a common finding in sequence learning (Terrace & McGonigle, 1994).

The age-related results in size sequencing would not have been predicted from the verbal reports. Both groups showed ambiguity and lack of specificity in describing the series afterwards and both groups made inaccurate judgments of the number of items in the set. At face value, this suggests that the superior performance by 7-year-olds was not due to a mediational variable such as verbal labeling or explicit counting.

Overall, the data were consistent with expectations derived from the classic task. However, it remained to be seen whether the 7-year-old group's performance would stand up if the set size were expanded. The relatively speedy of acquisition of 5-item sequences by younger children, moreover, was deemed to offer too sparse an error distribution to inform the computational modeling of putative learning mechanisms. To provide data bearing on these issues, we conducted an extended version of Experiment 1 employing a stimulus set containing seven items. The resulting Experiment 2 is discussed next.

Experiment 2

Transfer Pilot Study

The expanded set size required us to replace the square stimulus objects with rod-like stimuli so that all seven items could fit on the computer screen. Prior to running Experiment 2, we conducted a brief study to test whether changing the stimuli alone would affect outcomes. Specifically, we repeated Experiment 1 as described above with 12 5-year-olds (*Md* = 5.6), but employing rod-like rather than square stimuli. After training, we changed the sets size from five to seven items by adding on two at the end, for both the small-to-big and big-to-small directions. One 5-year-old was an outlier on the 5-item set requiring 78 trials to start criterion run. The remainder succeeded on a mean of 7 TSCR (*SD* = 5.8) on the 5-item set. When transferred to the larger set they required an average of 15.6 (*SD* = 14.4) additional trials to regain criterion performance. There was thus no suggestion that the rod-like stimuli themselves would cause any discrimination difficulties. The full assay of 7-item seriation is described below.

Rationale

We repeated the procedures of Experiment 1 with new participants and a set size of seven to evaluate whether there would be any diminution in the 7-year-olds' performance with an expansion in set size, and also to obtain a richer error database for modeling.

Method

Design

Once again, a nested design was used with age as a grouping factor, with two levels, 5-year-olds and 7-year-olds, and task (size vs. color) as a within-participant factor. Task order and direction of sequencing in the size task were counterbalanced as before.

Participants

Participants were 12 Primary 1 children (5 girls and 7 boys) aged between 5;4 and 5;10 (*Md* = 5;7) and 12 Primary 3 children (4 girls and 8 boys) aged between 7;4 and 7;11 (*Md* = 7;5) drawn from the same population as in Experiment 1.

Stimuli and Tasks

Stimuli were rod-like rectangles (see Figure 1B), 5-mm wide. For the stimulus set used for size sequencing, each rectangle ranged in height from 10 to 40 mm for the small-range set) and from 40 to 70 mm for the large-range set. For the color task, the stimuli were uniformly sized rectangles of 5 mm wide by 40 mm high shown in varying colors drawn from a parent set of 10. As long as the child was willing to continue, no limits were imposed on the total number of trials. In all other regards, the procedures were exactly as described for Experiment 1.

Results

All data were analyzed using the statistical tests as described for Experiment 1.

Choice

All children met the learning criteria. Six of the 7-year-old participants were spontaneous in their size sequencing success, starting criterion run on the first or second trial. Although none of the 5-year-olds showed the spontaneity found in the 7-year-old group, three of them were within the learning range of the older children. Two, however, required 109 and 120 trials, respectively. Although in neither case did the child appear to fail to understand the task (in that they showed gradual improvement from the outset), these two outlying data points were excluded from statistical comparisons. Table 4 shows the choice data in for both groups for TSCR and error touches for both tasks.

The missing data for the two outlier learners on the size task were replaced by the mean for the rest of the group. The corrected mean (and *SD*) for this purpose was 20.7 (13.8). A Shapiro–Wilk normality test found that the TSCR data were not normally distributed for the 7-year-olds for the size task, $W = .72$, $p = .001$, and group and task comparisons were again made using a Wilcoxon signed ranks test.

TABLE 4

Trials and Number of Error Touches (Mean and *SD*) for Both Age Groups During Size and color 7-Item Sequence Learning (Experiment 2)

	Size		Color	
Age	TSCR	Errors	TSCR	Errors
5 years	36.3 (38.7)	111.3 (152.2)	23.3 (12.6)	176.2 (143.0)
7 years	3.5 (2.8)	7.6 (12.4)	9.9 (7.4)	71.6 (76.4)

Note. TSCR = trials to start of criterion run.

TABLE 5
MEAN (*SD*) FOR ERROR TYPE DURING EXPERIMENT 2

Age	Forwards	Backwards	Forwards	Backwards
5 years	85.1 (118.8)	26.2 (41.3)	100.0 (70.3)	76.2 (75.9)
7 years	6.2 (9.6)	1.2 (4.3)	45.5 (41.3)	26.6 (35.8)

Age Comparisons

As hypothesized, a significant effect of age was found for size sequencing, $Z = -2.90$, one-tailed, $p = .002$, and also for the color task, $Z = -2.49$, one-tailed, $p = .006$. Both were large effects, $r = .84$ and $r = .72$, respectively.

Task Comparisons

No significant difference between performance on size versus color was found for 5-year-olds, $Z = -0.22$, one-tailed, $p = .410$, but performance on the size task was significantly better than on the color task for 7-year-olds, $Z = -2.35$, one-tailed, $p = .009$, $r = .68$, a large effect.

As Table 5 shows, forwards errors significantly outnumbered backwards errors for the size task on paired-samples t tests, $t(22) = 2.2$, $p = .049$, for size, and $t(22) = 2.69$, $p = 0.021$ for color. Both were intermediate effects, $d = .66$ and 0.77, respectively. For 7-year-olds, the color task showed a significant bias toward forwards errors, $t = 5.35$, $p = .001$, $d = 1.54$, a large effect; error was too low to run a statistical test for size.

Although learning for size versus color was not different for 5-year-olds, the larger number of errors overall allowed a deeper examination of task differences in terms of error touches. Even with outliers' data excluded ($N = 10$), this analysis revealed significantly fewer errors for size than for color sequencing, $Z = -2.19$, $p < .05$, $r = .44$, a large effect. It also revealed a significant correlation across the two tasks on both measures of learning difficulty for 5-year-olds ($N = 12$), $r(10) = .70$, $p = .01$ for TSCR and $r(10) = .58$, $p < .05$ for error touches. This relationship was even stronger when the two outlier participants were removed: $r(8) = .84$, $p = .002$ for TSCR and $r(8) = .67$, $p = .030$ for error touches. The error was too low for the size task for 7-year-olds to conduct a meaningful correlation.

A summary comparison of monotonic size sequencing across age and set size is given in Figure 4.

Reaction Times

RTs for the last four completely correct sequences were subject to cell replacement for outliers as for Experiment 1. The resulting functions are shown in Figure 3B. The Scheirer–Ray–Hare two-way non-parametric ANOVA was used to compare across age for both task conditions. The main effect for age was found for the size task, $F(1{,}164) = 40.63$, $p < .001$, $d = 2.72$, a large effect. There was no

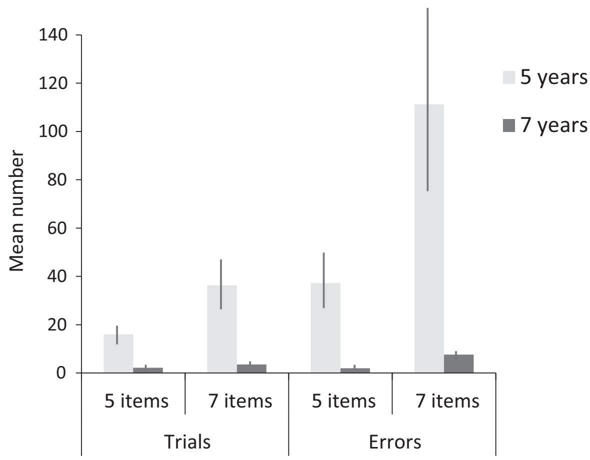

FIGURE 4.—A summary of learning difficulty on monotonic size sequencing as a function of age and number of items, depicting mean trials to start of criterion run as well mean number of error touches. Standard error is depicted by vertical lines.

interaction for age and position, $F(1,164) < 1$. For the color task, there was a significant main effect of age, $F(1,116) = 6.05$, $p = .035$, $d = 1.05$, a large effect, but no interaction between age and position, $F(1,116) < 1$. No main effect was found for task for the 5-year-olds, F $(1,164) = 1$, nor any interaction between task and position, F $(1, 164) < < 1$. For the 7-year-olds, there was a significant effect of task, F $(1,164) = 15.26$, $p < .001$, $d = 1.67$, a large effect. No interaction between task and position was found, F $(1, 164) < 1$.

Verbal Report

Due to the high levels of ambiguity in their answers in Experiment 1, 5-year-olds were not asked to report verbally on what they had just done or viewed during the task itself. The older participants were asked, but their answers were too ambiguous to tabulate using the coding categories used in Experiment 1. Specifically, responses in Experiment 2 seemed to be aimed at only an overall description of an ascending or descending series. Eleven of the twelve children described the series monotonically in terms of their training direction but ranged in degree of specificity. One child said, "biggest and going down." Three children mentioned both ends of the series (e.g., "smallest to tallest"). The remaining seven attempted to name the intervening items such as "smaller; smaller," and so forth, but only one specified seven unique items; the others ranged from five to eight. One remaining child split the series in terms of big and small items, and was the only one to use numerical comparatives: "started with smallest, second smallest, third smallest, fourth biggest, third biggest, second biggest, biggest." His number estimate (and that of four others) was correct. The rest ranged from "five or six" to "seven or eight." The color series, by contrast, was correctly recollected by all participants with the exception of one small inversion error. However, number estimates were correct only in 50% of participants, with both under- and overestimates (with a low of 6 and a high of 10).

Discussion

Expanding the set size by just two items had a major effect on the size-sequencing performance of the younger, but not the older group. The 5-year-olds displayed a trial-and-error solution that eventually culminated in successful performance, whereas the 7-year-olds were almost errorless from the start in applying a monotonic rule of sequencing. The contrast between the performance of the two age groups invites the question as to whether what appears to be a qualitative difference between the two age groups' strategies is, indeed, an indication of some conceptual change. An alternative is that it can be fully accounted for by a linear increase in the size of WM occurring between these ages. Our test for WM was in the form of an arbitrary color-sequencing task on which 7-year-olds, although affected by the set size increase, were much better than 5-year-olds. This is consistent with the possibility that improved WM alone could explain the size sequencing findings.

With a larger error database from which to analyze difficulty, moreover, we could consider this latter point more specifically in terms of whether the arbitrary color sequencing—our test of WM—might show a strong relationship with learning difficulty on the size-sequencing task. For the more extended learning required for Experiment 2 and the larger error database it generated, we did indeed find that the data from 5-year-olds showed a high correlation across the two types of task, consistent at least with the inference that variability in WM resource can explain individual variability in the ease or difficulty in learning to sequence sizes. However, two possible—and radically different—scenarios underlie the role of WM in a task in which a child is learning to sequence sizes, only one of which would suggest that WM resource alone could explain the age effect.

So first we have to ask how WM operates in the two sequential tasks. Both tasks showed a predominance of forwards over backwards errors. In the case of the arbitrary sequence, this bias is likely to be due to the fact that the correct items can be rehearsed (Hitch & Baddeley, 1976), whereas the order among the new ones has yet to acquired. This is similar to a standard case of list learning where the items are already known and understood by the participant. The forwards errors arise through uncertainty about the items not yet entered into the list being rehearsed. In the case of size, the task is not to memorize a list but to maintain a principled search. The forward errors could reflect uncertainty as to which item comes next or, alternatively, they could reflect in-principle confusion regarding the relational status of the items themselves. On this second scenario, the child is using a WM resource to try to differentiate items that started off as lumped together.

It may be valuable here to consider a comparable study on size sequencing with older children aimed primarily at comparing the children with autism spectrum disorder (ASD) with age-matched controls (McGonigle-Chalmers, Bodner, Fox-Pitt, & Nicholson, 2008). In particular, this research was designed to explore how impairment in EF in children with ASD might manifest itself in a sequencing task. The study established that children with neurotypical development and a mean age of around 9 years could learn to size-order sequences (of randomly distributed stars) in sets increasing from nine up to twelve items, whereas children with ASD were more limited in their ability to learn the longer sequences. It should be stressed here that there was no suggestion by the authors that any of these children was trying to grasp the concept of a series. Past the age of operational seriation, all children made some errors, but these were relatively low for sets of eight or nine objects. The reason for the group difference was interpreted as a constraint on the ability of the participants to execute a series—not to understand it. The question was, what were the executive factors at work? In this case, WM seemed to be the best candidate for both the group and set size effects. It was argued by McGonigle-Chalmers et al. that the constraint in the clinical group arose from having to keep WM "online" while possible alternatives were checked, which they described as "prospective working memory." This explanation also applied to the age-related changes within the neurotypical group where age was a significant covariate. Between the ages of 8 and 11 years there was an increase of two items sequenced before exiting the task. This is almost certainly due to well-known age-related increases in WM that can be sourced to both speed of processing as well storage factors (Bayliss et al., 2005). The data from the autism study thus shows how WM constraints can explain age-related difficulties in executing a sequence and we now have to consider if this is the only difference between the 5- and 7-year-olds in the current study?

If the WM factor for 5-year-olds is similar to that found by the neurotypical children in the autism study, both the effects of set size and the correlation across the two sequential tasks found in the current study would result from individual variability in terms of storage and processing capacity, and the age-related differences would be due to a reduction in that variability across age. In contrast, it could also indicate something quite different that we would not have found with children over the age of 7. This alternative interpretation is that children younger than 7 actually need to acquire the concept of a series before they can execute one. On this argument, the younger children would not only be susceptible to making executive errors due to WM constraints, but would actually have to learn to differentiate in a more precise way before a perfectly ordered size sequence could be executed. The need to learn to differentiate appropriately would explain the dramatic effect of set size expansion, as yet further item differentiation would be needed in that event. The 7-year-olds, by contrast, performed as if each and every item was already well differentiated, and all that remained for them to work out was the end from which the items should be sequenced. Thus, the evidence from Experiment 2 in the current study is compatible with the idea that 5-year-olds were learning to acquire the item differentiation needed to produce a principled size sequence, whereas 7-year-olds were simply executing a principled means of size sequencing that they already possessed. The RT data are also consistent with this interpretation, in that the 7-year-olds were significantly faster than the younger children in their sequencing even though these older children had engaged in virtually no learning and thus lacked the exposure to the task experienced by the younger group.

Finally, our short assessment of verbal descriptions of the items after sequence learning does not indicate that the change by 7 years is mediated by conscious access to the precise ordinal relations between and among the items such as second, third biggest, and so forth. The precision of their performance was not matched by the precision of their reporting. As to whether or not these ordinal relations can be accessed nonverbally is an issue to which we return in Chapter IV using a more transparent methodology.

Experiments 1 and 2 suggest that 7-year-olds can spontaneously deploy a nonverbal concept of linear monotonic change based on the repetition of a single asymmetric relation, whereas most 5-year-olds cannot. This conclusion notwithstanding, its full implications remain speculative. If there is an important qualitative difference in cognitive control underpinning the age effects, we need to identify what this is and how it emerges. This is the issue that is directly addressed in the simulation presented in Chapter III. Before turning to that topic, however, we considered the discrepancy between our own methodology and that employed by Piaget with his classic task with real objects.

Experiment 3

Introduction and Rationale

Age differences notwithstanding, we had successfully trained 5-year-old children to proceed from trial and error learning to correct seriation. The question motivating Experiment 3 was whether comparable success could be achieved with the select-and-place method devised by Piaget. There were several task discrepancies that could result in a lack of comparability. First, the touchscreen task deployed a nonimitative form of learning where the measured behavior was confined to item selection with no cue as to correctness other than the feedback from the computer. The classic task, by contrast, usually offered a model to copy or a human adult to imitate. Once an item has been selected and placed furthermore, it can be removed or replaced. As noted earlier this could have both an enabling as well as a disabling effect on performance. One means of aligning the two paradigms in this regard would be to train children to put items in order without an available model and also to disallow item replacement. Another would be to train stick seriation in a manner as similar as possible to that used on the touchscreen task. Given the fact that all 5-year-olds eventually achieved errorless seriation on the touchscreen, the question we ask here is whether comparable findings can be obtained with real sticks when we align the two methodologies as far as possible.

Earlier investigators have conducted training studies, but there has been no prior work answering our question directly. First, most training studies have been designed to improve the performance of younger (preschool) children to learn how much their cognitive development might be accelerated (Bingham-Newman & Hooper, 1974; Kidd et al., 2013; Kingma, 1986), rather than to learn whether those children would be capable of errorless seriation. A typical example of such a training study, described by Kingma (1987), is where kindergarten children were given various teaching elements, such as drawing attention to the differences between the items, or modeling the experimenter's own seriation with sets of varying numbers. Posttests indicated that children in the trained group improved more than those in an untrained group, but it is not clear from the data as presented what aspect of the teaching contributed to this, nor how many children achieved errorless seriation. In a training study carried out with preschool children by Swanson, Henderson, and Williams (1979) preschool children viewed an adult carrying out a correct seriation with up to six objects over several trials. Children who only observed the modeler showed no improvement, whereas those who were either allowed to try seriating while watching the modeler or were given practice after watching the tape, performed better than controls. Again, however, it is unclear how many of children in the active-seriator groups became completely successful.

Although these studies all involved the child in some way selecting and placing items, another set of investigators found that simply enhancing the experience with perceptual relations alone could improve seriation. Specifically, Timmons and Smothergill (1975) provided preschool children with pairwise comparisons of the different-length pieces taken from a set of to-be-seriated objects, although they did not actually train the children to seriate per se. Relative to control children, the children given the comparison experiences showed higher 6-item seriation posttest scores, although it is again unclear from the description which if any children were completely successful. Collectively, these studies indicate that at least a number of training procedures (albeit not passive imitation) can result in improved seriation performance, but they do not address the key question of interest here, namely, whether 5-year-old children can achieve principled seriation through specific training using the same set sizes and a methodology as similar as possible to the one used in Experiments 1 and 2. The purpose of Experiment 3 was to fill this gap.

Method

Participants

The participants in this study were 24 Primary 1 children (10 girls and 14 boys) aged between 5;4 and 5;11 (*Md* = 5;7) drawn from the same population as before.

Stimuli

The objects used for this task were wooden cylindrical rods from a set of 13, uniformly varying in height with 5-mm differences between successive pairs of items. They were divided into two sets of five and two sets of seven, each set comprising sticks from the “smaller” and also the “larger” end of the series. For testing, these were arranged upright on a wooden platform at the participant’s eye level in a random horizontal array. The children were invited to place the rods “in the right order” by taking them one at a time from the platform and placing them upright on a stand below the platform. The child was allowed to correct any misplaced item, but only if the error was registered at the point of placement; errors were not allowed to be corrected retrospectively. No model was provided, but the experimenter provided feedback as to whether the finished order was correct, inviting them either to “try again” if incorrect, or to “do it again” if correct. Participants were deemed to have met criterion when they could make four out of five completely correct series without trial and error. Two size ranges were used as for the touchscreen tasks, but it quickly emerged that participants were less tolerant of the stick training than the touchscreen training and once two correct series had been constructed with one size range, training proceeded with the other alone until it, too, produced two correct series.

Design and Procedure

Half the participants (Group A) were assigned to five-stimulus sets; half (Group B) to seven-stimulus sets. Children in Group A were then specifically trained on seven items by way of a transfer task. This manipulation was included to establish the extent to which a small set, once successfully ordered, would provide a basis for immediate expertise on a larger set.

Results

The results were assessed in terms of traditional criteria for seriation performance (Inhelder & Piaget, 1964) as well as by the learning measures used in Experiments 1 and 2. For the former, the finished series were divided into three categories: successful productions, trial-and-error successes, and failed productions as shown in Figure 5. Although there were more perfect productions than productions with errors, success on five items did not necessarily guarantee the same levels of success on seven items even when the latter was run as a transfer task.

Independent t tests showed significantly more attempted but unsuccessful orderings across set-size, during 7-item training as compared with 5-item training, $t\ (11) = 2.10$, $p = .02$, $d = .86$, a large effect. Similarly, as seen in Figure 5, the percentage of children succeeding on the first trial dropped to 50% on the transfer task and was at 33% on the 7-item task. Training effected an

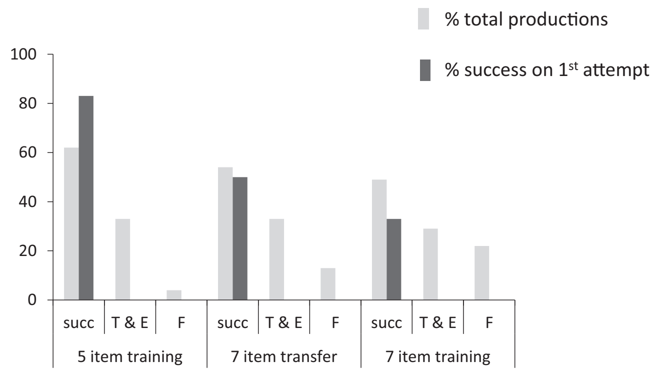

Figure 5.—Performance by 5-year-olds on a trained version of the select and place stick seriation task with five and seven items. It depicts the percentage occurrence of spontaneous seriation, trial and error success and failure (with the average number of attempts ranging from around five to twenty across conditions).

improvement in the 7-item training task; the number of successful productions during the first half of a participant's training session were significantly fewer than those registered in the second half on a t test for paired samples: $t(11) = 2.60$, $p = .025$, $d = .53$, an intermediate effect.

All children learned to construct four correct series using both size ranges of sticks, but learning was strongly affected by set size. Table 6 shows the mean numbers of trials to the start of the criterion run (TSCR) for all tasks. Normality assumptions were violated for the 5-item task, $W = 0.73$, $p < .05$, and so the scores were compared using a Wilcoxon test. The hypothesis that the 7-item task would take significantly more trials than the 5-item task was confirmed, $Z = -2.82$, one-tailed, $p = .002$, $r = .81$, a large effect. Paired-sample t tests showed a significant increase in TSCR from five items to seven during transfer, $t(22) = 2.38$, $p = .042$, $d = .68$, an intermediate effect. It should be noted that the performance recorded on the 7-item series from scratch did not qualify as "operational" using Piaget's criterion in that only a third of the group succeeded on the first attempt (this criterion normally requires at least half the group).

TABLE 6

Mean (*SD*) Trials to the Start of Criterion Run on Stick Seriation by 5-Year-Olds (Experiment 3)

5 Items	7-Item Transfer	7 Items
4.42 (5.18)	8.92 (6.72)	13.08 (6.71)

Discussion

Five-year-old children were successfully trained to seriate using actual objects without having to watch or copy a modeler, but their performance was sensitive to the effects of set size. Despite the fact that (a) children had the benefit of viewing their finished product, and (b) the training criteria were slightly less stringent than on the touchscreen than with the actual objects, the children in this study showed comparable increases (nearly trebling) in the trials-to-criterion measure when the set was extended by only two additional items. Even children who had already learned to seriate five items doubled their training effort to incorporate just two more. Not intended as a direct comparison with the touchscreen versions owing to the many differences between the stick and the screen-based task, Experiment 3 shows again that the number of items to be seriated is an integral factor affecting the challenges of nonspontaneous seriation.

At the same time that Experiment 3 confirms the lability and the trainabilty of seriation in children under 7 years, this experiment also confirms the difficulty in using the classic select-and-place task as a quantitative measure of seriation difficulty in trial-and-error learners. Sessions have to be videotaped and played back in real time to identify selection errors. The visual array representing the available stimulus pool as well as the array under construction is subject to constant changes as the construction proceeds. As the child works, there is a reduction in pool of items that remains that continually alters (i.e., reduces) the WM demands. The touchscreen task, by contrast, offers a touch-by-touch record of the sequencing, where errors occur, and how they are progressively corrected. WM demands remain constant throughout a trial in terms of the array to be interrogated and are uniform across participants. The touchscreen paradigm also offers the possibility of evaluating individual variability in WM in the form of an arbitrary sequencing control task performed under identical conditions of training.

Empirical Assay: Summary and Discussion

The development of size-sequencing abilities changes dramatically between the ages of five and seven even for set sizes smaller than those studied by Piaget. Our empirical assay has shown how this shift can be indexed by errors during acquisition and RTs. The near spontaneous seriation by 7-year-olds and their rapid RTs when seriating contrasts with the trial-and-error learning shown by most 5-year-olds, and also with the fact that these younger children were no faster in executing the size sequence than with an entirely arbitrary list of colors. We should be clear that we are not offering a definitive level of seriation ability in the younger group for either five or seven items. There is a likelihood that slightly different performance measures would have resulted from changing the interval difference among items, or even the type of stimuli used (Stevenson & McBee, 1958). Indeed, the shift from squares—in which area or the interval difference on either vector could be used to calculate a difference—to rods, in which the size difference was limited to the vertical dimension, may have aided some children and/or hindered others. In fact, a slightly enabling effect is exactly what was suggested by the short pilot study where we tested the use of rod-like stimuli with 5-year-olds. Piaget himself noted "we might have found a marked improvement in the seriation of length had we used fewer elements, or if there had been greater differences between the elements" (Inhelder & Piaget, 1964, p. 251). The point is that variability engendered by stimulus characteristics is likely to be a feature of trial-and-error learning in this study as in other training studies (Kingma, 1983a, 1983b; Kingma & Reuvekamp, 1984). Overall, the picture to emerge from the touchscreen experiments using children of 5 and 7 years is entirely consistent with other studies on seriation, as we review in further detail in Chapter V. Taking our 7-item task as a representative case, the results from twelve 5-year-old participants now provide us with a detailed data set of the variability, amount and distribution of error to be expected when children of this age are learning to seriate.

What is not clear from data alone are the cognitive mechanisms responsible for the age-related shift in performance, and it is important to spell out two distinctly different possibilities. The first is that WM restriction alone hinders accurate seriation by 5-year-olds. On this scenario, when the child is searching for the next correct selection, a restriction on processing capacity makes it difficult to hold two candidate selections in WM long enough to reject the wrong item. Indeed, a similar concept of attenuated attentional capacity has been applied to a failure by children on the autistic spectrum to perceive pattern variation (Greenaway & Plaisted, 2005). The WM constraint that accounted for variability in performance across the 5-year-old group could indicate that, as WM improves with age, this constraint declines to the point where it is effectively at zero eventually creating the sort of uniform performance shown by 7-year-olds in our experiments.

A very different scenario is that most of the younger children were actually acquiring rather than executing the size sequence. On this scenario, the interval values that are needed to order the set correctly are being gradually discriminated during training. This is not to say that these children are unable to detect these size differences—the differences involved are not challenging for children of this age (Peckham, 1933), but rather that they need to use them in order to distinguish between and among some of the internal items. This suggestion would be consistent with Piaget's contention that there is a shift in the classification of items as "big" and "small" to their classification by their precise ordinal position in the set. On this scenario, learning produces a new discovery about the ordinal properties of discrete sets that can then be applied to any sequential or ordinal size task. Trial-and-error differentiation would thus be essentially different from spontaneous seriation in that only the latter employs an already learned concept of unitized ordinal position.

How might we situate these two possible scenarios within contemporary EF approaches to cognitive development? Very broadly, these approaches divide into unitary models that seek a common developmental process to explain how children come to understand greater structural complexity, such as the CCS theory of Okamoto and Case (1996) mentioned in Chapter I, and those that attempt to fractionate the aspects of goal-directed behavior into key components (Miyake & Shah, 1999). Both of these approaches acknowledge the involvement of PFC in the development of goal-directed abilities (Diamond, 2013). The PFC is the most highly interconnected of all brain regions and latest to mature developmentally (Moriguchi & Hiraki, 2011). There is little dispute that the development of the PFC involves highly complex and slowly maturing brain changes involving (among other things) changes in gray and white matter volume, synaptic pruning and thickening (myelination) of long central nervous system neuronal connections (O'Hearn, Asato, Ordaz, & Luna, 2008). Equally complex are the changing interrelationships between and amongst the PFC and cortical and sub-cortical areas primarily responsible for memory, planning, and sensorimotor activity (Goldberg, 2017).

Despite the complexity of PFC functioning, both unitary and fractionated approaches to EF have embraced the idea that it can be considered in terms of single isolable components such as WM, RI, and Planning. The CCS theory, for example, makes extensive use of the concept of M space developed by Pascual-Leone (1970) and suggesting that the maturation of memory capacity can account for major changes in cognitive advance. Others have suggested a strong role for RI (Houdé & Borst, 2015). The neuropsychological community, however, face the challenge of explaining selective multifarious deficits in cognitive functioning due to brain injury and are thus more concerned with fractionation. Clinicians and psychologists have accordingly designed instruments to that might causally connect affected brain areas with specific aspects of EF. As noted in Chapter I, this has led to the development of neuropsychological instruments such as the CANTAB test battery (Robbins et al., 1994). Several such batteries designed specifically for children now exist (Young, Gurm, & O'Donnell, 2017).

Whatever general approach to EF is being taken, to situate our findings within such an approach we need to confront an obvious question regarding the apparently simple goal-directed task such as setting out a series in order of size. This question is whether maturation of any one of these components alone can explain the emergence of this sequencing ability. The alternative would be a more complex story implicating more than one of the EF components during development and possible interconnectivity among them. Either way, how might any EF components be explicitly implicated in seriation development?

One such component would be age-related changes in WM—the maintenance and manipulation of information over brief periods of time (Best & Miller, 2010). As these authors point out, WM demands are highest in tasks that require self-organization, such as the self-ordered pointing task (SOPT) in which participants have to plan a nonreiterative search through randomly presented shapes (Petrides & Milner, 1982). Development right up to adolescence is noted on the SOPT task and is highly related to the number of items in the search. There is a strong parallel with the seriation task as used here, implicating WM as a main, if not the single, source of developmental change.

A second component is RI to items already chosen. Although backwards errors by 5-year-olds represented only a quarter of the total error in the size task, improvements in inhibitory control could also help account for the developmental change. Its role as a single explanatory component is doubtful however due to the circularity noted in Chapter I of arguments that see stronger solutions arising from the inhibition of weaker ones. A third EF factor in self-organized tasks is planning—a factor that correlates with SOPT scores (Ross et al., 2007). Unlike WM and RI, this cannot be so easily regarded as a variable that could be subject to simple linear or maturational change in the context of seriation, as it appears that the plan itself undergoes a radical and apparently qualitative change. Younger children seem to operate according to a rough guide to sort from big through to small items; older ones to actively seek the next and only the next size in turn from a given end-point. An obvious interpretation is that a new plan becomes available to children as they get older; a plan that is qualitatively different from the one driving trial-and-error learning. In short, it could be the formation of this plan—not the changes in WM and RI alone—that gives rise to the transition in development.

In an EF context, therefore our two scenarios essentially devolve to the issue of whether we can ascribe our findings to the maturation of a single component process or whether we need to invoke a more complex relationship among components where new plans emerge as a consequence. Either way, memory development clearly has an important role, but WM is not the only aspect of memory that may be relevant to the development of size sequencing as we review next.

Long-Term Memory

Given the serial nature of the tasks we have used, the role and importance of WM is obvious, but it also has implications for the more permanent storage of information known as Long-Term Memory or LTM. An elusive and still somewhat disputed concept in memory research (Miyake & Shah, 1999), this is not as clearly manipulated or measured as WM as it can involve an unknown element of general background knowledge, which can be of a procedural sort (such as how to start a car) or of an acquired semantic nature (such as four is bigger than three). There is some agreement that WM must interact with LTM in most skilled tasks, and some have argued that WM is in fact an "activated" part of LTM (Kintsch, Healy, Hegarty, Pennington, & Salthouse, 1999). In the case of a developing skill such as size sequencing, the activated part of LTM would, therefore, contain some sort of growing information about what a size sequence is and how to construct one. But memory for the trained sequence whether of a working or long-term sort, are, in effect, the dependent variables of our task to which we have no direct access. On the basis of our evidence, we can only conclude that both seem to develop.

The answer to our questions is that the role of WM, RI, planning, and LTM are all likely to be implicated in seriation development, but they are hard to separate and operationalize any further in terms of behavioral outcomes alone. Adopting an EF approach either of a unitary or more fractionated version does not in itself resolve the enigma of the behavioral phenomenon of an apparent discontinuity in cognitive functioning. Unless we can see these components at work during the developmental transition, we are still unclear as to whether the changes are linear or whether they work in tandem to produce a new intervening variable that is responsible for the behavior we see in 7-year-old children but is itself invisible to inspection. Fortunately, these factors can be separately modeled in a moment-to-moment simulation of seriation learning, as each EF factor has a unique role in controlling choice. A simulation based on a representation of the actual perceptual environment and training feedback given in our tasks should be able in principle to show how LTM can be updated during a learning experience, affecting both the knowledge state of the learner as well as its procedures for solving the task. And from such a simulation, the question of whether or not a new solution emerges from learning (along the lines of our second scenario) can be put to a direct test.

We have reported the main learning criteria in this chapter in order to obtain an age comparison in terms of task difficulty and to situate our findings within the relevant literature. However, we also obtained detailed learning profiles from the 5-year-olds in terms of error distribution by ordinal position. As this gives us a profile of how and where error becomes reduced in the course of learning, this is the data that we shall use in the computational modeling of seriation learning discussed Chapter III.

III. A Computational Model of The Emergence of Sequential Size Understanding

Introduction

Sequential size understanding as measured by training on a 5- and 7-item monotonic sequence is characterized by apparent discontinuity across the ages of 5 and 7 years. As we have found, however, age-related changes in WM for arbitrary sequence learning also develops during this period. The question that confronts us in this chapter is whether linear developments in WM alone can account for the apparent discontinuity. The alternative is that age-related changes in WM interact with aspects of the sequencing task itself to promote an entirely new skill.

If we look to the cognitive modeling literature in general, a large variety of computational formalisms have addressed the issue of how the various gradations, spurts, and plateaus that characterize cognitive development emerge. However, characterizing these changes in information-processing terms remains a challenge (Klahr, 1992; Schlesinger & McMurray, 2012; Shultz, 2003; Sutton & Barto, 1998). Our focus in this chapter is to contribute to such an understanding via a novel computational model of the emergence of sequential size understanding.

What Must a Model of Sequential Size Understanding Deliver?

In broad terms, a computational model of sequential size understanding should deliver a coherent account of developmental progression including the emergence of spontaneous seriation, an adjudication between a continuous or discontinuous explanation and should align with reality in terms of psychological plausibility.

In specific terms, we can translate these objectives into the following set of questions:

1. Assuming that trial-and-error monotonic (size) seriation is learned through psychological processes, is it possible to build a computational model of this learning similarly constrained by psychological theory? In particular, can we elucidate how factors relating to EF such as WM and LTM play their part in this learning?
2. Using this analysis, is it possible to make transparent how spontaneous monotonic (size) sequencing may emerge directly from trial-and-error learning between the age of 5 and 7 years? Here we would use the representational devices within the same computational model.

3. Is it possible to predict how other corollary skills such as ordinal competence may emerge using the representational devices within a computational model that has already learned to seriate spontaneously?
4. How are the concepts of continuity and discontinuity in size sequencing development clarified by this model?

Prior to presenting our model, we summarize how existing computational models of Piaget's classical seriation task represent and explain developmental continuities and discontinuities.

Production Systems and Connectionist Models of Piaget's Classical Seriation Task

To date, two computational models of seriation (Mareschal & Shultz, 1999; Young, 1976) have focused on the classical Piagetian select-and-place size seriation task and specify information-processing devices that capture such seriations successfully. These were the production systems of Young (1976)[3] and the connectionist model of Mareschal and Shultz (1999)[4]. In the 1970s, and prior to Young's model, several production systems characterizations of seriation emerged adopting a similar methodology (Baylor, Gascon, Lemoyne, & Pothier, 1973; Baylor & Lemoyne, 1975). We choose here to analyze that of Young (1976) due to his exclusive focus on size (as opposed to weight) seriation, his "concern for empirical confirmation" (p. 16), as compared with the production systems by Baylor and colleagues, and the availability of his models for hands-on experimentation for interested readers (Scott & Nicolson, 1991).

Young's (1976) model provides a mechanistic story for the emergence of select-and-place seriation, fitting specific production systems to the action profiles of individual children. Young (1976) models both successful and unsuccessful seriation attempts of children aged between 4 and 6 years, during a standard length seriation task, converting each move, captured on videotape, into a production system condition-action rule set. Young views such production rules as forming a kit from which a specific production system can be derived, such that a child can be situated within in a "space" of seriation skills along three dimensions: first, selection (the choice of which block to work with); second, evaluation (whether or not to accept a block as a suitable addition to the line); and third, placement and error repair (where in the line a block should go). Seriation development is seen as the addition of one or more critical rules to each of these dimensions of competence. These skills come together collectively and collaboratively to generate operational seriation in the older child. Young (1976) acknowledges that production systems can benefit from external cues, and he discusses the benefits of making partially seriated sets available in the visual field. This is similar to the effect of "good form" on preoperational behavior as also noted by Piaget (see Chapter II). He also considers other "hints" to seriation success, such as showing the gaps in partially ordered lines and went on to incorporate these configurative features into his model. Young's model also caters for the growing capacity for error correction in older children, which coheres with the fact that Piaget regarded trial-and-error behavior as the penultimate stage before operational seriation.

As opposed to collecting and fitting child data profiles to model output, the connectionist model of Mareschal and Shultz (1999) represented the three broad stages of seriation development as identified by Piaget, from incorrect, to trial-and-error, and finally operational error-free behavior. Connectionist models can represent a perception-action loop if the perceptual component is seen as an input vector and the action component as an output vector. A series of such operations can simulate a serially ordered sequence of actions, which is the basis of a simulated action within the Mareschal and Shultz (1999) connectionist model. Their model is thus conceptually similar in approach to Young (1976) in which seriation is decomposed into a succession of independent moves based on the perceptual features of a set of stimuli. They provide an explicit and testable computational implementation of select-and-place seriation but in the form of a single cascade correlation network (Shultz, 2003). The model has a capacity for learning (via an error minimization algorithm) and development (via the automated recruitment of network nodes to increase its representational power), and learns to sort disordered arrays of six elements, progressing from error-prone to error-free item selection. Inputs are of the format {5 2 4 1 6 3} and outputs are of the format {1 5 2 4 6 3}. In this instance, the model determines that the stick in the fourth position should be moved to the first position of the array, having just been exposed to the correct sequence. The array is then readjusted to fill the missing position by being moved to the right, resulting in the array which is consequently presented as an input array in the next simulation iteration. This training regime is based on the theoretical assumption that the child can benefit from the inspection of correctly ordered sets created by a tutor or caregiver. The representations that emerge within the model as a result of this training regime can be seen as an emerging set ranking within the hidden unit values, a ranking which ultimately leads to seriation success.

In terms of what these models deliver, both show how an information-processing mechanism can be built to clarify how the young child's competence on the classical size seriation task develops. In Young's case this is by the principled addition of more rules, and for Mareschal and Shultz, by the principled addition of graphical nodes, resulting in more representational power. Both models are continuous rather than discontinuous, represented in the Young (1976) model by the linear addition of extra rules, resulting in expert seriation, and in Mareschal and Shultz (1999) model, by the addition of hidden layer nodes again resulting enhanced seriation performance. Although he mentions that individual rules are "a few of the (presumably very many) rules that the child has stored in his head" (p. 49), Young (1976) will not commit to the psychological realism of the production system as a whole, or make transparent where the additional rules come from that allow seriation expertise.

Young's (1976) model is thus tending toward the descriptive as opposed to providing an explanatory mechanism for development. This lack of learning and plasticity is common in early production systems, a problem alleviated by current architectures, as explained by Yule et al. (2013). However, Mareschal and Shultz (1999) make the case for the biological and psychological realism of their mechanism, in that on the addition of hidden units, corresponding accelerations to new developmental levels occur in their seriation task. Indeed, biological neural networks show the type of structural growth consistent with their model, in terms of the growth of new neurons in children and adults (Quinlan, 2003).

Both models stop short of explicitly addressing the growth of EF components, such as WM and LTM, and crucially the progression to spontaneity and ordinal competence. These models also vary regarding the empirical database by which they are informed. Mareschal and Shultz (1999) simply use a broad classification of errors made in each Piagetian stage as a success metric for developmental progression, and so their model is not informed by a detailed empirical data set. Although Young (1976) empirically tested the seriation behaviors that he modeled, his chosen age range barely extended into that of operational seriation and the principled seriation behavior observed by Piaget in children older than 6 years. It is perhaps not surprising that he characterized the productions as a “quite messy” collection of different rules.

The Mareschal and Shultz (1999) and Young (1976) models thus only partially address our questions of interest. Although of a different theoretical origin, task presentation, and computational architecture, our model shares the aim of these two models in specifying in explicit and testable terms the emergence of size sequencing skills. It differs from both in that it is aimed at capturing common computational constraints in describing how children can arrive at a principled ranking through perceptual learning and feedback alone without any physical placement of items. It differs from Young’s model in that it fits the resulting model output to empirical data that capture the progress from trial-and-error learning to principled seriation within a given individual and a given set size. It uses the simulation of the learning to show how the emergence of spontaneous size sequencing will be an inevitable outcome of such learning given the necessary memory resources. In contrast to Mareschal and Shultz, it does not use a discrete set of numbered units as given input/output relations, nor any form of imitative learning; rather it demonstrates how information inputted in broad binary terms eventually becomes unitized through learning by the individual alone. It also addresses the variance relating to the length of trial-and-error learning of younger children due to individual differences in WM. Finally, it makes transparent how ordinal competence can emerge from the transition to spontaneous seriation.

A Computational Model of Sequential Size Understanding

As noted in Chapter I, the model we propose is inspired by a synthesis of the strengths of dynamical systems, connectionist systems, production systems, Bayesian cognitive models, and cognitive architectural principles. Chapter IV contains the computational model of ordinal competence; this chapter's focus is on sequential size understanding. Our model does several things to help our understanding of how sequential size understanding emerges. First, simulating the 5-year-old child, the model infers the order of a set of shapes by trial and error, via a Bayesian ranking algorithm similar to Jensen et al. (2015). This process creates a gradually more precisely ranked representation of the items in LTM. We simulate this via the *heuristic search model*. Second, we simulate via a *transitional model* the representational change that leads to spontaneous sequencing and ultimately ordinal competence. Here, links between rank-ordered items representing correctly selected shapes get gradually stronger, this process being facilitated by a more robust WM. Thirdly, we simulate the spontaneous ordering of size-related sequences via the *principled search model*. Here, the model has discovered that a heuristic search across many possible actions is less efficient than an algorithm that selects stimuli based on the principled iteration of a "select smallest difference" rule. Crucially, the models are all informed by the developmental psychological literature, as explored in the next section in terms of our working assumptions regarding the built-in psychological constraints shaping the operation of the model.

Constraints on Sequential Size Understanding

As Smith and Breazeal (2007) note, core design principles are essential for a mechanistically realizable account of a behavioral phenomenon. There are several design constraints from psychological theory that should be incorporated into models of how children (and nonhuman primates) control potentially combinatorically explosive search spaces. Such constraints can provide the start conditions for how this control becomes manifest over time and experience. In the context of size sequencing, a first start condition for child and model is to be able to make initial broad classifications that allow a separation of items into the most and the least likely to be correct.

The first constraint, therefore, is an asymmetric, binary relational rule that divides sets into *GOOD* and *BAD* subsets. Applicable to all learning, it can also serve to dichotomize sets and collections of size objects into broad categories, such as *BIG* and *SMALL* (Leiser & Gillierion, 1990; Piaget & Szeminska, 1941). However, in the case of sequence learning the task is to move from a rough classification of likely candidates to a specific ranking where there emerges a single unique position for every item in the set. In the model, a core binary constraint behind the final ranking is treated as a selection between *GOOD* and *BAD* choices, and is applicable to every binary decision whether for size or for color.

The second constraint is the nature of the environmental information available to govern *GOOD* and *BAD* choices. For our arbitrary sequence, this is simply a set of different colors with no intrinsic relationships, whereas for size it is a set of discriminably different values with no other differentiating features.

Here the task is to move from a rough classification of the set (e.g., *BIG* and *NOT BIG*) to a set of unique values. The biological constraint assumed to operate here is the default process of relational comparison (Lawrenson & Bryant, 1972; McGonigle & Jones, 1978), whereby perceptual generalization will seek out a plausible candidate for the next item to be reinforced on the basis of similarity to the one already correctly selected (Reese, 1968).

The third constraint is also selective to the size sequencing condition and recognizes the early emerging tendency to classify size items into *BIG* or *SMALL*. This is known to invoke perceptual reference points that determine the starting point of this dichotomization (Clark, 1970, 1973). According to Clark, the ground plane is a natural reference point for vertical height, identifying stimuli near it, as well as those near the "skyline" (Clark, 1970; McGonigle & Chalmers, 2002). Similarly, Bryant has argued that a natural tendency in children's relational discrimination is to use an external frame of reference as the primary binary comparison (Bryant, 1972), in which case selecting the biggest item would be an easy first step. Theoretical reasons apart, a common observation in seriation research as well is the selection of an *extremum* or end stimulus (Leiser & Gillierion, 1990). The rapid learning of these end stimuli is also what we found in Experiments 1 and 2 (Chapter II). This is, therefore, an additional constraint operating on the serial ordering of size differences.

The fourth constraint is the inhibition of selected stimuli from further selection. Bullock (2004) has shown that small numbers of perceived stimuli are represented in a one-to-one manner with patterns of activity in the PFC, patterns that start to disappear after stimuli are manually selected in a task environment. This constraint is part of a general serial ordering mechanism, which produces a tendency to forwards rather than backwards errors (McGonigle-Chalmers et al., 2008; Terrace & McGonigle, 1994).

In sum, the four constraints enumerated above are part of a general mechanism dedicated to controlling serially ordered behavior. It does this by reducing the potential number of sequential combinations available at any point in a set. This reduction is effected through the general selection and inhibitory learning constraints. Further constraints apply over and above these that are specific only to the size task. These are the binary relational rule and end-point bias constraints.

Specific Operating Assumptions

We see all complex biological systems as being confronted with the adaptive problem of economical information management. The design primitives just outlined in the previous section form part of our operating assumptions, but we see these as working in combination with WM and LTM data storage mechanisms. Before we outline the model, we describe how we all of these components can work together in a co-operative interchange.

The immature system parses the sets and collections of objects it encounters in a pairwise fashion. It iteratively selects a referent stimulus and compares candidates against it, and then selects from these candidates. This pairwise comparison is a basic procedure that evolution has discovered to reduce the potentially lethal combinatorial explosion problem inherent in any multistimulus scenario. Should the task demand it, as in a sequencing scenario, it can then allow the cumulative effect of these comparisons to facilitate the creation of linear orders in LTM.[5] By LTM, here we mean the semipermanent store of the series that is constantly being updated as a consequence of correct selections.

As the system grows, a WM increase facilitates the discovery that, when ordering sizes, the correct choice always coincides with the selection of the smallest interitem size difference. Representationally, this has the effect of augmenting a stable rank order already created in LTM with *links* between correctly ordered set elements. For a 5-item set, the ranked representation augmented with links can be seen as $\{A > B > C > D > E\} \bigwedge \{A\ r\ B\ r\ C\ r\ D\ r\ E\}$, where r represents the interval relationship, linking A to B, B to C and so on. This relationship can be used to short circuit the probabilistic selection procedures of the immature system and allow spontaneous stimulus selection according to a "select smallest interval" rule that drives both expert size seriation and ordinal matching.

These operating hypotheses informed the final versions of the heuristic, transitional and principled search models presented in this chapter, the ordinal model presented in Chapter IV, and earlier drafts of these models, as briefly summarized in the following section.

The Evolution of the Computational Model Into Its Current Format

The computational methodology chosen combines aspects of Bayesian cognitive modeling (Lee, 2013), dynamical systems modeling (Van Geert, 1994), and the cognitive architectural approach, housed within a procedural simulation program (Yule et al., 2013). We chose to build our own modeling framework, due to concerns that the number of free parameters and predefined, and possibly hidden, theoretical constructs would have been very large should a ready-made cognitive architecture (e.g., ACT-R, SOAR) have been used (Miłkowski, 2013). Despite this concern, there are nevertheless many parameters in our model relative to the simple parameters defining the behavioral change. This is because the model can externalize invisible processes that can cause major representational change, but that are expressed in the data simply as change from failure (errors) to success (few or no errors). These include the rapid scanning processes of visual interrogation, the constant interplay between WM and LTM representations of the task and elusive transitional states. The last may contain discoveries that, we shall argue, propel the system into a new way of behaving. The discoveries can be represented in the model independently from the algorithm that starts to deploy them.

Crucially, the design of the model was driven by the empirical observations described in Chapter II, in addition to the psychological constraints outlined above. First, the error patterns generated by 5-year-old children suggested that they were representing stimuli as a weak internal ranking—even for the size-orderable stimuli. (This form of weak ranking was also proposed within the production system model of Harris and McGonigle (1994) to explain representations underlying triadic transitive choice for 5-year-old children.) Strong end-point learning, however, indicated that these end stimuli would likely be the reference stimuli to initiate pairwise comparisons. The weak ranking led us to consider probability theory as a way of representing stimulus uncertainty, specifically Bayesian models of ranking using Gaussian probability distributions, which were considered because of their use of serial pairwise comparisons to allow a probabilistic ranking to emerge rapidly and reliably (Koller, Friedman, & Bach, 2009). Following this point of conceptual coherence with our empirical observations, and our mathematical proof as to the emergence of rankings via paired comparisons, we simulated the emergence of a Gaussian probabilistic rank in the LTM of an artificial agent in a pilot study.

Although it was able to form a ranking within the LTM of the artificial agent and to generate error patterns comparable to that of the 5-year-old child (though not to a statistically significant level), the pilot study was abandoned for the following three main reasons. First, the Gaussian distributions were inappropriate for the data to which they were being fitted. The model produced a stimulus ranking in which the Gaussian distributions representing the end-points had more uncertainty than the middle elements, yet the data from 5-year-old children showed a strong bias toward the end-points, to which many of the errors were funneled. Second, the Bayesian inference algorithm that operated against the Gaussian distributions (approximate Bayesian inference using expectation propagation) was highly complex mathematically, and greater transparency was sought to allow the model to be understood. Thirdly, the use of Gaussian distributions could result in stimuli having a negative ranking, which was hard to rationalize psychologically.

However, the pilot study allowed us to demonstrate precisely how serial paired comparisons could be successfully used within a model of serial order to form rank orders with an inherent uncertainty. At this point, we adopted the *Beta* probability distribution as the representational basis for stimulus set ranking for the following two reasons. First, the *Beta* distribution represents the binary set division thought to be at work as seriation emerges, its parameters (α, β) representing two proportions of a set. Second, the *Beta* distribution has a mean value defined by $\alpha/(\alpha + \beta)$, which has the property of increasing in probability mass as $|\alpha - \beta|$ increases, allowing a natural representation of end-point bias. In sum, the choice of the *Beta* distribution alleviated the concerns we had using Gaussian distributions and formed the basis for the final draft of the model.

The pilot study also allowed us to understand precisely how many components would be required for the model, the type of information processing these components would have to do, and how they would interact. Thus, we saw it as essential to have a system that was complex enough to represent a task environment and an agent that could scan, evaluate, select, learn about, and inhibit stimuli while in dynamic interaction with this environment. We represented these operations in a set of Equations (1)–(6) and associated data structures. The model design thus avoided introducing any constructs or calculations that were not necessary. The most complex part of the model (the Bayesian inference mechanism) was made transparent within the bounds of a design decision to use continuous probability distributions. This was achieved using *Beta* distributions, which allowed transparency of inference via direct manipulation of parameters—something not possible with Gaussian distributions. It also allowed a fit of model to child data while taking account of their natural representation of end-point biases. In sum, at the end of the pilot study, we had a design for a plausible and transparent model with the requisite level of complexity.

The Simulated Task and Agent Architecture

We represent the child's information processing as a simulation, separating a simple yet appropriately designed experimental environment and agent (Cooper, 2016; Yule et al., 2013) in which an artificial agent interacts with a task environment containing two-dimensional stimuli (Parisi & Schlesinger, 2002). The agent is designed to be able to perceive, learn, and act upon stimuli within this task environment in an analogous way to the child. The task environment is designed to replicate the experimental structure with which the young child engages in the size and color experiments detailed in Chapter II. In order to simulate the behavior and learning of a 5-year-old child interacting with a touchscreen environment for sequential size and color conditions, the artificial agent is the virtual subject of a series of simulations, each comprising 40 trials, in which sets of shapes must be ordered correctly. The aim within each trial is thus for the simulated agent to repeatedly select the shapes presented to it until it selects the target sequence correctly (e.g., 13323435 counts as a complete trial). The following sections use a 5-item sequential size task to illustrate the data structures and information processing within the task and agent for these models.

Task Environment Representation

The representation of the task environment within the model was informed by the structure of the touchscreen experimental paradigm, and so follows the stimulus and trial structure therein as closely as possible. Thus, the task environment is a one-dimensional array containing a set of stimuli C_n with size, shape and color attributes. Each element of C_n has a unique position i within the task array. C_i has the attribute set $\{size, shape, colour\} = \{100 - 5i, rectangle, blue\}$. The remaining elements of C have the attributes $C_1 = \{95, rectangle, blue\}$, $C_2 = \{90, rectangle, blue\}$ and so on for the remaining stimuli. The task array shuffles at the start of each trial, producing a permutation of C_n. This means that the stimulus array appears different to the perceiving agent at the start of each trial. The frame of the screen on which the stimuli appear serves as a referent for the selection of plausible start points C_1(*minimum vertical difference*) and C_5(*maximum vertical difference*).

Agent Representation

The task environment is perceived by a single, simulated artificial agent that can compute the size difference between any two distinct elements of C at one point in time. The size difference being detected is on the vertical axis. Time is represented as the discrete iterative steps of an individual simulation.

The agent's WM is a list that holds the results of perceptual computations, each element of this list being subject to probabilistic decay to determine its availability. A real value (0–1) determines the availability of the WM element. For example, if the list element has a value of 1, the probability of its availability is 1. This is a similar concept for WM to the buffers used by Yule et al. (2013).

The agent also has an LTM, in which the agent maintains a set of continuous probability distributions, $Beta_n$, which relate to items in the stimulus array C_n in a 1:1 manner. These distributions represent the agent's uncertain "belief" in the target order of the stimulus in the array by situating it mentally along the unit interval (0–1) (e.g., for a 5-item set, $\{Beta_1 > Beta_2 > Beta_3 > Beta_4 > Beta_5\}$). The agent also maintains, in parallel, a set of relations within LTM, representing relational interval links between these distributions (e.g., for a 5-item set, $\{Beta_1 r Beta_2 r Beta_3 r Beta_4 r Beta_5\}$).

The agent is controlled by algorithms which are made up of the following procedures: *SCAN, INHIBIT, HEURISTIC SELECT, INFER RANK, INFER LINK,* and *PRINCIPLED SELECT.* Their presence in the agent's control algorithm varies depending on the model. Common to all models we have (*SCAN*, *INHIBIT*). The heuristic search model uses (*HEURISTIC SELECT, INFER RANK*), the transitional model uses (*HEURISTIC SELECT, INFER RANK,* and *INFER LINK*), and the principled search model uses (*PRINCIPLED SELECT*).

Agent and Task Interaction: SCAN

The agent interrogates the stimulus array of length n utilizing a compare-and-contrast scanning procedure.[6] At the start, a plausible end-point (big or small) is sought, and if confirmed, the target direction is set to "biggest to

Target stimulus

C_2	C_3	C_4	C_5	Referent stimulus
$C_1\ C_2$	$C_1\ C_3$	$C_1\ C_4$	$C_1\ C_5$	C_1
	$C_2\ C_3$	$C_2\ C_4$	$C_2\ C_5$	C_2
		$C_3\ C_4$	$C_3\ C_5$	C_3
			$C_4\ C_5$	C_4

FIGURE 6.—An illustration of the paired comparisons made by the simulated agent, with each cell representing the comparison made between a referent stimulus and a candidate stimulus. The referent stimuli are represented by the rows $C_1 - C_4$, and the candidate stimuli are represented by the columns, $C_2 - C_5$.

smallest" or "smallest to biggest." This selected end-point becomes a referent for the next comparison and the items in sequence are then scanned in the direction in which they are encountered in the shuffled array. Hereafter, we shall assume that the direction is "biggest to smallest"; the model is the same for both directions.

The uncertain aspect of the interitem scans is represented by a *Normal* distribution, the mean of which is $Ci - Cj$ where $\{1 \leqslant i, j \leqslant n)$ and the variance of which is a fixed value. The scanning state space is shown in Figure 6, which assumes the 1st item, the referent, has been selected correctly already. The agent compares each stimulus to a referent and on discovering the size difference, represents these comparisons in WM for further processing.

The equation for representing a single comparison within a scan of the shuffled array is as per (1), in which *NormalDifference* is a normal distribution with mean |*referent* − *candidate*| and the standard deviation is σ.

$$SCAN_{candidate} \leftarrow \log\ (NormalDifference(|\ C_{referent} - C_{candidate}\ | \times 5, \sigma)) \qquad (1)$$

This equation represents the size difference between the referent and the candidate as a multiple of five, this being the model analogue of the actual empirical interstimulus difference in millimeters. In the color condition, the size comparison term is not available, and the term *NormalDifference* $(|referent - candidate| \times 5,\ \sigma)$ is replaced with $NormalDifference(k, \sigma)$, where k is a constant. The color information is not used to represent an arbitrary string in this condition; we use the stimulus index (1–7) assigned to it within the simulation code, which represents the target order and is not perceivable to the agent, to represent the target order within the sequence of colored shapes.

Agent and Task Interaction: INHIBIT

Within each trial, and on correct selection, the stimulus is inhibited from further selection within the trial, as if a button has been pressed on the agent's representation of the stimulus that stops it from becoming a candidate during subsequent stimulus evaluations. Equation (2) simulates such stimulus inhibition.

$$Inhibitionselected \leftarrow InhibitionLevel \tag{2}$$

This change is within-trial, in that inhibition levels are reset at the start of a new trial. Notably, the level of inhibition applied to the stimulus once selected is a small constant which varies for the size and color conditions (0.00000001 for size, 0.1 for color, these two figures arrived at by parameter optimization), which means that the inhibition is not absolute; there is still a small probability of the stimulus being selected again once this constant is applied and subsequently used within Equation (3).

Agent and Task Interaction: HEURISTIC SELECT

This procedure utilizes the probability distributions contained within the vector $Beta_n$, size information derived from the "compare and contrast" comparisons in *SCAN*, and an inhibition factor, to score each candidate stimulus. This score is then converted into a discrete probability distribution that relates to items in the stimulus array in a 1:1 manner, and that inform an action selection mechanism. The selected stimulus becomes the new referent.

Each stimulus can be thought of as having an uncertain position along the unit (0–1) interval, represented by a *Beta* distribution[7] the set of which makes up the vector $Beta_n$. At the start of a simulation, the agent has little understanding of where along this unit interval each of the stimuli belong. For the size condition, it represents the biggest stimulus as being closer to 1, and the smallest stimulus as being closer to 0, the other stimuli being randomly placed. For the color condition, they are all randomly placed, as no end-point biases apply in this condition.

The uncertainty the agent has about stimulus position decreases as the simulation progresses and the stimuli gradually form a rank order. The *SCAN* information is combined with $Beta_n$ and an inhibition level (a very small constant applied to correctly selected stimuli) to generate a score for each stimulus, which are collated into a list and the *softmax*[8] function applied to them.

$$BetaPrecision_{candidate} \leftarrow \frac{Beta.\ \mu_{candidate}}{Beta.\ \sigma^2{}_{candidate}} \tag{3a}$$

$$Score_{candidate} \leftarrow \left(\frac{BetaPrecision_{candidate}}{SCAN_{candidate}} \right) \times Inhibition_{candidate} \tag{3b}$$

$$Score_{candidates} \leftarrow softmax\,(Score_{candidates}) \tag{3c}$$

In Equation (3b), the *Beta* distribution precision[9] term is divided by *SCAN* results, the result of this division being multiplied by the inhibition value attached to each stimulus. In the size condition, we see higher scores for candidate stimuli (that have not already been correctly selected) that are closer to the referent in size, representing a small interval difference, and that have more precise *Beta* distribution representations. The scores are converted into stimulus selections via a *softmax* action selection function, and so the stimulus with the highest score tends to be selected more often. Crucially, the scores generated do not correspond to the target order at the start of a simulation due to the agent not yet having formed a precise ranking of the stimulus set. This situation leads to errors being more frequent in trials at the start of a simulation. Uncertainty in heuristic selection comes from a combination of the noisy *SCAN*, the $Beta_n$ vector, and the temperature factor within the *softmax* action selection mechanism.

Agent and Task Interaction: INFER RANK

If the correct stimulus is selected, there are two learning processes, the second (*INFER LINK*) being dependent on the first. *INFER RANK* is a learning process that increases the probability of that stimulus being selected at that point in the set via changing the probability distributions contained within the vector $Beta_n$. This procedure is unique to the heuristic and transitional models. We can unpack this as follows. Within each trial, and after correctly selecting a stimulus, the agent classifies all stimuli selected correctly to date as *GOOD*, and the remaining ones *BAD*. This classification has the effect of reducing uncertainty about the set order for the agent, and a rank order forming. For vector $Beta_n$, at the start of a simulation, $Beta_1$ is weighted toward 1 (*BIG*), and $Beta_5$ toward 0 (*SMALL*), representing the end-point bias constraint. A gradual "filling up" of two buckets held within each distribution occurs, the α parameter bucket of *GOOD* stimuli and the β parameter bucket of *BAD* stimuli. Adding to the *GOOD* bucket pushes the mean of the *Beta* distribution to 1 and adding to the *BAD* bucket pushes the mean toward 0 and adding to *any* bucket decreases its variance (i.e., makes it more "pointed"). These buckets are initially empty apart from the end-point biasing, and we assume that the agent has a general preference for *GOOD* over *BAD* stimuli. As this learning process occurs after each correct selection over many trials, a rank order of *Beta* distributions of the format ($Beta_1 \gg Beta_2 \gg Beta_3 \gg Beta_4 \gg Beta_5$) emerges from the initially weakly ranked set ($Beta_1 > \{Beta_2 \cong Beta_3 \cong Beta_4\} > Beta_5$). The two "bucket" parameters, α and β, are incremented by samples from a *Normal* probability distribution (*NormalIncrement*) [Equations (4a) and (4b)].

$$Beta\,.\,\alpha_{1 \leqslant i \leqslant selected} \leftarrow Beta\,.\,\alpha_{1 \leqslant i \leqslant selected} + NormalIncrement(\mu, \sigma) \tag{4a}$$

$$Beta\,.\,\beta_{selected+1 \leqslant i \leqslant n} \leftarrow Beta\,.\,\beta_{selected+1 \leqslant i \leqslant n} + NormalIncrement(\mu, \sigma) \tag{4b}$$

The precision of the *Beta* distributions is controlled by varying the mean and standard deviation values of the *NormalIncrement* distributions. A large value with a low standard deviation results in larger, more precise increments, and a small value with high standard deviation results in smaller, noisier increments. The rank order ultimately allows sequencing success, as these *Beta* distribution variables are used in the heuristic search model [Equation (3a)] when the agent is deciding on the next stimulus to be selected.

Agent and Task Interaction: INFER LINK

If the correct stimulus is selected, and sufficient WM is available, the variance of an individual $Beta_n$ vector element is sufficiently low, and the size difference detected between the referent and the stimulus is noticed to be the minimum in the set of possible size differences, then the relational link between the referent and the stimulus is strengthened, as shown in the following equation:

$$L_{referent,selected} \leftarrow L_{referent,selected} + LinkIncrement \quad (5)$$

This procedure is unique to the transitional model, and the combined presence of precise *Beta* distributions joined up by links representing the agent's knowledge to "select the smallest interval" to get the right answer represents an important change in the agent's LTM representation of the series as each item will now de facto have unique marker. We can describe this as the items now correctly ordered as having a *slot* in *LTM*.

Slots

Slots form in *LTM* due to repeated experience with ordering sets and collections that vary along a dimension such as size. A slot is a content-free placeholder that represents a unique ordinal position in a set, rather like mailboxes within an apartment block. The content of a slot is tied to the agent's encounter of a smallest interval as part of a serial ordering exercise executing in real time, and each slot can only hold one indicator that such an interval has been encountered, for example, [*x*], [*x*], [*x*] … where *x* represents such an encounter. As such, a slot can hold an ordinal position in the form of a count (3rd from big end-point) should an ordering task demand it. Slots thus serve a dual function indicating that the "select smallest" rule must be satisfied, and also being capable of temporarily storing the ordinal position—or cumulative tally—of these computations should that be necessary.

Agent and Task Interaction: PRINCIPLED SELECT

This procedure utilizes the size information derived from the compare-and-contrast comparisons in *SCAN*, and an inhibition factor, to score each candidate stimulus. The selection is the minimum size difference in this list of scans (Equation (6)). The selected stimulus becomes the new referent. Uncertainty in selection comes from the noisy *SCAN* only.

$$Selected_{candidate} \leftarrow MINIMUM((SCAN_{candidates}) \times (Inhibition_{candidates})) \quad (6)$$

A General Hypothesis Linking Uncertainty to Representational Change

The sources of uncertainty in the agent and task interactions in the models discussed so far are as follows:

1. The scan of the stimulus array, *NormalDifference* $(|C_{referent} - C_{candidate}| \times 5, \sigma)$, which is inherently noisy.
2. The uncertain positioning of each stimulus along the unit interval ($Beta.\ \mu_{candidate}$).
3. The variance of each stimulus ($Beta.\ \sigma_{candidate}$).
4. The discrete probability distribution as generated from the *softmax* function, with a variable temperature parameter.
5. The alteration of the *Beta* ranking via $NormalIncrement(\mu, \sigma)$;.
6. The enhanced WM availability, the probability of which is represented by a real value between 0 and 1.
7. The alteration of the link strengths via *LinkIncrement*.
8. The inhibition of selected stimuli via *InhibitionLevel*.

The manipulation of this uncertainty allows error data generated by 5-year-old children to be compared to model error data, as we will see in the heuristic search model simulation section below. However, this uncertainty is also the source of representational stability, and allows us to state a general hypothesis as to how sequential size sequencing competence changes from trial and error to expert. The following diagrams illustrate the hypothesized simulated task and agent architecture, featuring a 5-item size-related sequence $\{C_1 > C_2 > C_3 > C_4 > C_5\}$. We can see here the progression of the agent's LTM representations from a weakly to a strongly ranked *Beta* vector, facilitating error-free performance within the *heuristic search model* (Figures 7 and 8).

We can now see the *transitional model*, powered by enhanced WM, setting up the "smallness" links needed for the principled search model, which also represent the slots needed for the ordinal search model (Figure 9).

Next, we see the operation of the *principled search model* illustrating size sequencing expertise via minimal size difference selection. The LTM is not used in this model, the agent having discovered the utility of the "select smallest difference" rule (Figure 10).

General Simulation Methodology

Each of the simulations were implemented using the R statistical programming language (R Core Team, 2017), version 3.3.3, and used the e1071 (version 1.7) and hash (version 2.2.6) packages.

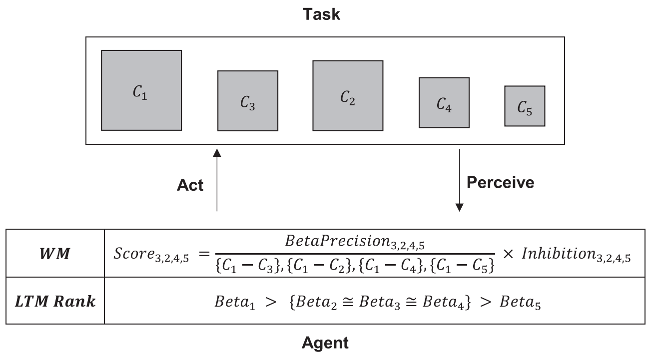

FIGURE 7.—Task and agent architecture overview at the start of a heuristic model simulation, showing the agent interacting with a 5-item monotonic set $S = \{C_1 > C_2 > C_3 > C_4 > C_5\}$. Following the direction of the randomly shuffled array, the agent scans and scores the encountered pairs with reference to the last item selected (C_1 in this scenario), stores them in *WM*, and then selects probabilistically from these scores. On selection, the agent perceives positive or negative feedback. The *Beta* distribution ranking (*LTM Rank*) starts very weakly, with only end-point biases, and changes from "weak to strong" should positive feedback be received and the variance of the *Beta* distributions decrease. As the scores will not represent the target ranking at this point, as they depend on the rank order and the variance of the *Beta* distributions in *LTM* (*BetaPrecision*) and so the selections are full of errors, such as sequence 2234321243445. LTM = long-term memory; WM = working memory.

Heuristic Search Model Simulation

The aim of this experiment was to replicate how children generate the error patterns we see in Chapter II, for both size and color 7-item sequences and so provide an answer to question 1 posed at the start of this chapter. The

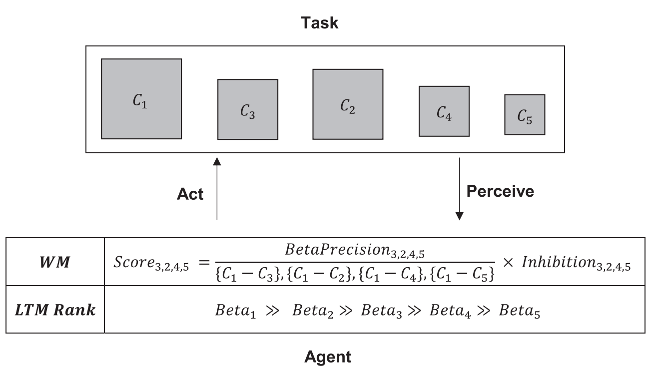

FIGURE 8.—Task and agent architecture overview at the end of a heuristic model simulation. The *Beta* distribution ranking (*LTM Rank*) is now strong, having been reinforced on positive feedback on previous trials, and so the variance of the *Beta* distributions has decreased. The scores now align the target ranking, as they depend on the rank order and the variance of the *Beta* distributions in *LTM* (*BetaPrecision*). As a result, the selections are now being more accurate, with perhaps the occasional error, such as sequence 124345. LTM = long-term memory.

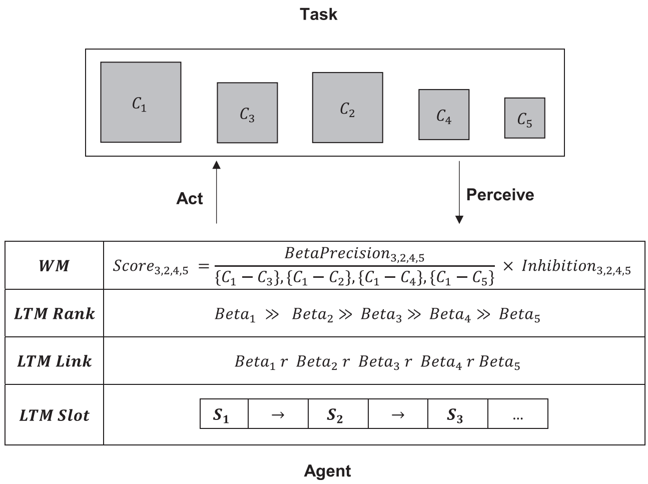

FIGURE 9.—Task and agent architecture overview at the end of a transitional model simulation. The *Beta* distribution ranking (*LTM Rank*) is strong, having been reinforced on positive feedback on previous trials within the heuristic model simulations. The increased capacity of *WM* allows the discovery that correctly ordered set elements are linked by relational (size) differences that are smallest, a discovery that is persisted via the generation of inter-stimulus links in *LTM* (*LTM Link*). The representational effect of the ranks and links forming are slots (*LTM Slot*). LTM = long-term memory; WM = working memory.

hypothesis is that there will be a correspondence with the child data for the same condition, thus validating the heuristic search model's design and operation. Ten simulations of 40 trials each were executed, and the error data generated analyzed for statistically significant differences against the child data. The two child subject outliers reported in Chapter II were excluded. The task representation was analogous to that provided to the 5-year-old

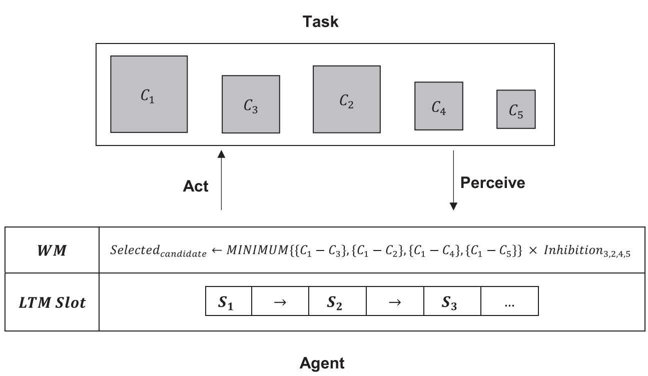

FIGURE 10.—Task and agent architecture overview of the principled search model simulation. The *Beta* distributions now rank with a low enough variance, and the links connecting correctly selected stimuli are now strong, indicating the presence of slots (*LTM Slot*) to inform choice. LTM = long-term memory; WM = working memory.

child participants and applied the same training and feedback. The simulation pseudo-code is as follows:

```
simulation ← 10
trial ← 40
n ← | C_n |
for simulation simulations
  Beta_n ← GenerateBetaDistribution
  for trial trials
    SizeDifference ← SCAN (C_n)
    selected ← HEURISTICSELECT (SizeDifference)
    If selected = target then
      Beta_n ← INFERRANK (Beta_n)
      inhibited ← INHIBIT (selected)
    else
      errors ← errors ∪ selected
  next trial
next simulation
```

The equation parameters were set as per Table 7.

For the size condition only, a bias of 0.25 was assigned to the α parameter of $Beta_1$ and the β parameter of $Beta_7$, biasing them to ranked positions of 1 and 7, respectively (in the color condition, this bias value was 0). In order to optimize the simulation parameters for both the size and color conditions, the mean squared error (MSE) was calculated by the function that compared the number of errors made by the child with the number of errors made by the computational model.

TABLE 7

THE VARIABLES AND THEIR PARAMETER VALUES USED FOR THE HEURISTIC SEARCH MODEL SIMULATION, FOR THE SIZE AND COLOR CONDITIONS

Equation	Variable	Size Value	Color Value
N/A	Stimuli	7	7
N/A	Simulations	10	10
N/A	Trials	40	40
1	$NormalDifference.\ \sigma^2$	5	5
2	$Softmax.\ Temperature$	1	1
4	$NormalIncrement.\ \mu$	0.013777778	0.06
4	$NormalIncrement.\ \sigma^2$	0.012	0.5
6	$InhibitionLevel$	0.00000001	0.1

For the size condition, all 10 simulated agents completed all trials, resulting in successful seriation of the sets at the end of 40 trials. The total number of errors made by the 10 simulated agents was 258. A one-sample Kolmogorov–Smirnov test showed that the error distributions generated by the child, $D = 0.30$, $p < .001$, and by the model, $D = .36$, $p < .001$, were both not normally distributed. Consequently, a two-tailed Wilcoxon's signed rank test for two-sample paired data was used for analyses. To establish comparability between model and child data, the test was applied to the matched conditions (model size with child size) but also to the opposite pairings (model color with child size). This showed a similarity in terms of no significant difference for the matched conditions only, $Z = -1.43$, $p = .15$, but a significant difference for the unmatched conditions, $Z = -2.17$, $p = .03$, $r = .24$, an intermediate effect.

For the color condition, all 10 simulated agents also completed all trials, resulting in successful seriation of the sets at the end of 40 trials. The total number of errors made by the 10 simulated agents was 477. A One-sample Kolmogorov–Smirnov test showed that the error distributions generated by the child, $D = .28$, $p < .001$, and by the model, $D = .22$, $p = .02$, were both not normally distributed. A Wilcoxon signed rank test for two-sample paired data was thus again carried out, which showed that the errors generated by the child were not significantly different from those generated by the model, $Z = = -.52$, $p = .61$. A significant difference was found, however, when the errors for the size condition generated by the model were matched against the child color data, $Z = -3.63$, $p < .001$, $r = .41$, a large effect (Figure 11).

We can understand how the model generates these error patterns by inspecting the set of *Beta* distributions. The precision values of the *Beta* distributions are a measure of how explicit a rank they form; low precision values are indicative of a weaker rank, and high precision values a stronger rank. They were analyzed after 1 (the start of a simulation), and 40 trials (the end of a simulation) for the size and color conditions. The cumulative effect of the learning process is illustrated by the *Beta* distribution plots from a single simulation in Figure 12 for the size condition, and Figure 13 for the color condition. For the size condition, the ranked precision values at the limit of 40 trials were: 732.08, 159.56, 86.58, 60.46, 49.40, 39.13, 32.78.

For the color condition, the ranked precision values were: 987.28, 521.11, 357.46, 270.41, 212.18, 180.68, 162.06.

Thus, the *Beta* probability distributions underlying the size condition model performance are ranked with less precision than those underlying the color condition (Wilcoxon's signed rank test, $W = 28$, $p = .02$). $Beta_n$ distribution precision within the size condition derives from precise, smaller increment values over time, and within the color condition from noisy, larger increments. This results in less precision in the size condition *Beta* distributions, and more precision in the color condition *Beta* distributions.

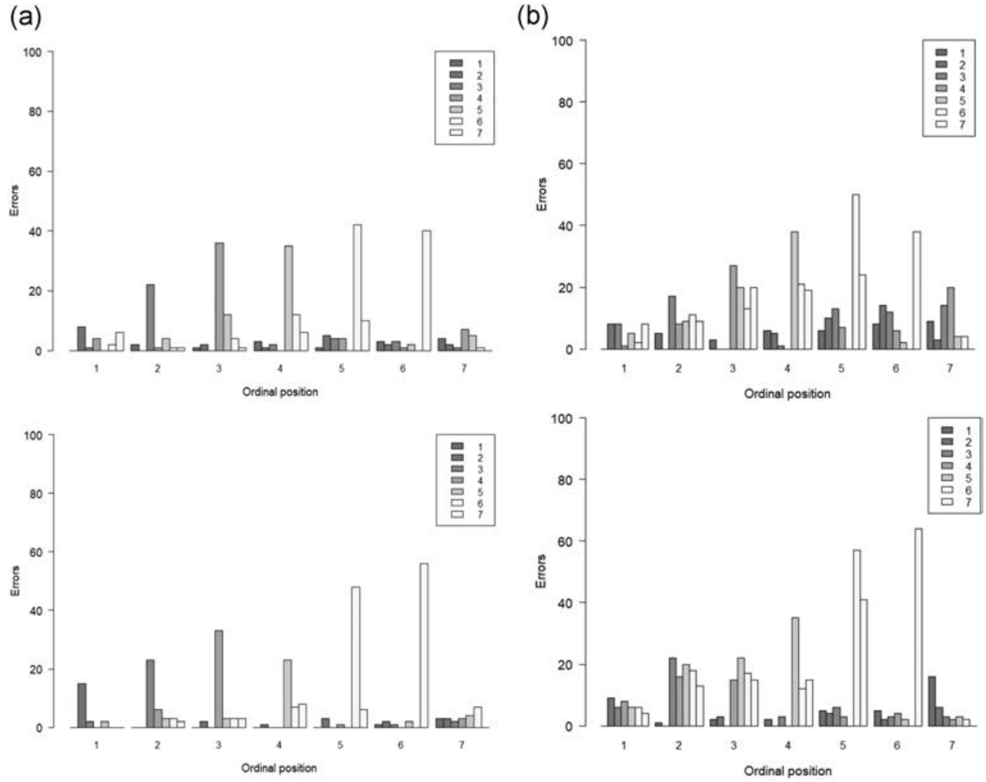

FIGURE 11.—(A) The total error counts (y axis) for child (top graph) and model (bottom graph) each stimulus attracts at each ordinal position (x axis) for the size sequence tasks, split up to show the error counts at each position that is not the target. The legends show an increasingly light shade for each stimulus position 1–7. (B) The total error counts (y axis) for child (top graph) and model (bottom graph) each stimulus attracts at each ordinal position (x axis) for the color sequence tasks, split up to show the error counts at each position that is not the target. The legends show an increasingly light shade for each stimulus position 1–7.

This is counterintuitive, as approximately half the number of errors is made by the agent within the size condition as compared with the color condition. However, the agent in the color condition does not have access to the size difference information, which appears to give momentum to the agent in the size condition in that there are very few backwards errors. It would appear that smaller, precise inferences appear to suffice in the size condition. In a sense, the size relation is propelling the agent forward through the task space according to the direction of the size relation, that is, from "biggest" to "smallest." The small value of *InhibitionLevel* within the size condition, which has the effect of strongly inhibiting a stimulus from further selection, is enhancing the direction the agent takes as time proceeds. The detrimental effect of having no size relation to afford a consistent direction of information processing can be seen within the error patterns made in the color condition, in which there are many more errors and more backtracking occurs.

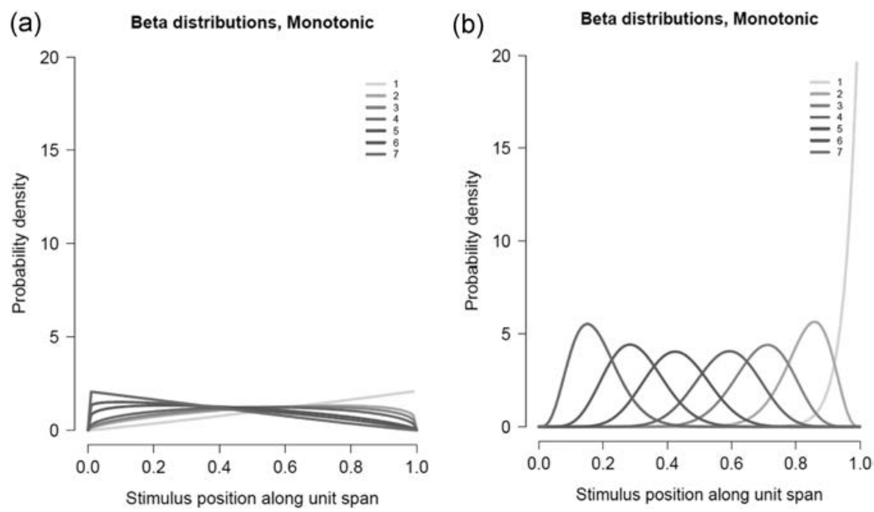

FIGURE 12.—*Beta* distributions at the end of 1 trial (a) and 40 trials (b), for a single simulation in the size condition. Probability density is on the *y* axis, and stimulus ranking (0–1) is on the *x* axis. In graphs (a) and (b), the legends show an increasingly light shade for each stimulus position 1–7.

The larger value of *InhibitionLevel* in the color condition means that more incorrect selections in the backwards direction are made, and the simulated agent is completely reliant on forming as precise an internal representation as quickly as possible; hence the large noisy increment values.

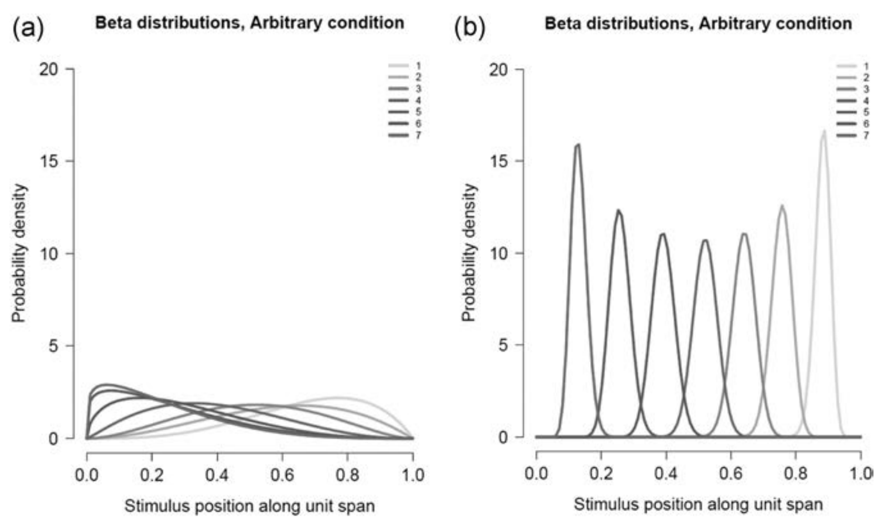

FIGURE 13.—*Beta* distributions within the variable $Beta_n$ at the end of 1 trial (a) and 40 trials (b), for a single simulation in the color condition. Probability density is on the *y* axis, and stimulus ranking (0–1) is on the *x* axis. In graphs (a) and (b), the legends show an increasingly light shade for each stimulus position 1–7.

The precision differences give us a clue as to the division of labor in the information processing going on as the simulations proceed. It appears that the availability of size relationships allows a precise, quantitative view of the benefits that can be derived by the agent explicitly using the size differences between the stimuli in solving the problem. Thus, counterintuitively, maintaining a more precise internal representation does not necessarily allow sequences to be ordered with fewer errors than when maintaining a less precise representation. That is, in the color condition, the agent does not have the benefit of having informative size relationships available "on demand" to help solve the sequencing problem. Thus, although both types of sequence can be solved by the agent's formation of an explicit ranking of stimuli, as indicated by *Beta* distributions in LTM, in the case of the size condition, this representation was only utilized by the agent in proportion to the opportunities for leveraging the size relational properties of the stimulus array. That is to say that in the size condition, the agent was able to offload the expensive probabilistic calculations involved in the heuristic search to a "select smallest interval" rule. This rule allows more efficient computations to take place as LTM representations are not required to be cross-referenced; selecting the correct stimulus based on perceptual information is sufficient.

This unique aspect of the size condition supplies the rationale for the agent to transition into size sequencing spontaneity.

Individual Differences in Learning Speed

We noted that there was variability in the performance of the 5-year-old children, for both the size and color conditions. Such variability is a feature of preoperational performance as noted in Chapter I. To address this, we split the error data into fast and slow learners for the 5-item condition (size and color), on a 50% basis, where fast learners had the least errors and slow learners the most. We hypothesized that this performance difference would be captured by varying a single parameter within the action selection mechanism, the *temperature* parameter of the *softmax* function. For high temperatures, all candidate values within vector $Score_n$ have nearly the same probability; for low temperatures, all candidate values align closely to the actual vector values. Thus, an increase in temperature within this function has the effect of making action selection against a set of values more random; a decrease in temperature more closely aligns action selection to the values in the set. For example, given a set of scores {1, 2, 3, 4, 5}, a very high *softmax* temperature may return set {0.2, 0.2, 0.2, 0.2, 0.2}, and a very low *softmax* temperature may return set {0.01, 0.09, 0.2, 0.3, 0.4}.

We hypothesized that the number of errors and the shape of the error distributions as generated by heuristic models with high temperatures would be like those generated by the poorly performing 5-year-old children. We hypothesized further that the number of errors and the shape of the error distributions as generated by heuristic models with low temperatures would be

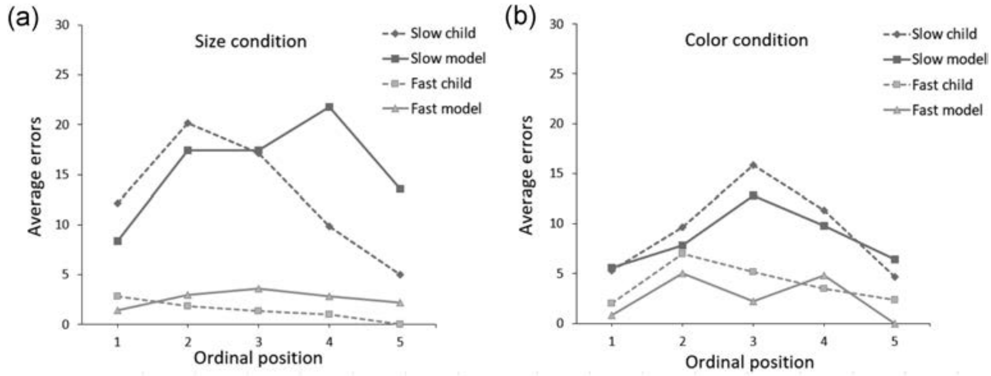

FIGURE 14.—The effect of varying the *softmax* function temperature on the error distributions of the model, within the size (a) and color (b) conditions. Slow (high temperature) and fast (low temperature) models are compared with slow and fast children.

like those generated by the better performing 5-year-old children. We set the temperature for the heuristic models as specified above for the size conditions to 0.5 (low) and 2.5 (high), and for the color conditions to 0.25 (low) and 1.5 (high). We ran six simulations each for the size and color conditions.

For the size condition, the model generated errors that were similar to the errors generated by the children, with no statistically significant difference in error distribution for the high temperature condition: $t(8) = 0.81$, $p = .22$, and mildly significant differences for the low temperature condition: $t(8) = 2.1$, $p = .04$, and again generated a qualitative similarity of error distributions (Figure 14A).

For the color condition, the model generated errors that were again similar to the errors generated by the children, with no statistically significant difference in error distribution for the high temperature condition: $t(7) = 0.37$, $p = .36$, or the low temperature condition: $t(8) = 1.04$, $p = .16$. A qualitative similarity of error distributions was again in evidence (Figure 14B).

The hypotheses were thus confirmed in that the errors generated by a poorly performing 5-year-old (size and color conditions) were similar to those generated by heuristic models set to have a noisier action selection mechanism, and no other difference to the standard models previously reported (which had temperatures set to 1). These individual differences in performance have implications for the speed at which the agent can progress from heuristic to principled search, in that they suggest that the higher *softmax* temperature, the slower such a progression.

Transitional Model Simulation

The aim of this simulation is to provide an answer to question 2. In the transitional model, the agent forms representations that allow it to achieve spontaneous, error-free sequencing via a recursive "select smallest difference" rule in the principled search model. With enhanced WM, the agent can process more information by virtue of its WM remaining open for longer (Ashby, Ell, Valentin, & Casale, 2005). With this extra WM power in place, the agent discovers that the correctly selected stimulus is consistently separated from its referent stimulus by the smallest amount among the set of possible size differences. This discovery is persisted in the form of links between the elements of *Beta* representations in LTM which have achieved stability, being of sufficiently low variance.

For each of two conditions, 10 simulations of 40 trials each of the models were executed. The conditions were *WMAvailability* = 1 and *WMAvailability* = 0.5, which are the probabilities of WM being available for each candidate stimulus at the point of interrogation. The task representation was analogous to that provided to the 5-year-old child participants and applied the same training and feedback. However, we are now simulating a child between the ages of 5 and 7 years old, who has variable WM capacity enabling the discovery of the invariant size relation of "smallest interval" between correctly selected stimuli.

In order to represent the growing knowledge that the correctly selected stimuli are always those with the smallest difference from the one previously chosen, we introduce variable *ThresholdOfVariance* to indicate that *Beta* distributions are of sufficiently low in variance (i.e., pointed, forming a sharper boundary with adjacent distributions), and variable *ThresholdOfLinkStrength* to indicate that the interval relations between adjacent *Beta* distributions are sufficiently strong (i.e., the change in relation from $(Beta_1 ? Beta_2)$ to $(Beta_1 r Beta_2)$). Once the *ThresholdOfVariance* has been passed for a pair of *Beta* distributions (referent and selected), it has stability in LTM, a stability we see as a necessary condition for links to form between two elements of the probabilistic rank. This stability indicates link strengths can be now changed on a successful detection of selected interval "smallness." Once the *ThresholdOfLinkStrength* has been passed for link joining a pair of *Beta* distributions (referent and selected), the "smallest interval" knowledge is now in place creating the unique slots or unit placeholders. The simulation pseudo-code, which represents this process of threshold breaching and slot formation, is as follows:

```
simulation ← 10
trial ← 40
n ← | C_n |
for simulation simulations
 Beta_n ← GenerateBetaDistribution
Links ← GenerateLinks
Slots_n ← FALSE
for trial trials
 SizeDifference ← SCAN (C_n)
 selected ← HEURISTIC SELECT (SizeDifference)
```

If *selected* = *target* **then**
$Beta_n \leftarrow$ ***INFER RANK*** $(Beta_n)$
inhibited ← ***INHIBIT*** (*selected*)
If $WMAvailability_{selected} < RandomNumber\ (0,\ 1)$ **then**
If $Beta_{referent,selected}.\ \sigma^2 < ThresholdOfVariance$ **then**
If $SizeDifference_{selected} = MINIMUM\,(SizeDifference)$ then
$Links_{referent,selected} \leftarrow$ ***INFER LINK*** (*referent*, *selected*)
If $Links_{referent,selected} > ThresholdOfLinkStrength$ **then**
$Slots_{selected} \leftarrow TRUE$
else
$errors \leftarrow errors \cup selected$
next *trial*
next *simulation*

The model parameters were set as per Tables 7 and 8.

Results and Discussion

The error distributions and *Beta* distributions are comparable to those detailed in the heuristic search model results section, as the same parameters as per Table 7 apply to this experiment. For the two *WMAvailability* conditions, Figure 15 shows the end of simulation link strengths for stimuli $L_{referent,\ selected}$.

Figure 15 shows the strengths of the links at the end of the simulations for each WM level. For *WMAvailability* = 1 the link, strengths have reached the threshold, which means that the agent can now use a “select smallest difference” rule to achieve spontaneous size seriation, which is used by the principled search model. This represents a child approaching 7 years old, perhaps ready to transition into expertise. For *WMAvailability* = 0.5 the link, strengths have not reached the threshold, which means that the agent cannot progress to the principled search model. As we noted above, it is very hard to use the empirical data to identify a transitional child for purposes of data comparison. This has to be a virtual participant at perhaps 6 years old, sometimes making

TABLE 8
The Variables and Their Parameter Values Used for the Transitional Model Simulation

Equation	Variable	Value
5	*LinkIncrement*	0.05 + random noise in range (−0.01 to 0.01)
N/A	*ThresholdOfVariance*	0.01
N/A	*ThresholdOfLinkStrength*	0.15
N/A	*WMAvailability*	(1, 0.5)

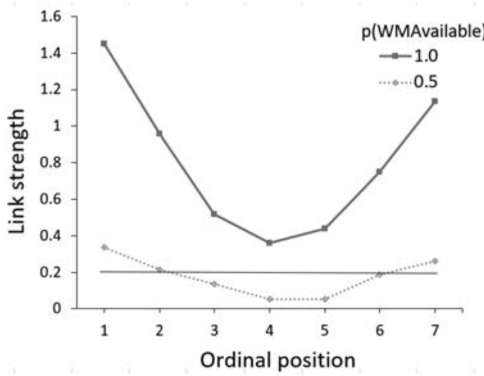

FIGURE 15.—The interstimulus link strength (*y* axis) at each stimulus position (*x* axis) for two different values (0.5 and 1.0) representing working memory (WM) availability probability. The threshold interstimulus link strength of 0.2 is represented by a horizontal line.

discoveries about the utility of using the smallest size relationship, but not having the WM capacity to do so every time. The empirical output from this situation would be very difficult to distinguish from trial-and-error behavior.

Principled Search Model Simulation

The aim of this simulation is to continue to provide an answer to question 2 posed at the start of this chapter. In the principled search model, the agent achieves spontaneous, error-free sequencing via a recursive "select smallest size difference" rule and an LTM representation of the series as a set of unique slots. Given a size sequencing problem, the principled search agent now attends to the smallest difference between each stimulus. This reduces load on WM and reduces dependency on the heuristic search mode, which is inherently noisy. The selections should now be much more accurate, with only variance in the perceptual scanning contributing to errors. Ten simulations of 40 trials each were executed for both conditions in the computer simulation of the model. The task reprecrovided to the 5-year-old child participants and applied the same training and feedback. The simulation pseudo-code is as follows:

simulation ← 10
trial ← 40
$n \leftarrow | C_n |$
for *simulation* simulations
    for *trial* trials
        SizeDifference ← ***SCAN*** **(**C_n)
        selected ← ***PRINCIPLED SELECT*** **(***SizeDifference*)
If $Slots_{selected} = TRUE$ **then**

If *selected* = *target* **then**
    inhibited ← ***INHIBIT*** **(***selected*)
else
    errors ← *errors* ∪ *selected*
next *trial*
next *simulation*

The model parameters are as per Table 7 for the size condition column. To represent an agent that has already learned the smallest difference rule according to the learning described in the transitional model, we changed with the variable *NormalDifference*. σ^2 from 5 to 1.5. This decrease in inter-stimulus variance represents the growing confidence by the agent in the commonality of the correct interitem size differences encountered across the trials. Without a microanalysis of error and stimulus scanning we have no data to precisely inform these values.

Results and Discussion

Minimal errors (10) are generated by the model; those that are present are the result of a small amount of stimulus size confusion on selecting the minimal size difference between referent and target. This can be compared in principle to the very low and randomly distributed errors found in the sequencing data from 7-year-olds, but in both cases the numbers (errors) are too small to make any statistical inferences. The principled search model has thus shown precisely how an agent equipped with a rule derived from the transitional model's operation can achieve spontaneous, error-free sequencing.

General Discussion

In this chapter, we have proposed a set of cognitive models to show how an artificial agent equipped with suitably designed perceptual, learning, and action selection procedures can progress in size sequencing competence. Collectively the models demonstrate how an agent may progress in size sequencing behavior from unprincipled and messy, relying on broad classifications of *BIG* and *SMALL* to inform serial selection, to principled and accurate, using a set of discrete units to inform such selection. In so doing we answered—in the affirmative—some of the specific questions posed at the start of this chapter (we come to ordinal competence in the next chapter). As for the more general criteria that we highlighted, we are now in a position to offer the following answers.

How Do These Models Represent Developmental Progression in the Child?

We hypothesize the following developmental progression. The *heuristic search model* represents a 5-year-old progressing from trial-and-error size and color sequencing, to error-free sequencing. The agent begins with a set of *Beta* distributions that gradually form a ranking, each stimulus being positioned in *LTM* along the unit interval. The speed with which the agent progresses along this developmental path can be decreased by an increase to the *softmax* temperature. However, the heuristic model is not equipped with enhanced WM to allow progression to seriation spontaneity, which is a feature of the *transitional model*. Extra WM power allows the transitional model agent to inspect the candidate size differences after a correct selection has been made, and notice that the size difference that applied to the correct selection is the always the smallest one. The transitional model represents a child between 5 and 7 years of age that by virtue of WM maturation is ready to make this discovery of relational invariance. This discovery is represented by the agent reaching two thresholds, one after the other. The first one is stability of serial order representation in LTM (low *Beta* distribution variance), and second one is a sufficiently high strength of link representing "smallness" between the adjacent parts of this serial order representation. The *principled search model* uses the representations in LTM created by the transitional model, which indicate to it that it can use a "select smallest difference" rule instead of relying on building up representations in LTM, characterizing a 7-year-old child that can seriate spontaneously.

However, computational models must also be faithful, transparent, plausible, and grounded (Mareschal & Thomas, 2007; Yule et al., 2013). These criteria come into focus when real-world data is not available. A *faithful* model abstracts from reality, in that it does not introduce artifacts that are invented by the programmer to allow a better fit to data. A *transparent* model is understandable to those who are going to interact with it, as a reader or as a user. A *plausible* model's mechanism should align with the real-world in terms of psychological and biological plausibility. A *grounded* model makes contact with data and theory that exists in the domain of interest, which, in this case, is developmental psychology. We also note some limitations of the models. To do this, we evaluate the routines common to all models (*SCAN* and *INHIBIT)* and then each of the other routines as apply to the models *(HEURISTIC SELECT, INFER RANK, INFER LINK* and *PRINCIPLED SELECT).*

Common Routine Evaluation (SCAN and INHIBIT)

SCAN represents the visual scanning of stimulus array, two elements at a time, which has long been acknowledged in active machine vision research as an economical way of a biological agent extracting information (Ballard, Hayhoe, Pook, & Rao, 1997). The natural logarithm of the size difference (size condition) or constant (color condition) *Normal* distribution sample is taken, a common procedure in representing perceptual scenarios (Page, Izquierdo, Saal, Codnia, & El Hasi, 2004). Likewise, inhibition on action selection (*INHIBIT*) is a commonly used assumption in executive control research (McGonigle-Chalmers et al., 2008; Terrace & McGonigle, 1994), which we represent by a constant real number as opposed to anything more complex. Although simple, these functions are plausible. We see inhibition on action selection as something that works in individuals of a certain age and have not been more specific than that in terms of its variability. Extensions to this routine could incorporate individual differences where RI might be seen to be an important participant variable (as in ADHD for example).

Heuristic Model Routine Evaluation (HEURISTIC SELECT and INFER RANK)

The *HEURISTIC SELECT* procedure selects stimuli noisily, based on a scoring function [see Equation (3)] that divides the *Beta* distribution precisions in LTM with minimal size differences in WM, and multiples this value by an inhibition factor, which is also stored in WM:

$$Score_{candidate} \leftarrow \left(\frac{LTM_{candidate}}{WM_{candidate}} \right) \times WM_{selected}$$

The scores are put through a *softmax* function, which inform a discrete probability distribution to simulate selection of those stimuli with the highest score. Increasing the temperature parameter makes the resulting probability distribution less aligned to the original values, in effect creating a much noisier set of stimuli from which to choose. This increase resulted in more errors with the heuristic search model and suggests that a generally noisier system at the action selection level is enough to simulate a poorly performing 5-year-old. *Softmax* action selection is widely used in cognitive modeling. Jensen et al. (2015) used it within one of their transitive inference simulations as an action selection routine. Also, we represent each aspect of the agent's memory (WM and LTM) to inform action selection here, which seems intuitively plausible and faithful.

INFER RANK represents the agent's growing representation of serial order in LTM. It defines a set of *Beta* probability distributions that are **all** changed according to a Bayesian inference procedure on every stimulus selection, resulting in a shift of their probability mass along the unit (0–1) interval and ultimately a ranking. The way we carry out this Bayesian inference is similar to the algorithm used by Jensen et al. (2015), but much simpler. Their study compares binary choice data from both rhesus macaque (*Macaca mulatta*) and human adult participants to binary choice model data. They argue that *Beta* distributions afford optimal computational economy in representing stimulus order in transitive inference tasks in that they require a minimal amount of memory to allow efficient learning routines. We agree with this point; such economy in space and time is favored by complex biological systems (Ballard et al., 1997; McFarland & Bösser, 1993). They concede, as we do, that there is no biological evidence for **all** neural representations of a stimulus set being updated on stimulus selection; the updating of all the *Beta* distributions across the set is thus not a standard way of representing learning in cognitive models. As Jensen et al. (2015) argue, the implicit updating of all stimuli in a set via a plausible Bayesian inference routine is a hypothesis to be falsified by neuroscientific data. In support of this, we note that the collective updating *Beta* distributions, as per Equations (4a) and (4b), is mathematically necessary for them to form a rank. Lastly, this model was validated by child data, generating errors that were similar in their distributions across item position to the corresponding empirical condition (size vs. color).

Transitional Model Routine Evaluation (INFER LINK)

INFER LINK runs when WM robustness (a variable in the range 0–1 representing the probability of availability of a list element) is increased, resulting in the agent having more time to inspect interstimulus relationships. This variable represents a varying level of WM maturation. *INFER LINK* represents the joining up of an existing ranking of *Beta* distributions in LTM, each link representing the "smallest interval" property of a correctly ordered size-related set. Our representational structure aligns to the spreading activation networks proposed by Cooper and Shallice in their DOMINO model of executive control (cited in Yule et al., 2013). Moreover, the explicit manipulation of the numerical values to change graphical link strengths with real values was used by Cooper (2016) to represent number bias within a model of random number generation. More generally, we see our LTM representation as consistent with the evolving directed graphical structures used by Cooper and Shallice within the DOMINO model to represent routine, unconscious action selection. Notably, the WM availability representation is highly simplified, the concepts of capacity and time collapsed into a single variable and mapped to each list element. However, we argue that the transitional model is nonetheless faithful and plausible, especially as it is extensible to representing principled selection and ordinal competence.

Principled Model Routine Evaluation (PRINCIPLED SELECT)

The *PRINCIPLED SELECT* routine represents the operational output of the transitional model. Here, the agent uses the links formed within the transitional model to rationalize rule usage. Rank variance and link strength thresholds having been reached and available to the model, the "select smallest difference" rule can be executed. The rule allows the agent to use online sources of perceptual information to solve the size sequencing problem, without having to rely on offline LTM representational look-up. This representational interplay is consistent with models of executive control, both cognitive (e.g., DOMINO; see preceding section) and robotic (Holland and McFarland, 2001; McFarland & Bösser, 1993).

Continuous or Discontinuous?

What we have presented is a **discontinuous** model of representational change, in that the information-processing routines within the principled search model are new skills, deriving from discoveries made about the utility of using the smallest available interstimulus relationships in the transitional model. This discovery is facilitated, but not caused, by a continuous increase in power to WM in the transitional model. As WM is limited in capacity and subject to decay as time progresses (Ashby et al., 2005), it is proposed that an agent who has already learned to sequence via trial and error will seek out and use the "select smallest size difference" rule. Our discontinuity position is consistent with the position taken by Case et al. (1993). They argue strongly for developmental progression involving the interplay between WM and the learning of complex executive processes: "the size of children's mental power or working memory … sets a limit on their ability to coordinate their existing schemes into a more sophisticated pattern" (p.158). Arguably, the enhanced robustness of WM allows more information to be managed by the "transitional" child prior to action; as a consequence, they can cope with more complex information processing.

Do the Models Align to Reality?

We used error patterns generated by the 5-year-old child as a source of validation data for the heuristic search model and were unable to do likewise for the transitional and principled search models. The transitional model is a hypothetical mechanism representing the perceptual learning of a child between the ages of 5 and 7 years. Although some of our faster learning 5-year-olds may have been in this category there is no way to determine this from the data. Similarly, the principled search model has no validation data set in this case simply because 7-year-old children generated so few errors on these tasks. It could thus be argued that evidence for the models is weak because its validation depends on a combination of error-matching statistics and the simulation of an expert agent who avoids error through the deployment of an algorithm.

As the lack of a rich validation database is always likely to be a problem with modeling spontaneous behavior, it is important to have other criteria to evaluate how well a model maps on to reality. For these, we would propose faithfulness, plausibility, and transparency (Mareschal & Thomas, 2007). The model we propose is grounded in psychological theory and has been validated via a modeling approach used in a separate domain. This validation applies to the Bayesian inference algorithm that we used, and also to competing machine learning models that it allows us to rule out. Specifically, experiments that compared the performance of several machine learning models to that of rhesus macaque (*M. mulatta*) and human adult data on transitive choice tasks were carried out by Jensen et al. (2015). They rejected a number of models that did not fit their data, such as a reinforcement learning model, for example, favoring instead a model they called "Beta-Sort." This was a Bayesian inference algorithm conceptually similar to the learning algorithm used in our heuristic and transitional models. Limitations notwithstanding (see the following section and also Chapter V), we see this coherence as providing validation for our models.

Limitations of the Modeling

The modeling described above has limitations in how faithful it is to a biological system which perceives, learns and acts in real time, and is serial as opposed to parallel in design and implementation. Indeed, RTs were not modeled, only errors, and only within the heuristic search model did they serve a validation function. However, the models should be seen as blueprints for yet more faithful and plausible ones. Current cognitive robotic architectures could represent the agent's control equations in addition to representing a perceptual-motor loop in real time (De La Cruz, Di Nuovo, Di Nuovo, & Cangelosi, 2014; Sandamirskaya & Schöner, 2010). There is also coherence with the probabilistic dimension of our models and aspects of the *predictive coding* view of neural functioning (Clark, 2016). We shall return to a fuller discussion of limitations and potential improvements in Chapter V, as these points also apply to the ordinal search model that will be presented in support of the ordinal competence theory proposed in Chapter IV.

IV. The Development of Ordinal Size Understanding and a Computational Model

Introduction

We have looked at the size-ordering abilities of children using sets of five and seven items. We now ask how easy or hard is it for a child between the ages of 5 and 7 years to identify individual items within such sets by ordinal size, that is, identifying which is, for example, the biggest, smallest, second biggest, and so on? Seen simply as a matter of counting from the biggest or smallest item in a set, reaching the requisite number would be well within the capabilities of a preschool child (Fuson, 1988). The identification of the ordinal size of items within a set of seven items requires a count to four at most to identify each and every item uniquely. But it is precisely the ability to initiate a principled search from one or other end-point of a series that we have found wanting in 5-year-old children. Indeed, on the Piagetian position that the concept of ordinality is bound up with the ability to seriate, we would expect no greater success in ordinal identification than in spontaneous seriation among children younger than about 6 or 7. As described in Chapter I, Piaget's serial and ordinal correspondence tasks showed difficulties in these younger children, when, for example, a doll of a given size within a jumbled set of dolls had to paired up with a walking stick of the equivalent ordinal value within a jumbled set of toy walking sticks. He showed that numerical counting to establish equivalence in this situation is unlikely to be performed either spontaneously or correctly by preoperational children. The experiments described in Chapter II also suggest that the spontaneous size sequencing required for ordinal identification is not reliably found until the age of 7.

The computational model outlined in the last chapter also predicts difficulties in ordinal size identification until the age of spontaneous seriation. It showed that the insights that could lead to principled size sequencing would also be necessary for the application of a possible search-and-stop procedure to enable ordinal matching. But as the model may require an additional processing step to this procedure, principled searching may be necessary but not sufficient to enable ordinal matching. On the grounds that it may actually require more than seriation in its execution and development, we are aware that ordinal identification may not be the inevitable accompaniment to seriation success that Piaget suggests. To be clearer about this relationship across the two skills, we need to test ordinal abilities directly using similar materials and methodology to those described in Chapter II. We report our experiments on ordinal understanding in this chapter.

The literature beyond Piaget's own research on this topic is surprisingly sparse. Ordinal understanding of size is a highly underresearched topic despite its obvious connection with counting, numeracy (number understanding), and measurement. As noted at the start of Chapter I, the connection between counting and ordination is the concept of a unit. The estimation of amount through counting or direct measurement requires unit iteration. If children are unable to differentiate sets of discrete sizes such that they can distinguish every ordinal value, then we would have to question the level of their understanding in related domains. For example, despite early exposure to the activities of counting and measurement in the early school years, there may be a conceptual limit on what children can derive from these activities. In the introduction to this chapter, we trace what is known about the development of ordinality, explain how research on ordinality raises questions about the understanding of unitization, and why, despite its centrality, the study of ordinality has nevertheless fallen by the wayside of mainstream research.

Relational Discrimination Research

In a tradition entirely unconnected with the Piagetian one, considerable debate raged in the early part of the 20th century over whether the human perceptual system was geared to processing stimulus relations or absolute values. One method employed to study this question was a task known as *transposition*. Put simply, this was a relational discrimination task where the participant (a child, monkey, or other animal) was trained to always choose, for example, the larger of two sizes whatever their absolute values. Following training, the test set involved new size values and the question was whether or not the trained relation would be "transposed" (a relational response) or whether it would adhere as far as possible to the original stimulus values (an absolute response). The central argument among theorists was whether transposition behavior supported a behaviorist or a more "cognitive" account of perceptual learning (see Reese, 1968, for a review). The behaviorist explanation was founded on the proposition that stimulus choice results from learning to attach a response to the "energic" properties of the stimulus, independently of any other (Spence, 1936). In the case of training on size relations, this would amount to learning that a particular absolute value (e.g., 3 cm) is always correct. This seemingly simple question was made endlessly complicated owing to evidence of limited transposition in humans and nonhumans alike. As far as children were concerned, the results were equivocal and varied depending on the age of the child and numerous other variables. Behaviorists such as Spence were able to account for all cases of limited transposition in terms of the net excitation of stimulus-response attachments. Developed mainly within a comparative context, the behaviorist account was particularly challenged in terms of whether it was applicable to children (Kendler, 1975), and even Spence acknowledged that a species capable of verbal mediation would respond relationally. Bryant (1974) argued, furthermore, that absolute responding in children was an artifact of the way the stimuli were presented and that child participants were actually responding to the relationship between the stimuli and the background frame on which they were displayed. The debate became finally settled, at least for human and nonhuman primates, when there was incontrovertible evidence that absolute values are actually much harder to compute than simple binary relations (Lawrenson & Bryant, 1972; McGonigle & Jones, 1978; Thomas & Crosby, 1977).

The relevant point here is that in the course of this debate, some early investigators trained children on the middle-sized relation as well as on bigger and smaller. The primary objective was usually to see if it, too, would be transposed to new values, but the studies revealed quite startling difficulties by children up to the age of 7 in simply learning in the first place to select the middle-sized from a set of three differently sized objects (Yeh, 1970; Zeiler & Friedrichs, 1969; Zeiler & Gardner, 1966). Buried within the depths of behaviorist research, these findings were reprised in a supposedly neo-Piagetian context by Siegel (1972). Described as seriation, this was in fact an ordinal size training experiment with children, using nonverbal learning methods. In this experiment, different groups of children between the ages of 3 and 9 years were given one of various relational discrimination tasks using vertical bars as stimuli. The absolute sizes of these items varied across trials. Siegel used the standard reinforcement methods of the time where selecting the correct stimulus was rewarded with candy. Each child was given a 2-, 3-, and 4-item set but had to learn only one of the possible ordinal values within each. The most demanding task turned out to be a 4-item series in which the trained size was second smallest. Correctly selecting the inner positions (middle-sized for 3-item sets and second smallest in the 4-item set) was harder for all participants than for participants who had to select the end-points, biggest and smallest. Even 6-year-old children required around 40 training trials to identify the middle-sized within 3-item sets and the second smallest within 4-item sets, but there was a dramatic drop in errors by the age of 7.

Two notable features of the discrimination learning studies cited above is that single size relations were allocated to the participants independently of any other size within the training set, and the second is that the stimulus sets were not seriated either for or by the child as part of the task. A single contrastive case is provided by Brainerd (1978) who found that children younger than 6 years could match a particular size, out of five, to a numeral from 1 to 5, but the selection made was from was a highly discriminable seriated array of five objects. This last finding raises the very issue addressed by Piaget, which is how does ordinal size understanding relate to seriation? They may be completely interdependent, as Piaget argued, or sequential ordering could be the prior and separable foundation for the other, as we shall argue later in this chapter. We start by restating Piaget's position and then considering the evidence as it stands.

Ordinal Understanding in the Piagetian Literature

For Piaget, transitivity, seriation, conservation of amount, and class-inclusion tasks tapped the child's ability to coordinate logical and spatial relations, and were assumed to require reversibility of thought. Although, as noted in Chapter II, this has proved to be an elusive concept in the development of seriation (Leiser & Gillierion, 1990), Piaget's use of the concept provided a persuasive argument defining the true grasp of ordinality. Where the objects can be ordered by size, the ordinal position would not just mark its number but also its rank within an apparently coordinated set of reversible relations where the child becomes aware that "a given element, say *E*, is both longer than those already in the series ($E > D, C$), and shorter than the ones yet to follow ($E < F, G$)" (Inhelder and Piaget, 1964, p. 257). Spontaneous seriation at that age was thus said to follow from the perceived necessity that every item must have a unique position within the set as a whole. This unique ordinal value was argued to be understood as interchangeable with its cardinal value by operational children in that "they understand that the Nth position corresponds to a cardinal value N which is at the same time greater than that of the elements A…(N-1) and less than the elements (N+1…..T)"—where T represents the final element (p. 134). A corollary to this is that every ordinal position of an object in a given set, will also be seen to be logically equivalent to an item occupying the corresponding position in another set—even when the sets are composed of different items and are not spatially aligned.

Two classes of task emerged from this analysis. One was based on numerical correspondence alone, where asymmetric relations of difference played no part. But where understanding ordinal value was being measured, seriation (e.g., ordering by size) had to occur for those values to be determined. The first class generated the number conservation task, which received considerable attention in the subsequent literature; the second generated tests of ordinal size identification, which received almost none.

Number Conservation

In Piaget's classic number conservation task, children are asked whether identically numbered sets of objects laid in matching rows one above the other have the same number after one row is spaced out (Piaget, 1952b). Piaget saw the ability to conserve number in the face of this transformation as a triumph of operation over perceptual intuition. Although it does not require the understanding of an ordinal value within the sets being compared, he sees both abilities as related to the understanding of cardinality, such that "an element 'n' shall be seen to be permanently after the (n-1)th and before the (n+1)th" (Piaget, 1952b, p. 155). The permanence of this cardinal value should be impervious to spatial layout. Any numerical estimate that took account of a perceptual layout or transformation was, therefore, and by definition, a failure to understand the true nature of a cardinal number. In

Piaget's conservation experiments, he exposed the conflict experienced by preoperational children when they could establish cardinal correspondence through overt counting, but failed to judge equivalence if the objects counted were visually displayed in a manner suggesting that one had a greater amount. Stage II children (around 5–6) he describes thus:

> *Every one of these children concludes that there is equality if the same number of elements is dropped, one at a time, into two containers irrespective of the shape of the containers. But when the child afterwards, considers the result obtained when the shapes are different, his belief in the equivalence is shaken by an evaluation based on the perceptual relationships.* (Piaget, 1952a, p. 32)

Piaget's explanation of how "lasting equivalence" is achieved involves a rather convoluted argument that recruits the notion of complementarity between classes of like elements, that disregard differences, and asymmetrical relations that disregard equivalences:

> *Finite numbers are therefore necessarily at the same time cardinal and ordinal, since it is of the nature of number to be both a system of classes and of asymmetrical relations blended into one operational whole.* (p. 157)

Without straying into Piaget's argument about the relationship between the logic of classes and the logic of asymmetrical relations, the key point here is that there should be an age at which children will not only conserve number, but also be able to solve class-inclusion problems. Successful performance on conservation and class-inclusion tasks in children of operational age appears to be fairly robust in that there are no disclaimers regarding children around seven and older. The claims regarding younger children on both tasks, however, have been persistently challenged. With regard to number conservation in particular, the staged progression of preoperational aged children as detailed by Piaget has been subject to revision in the light of subsequent task variation (McEvoy & Mona O'Moore, 1991; Siegler, 1995). This can be due to the number and/or familiarity of the items used (Gelman, 1972; Hanrahan, Yelin, & Rapagna, 1987; Kahn & Garrison, 1973) or the way in which the test question is put to the child (Neilson, Dockrell, & McKechnie, 1983). When younger children fail, moreover, some have argued that this is not through a lack of understanding but simply through their choice to use a more direct strategy for judging equivalence (Bryant, 1972; Gelman, 1972). Despite this, Piaget's assertion that number conservation becomes reliable and spontaneous between 6 and 7 years of age is still uncontested.

Ordinal Size

The other class of tasks bearing on ordinal understanding was specifically focused on the organization of asymmetric relations such as size. This has been a relatively neglected area of research with the single exception of transitivity research as discussed in Chapter II. As described there, transitivity testing never really extended into the comprehension of explicitly seriated sets and the corresponding ordinal size relations such as second biggest, middle-sized, and so forth. Piaget's own tasks were more wide-ranging focusing on the ordinal understanding of internal positions that accompanied single set seriation, and also on tests of ordinal and cardinal correspondence of individual items across two sets. The first of these was based on single element insertion

Insertion tasks. For single set seriation, Piaget's test simply followed the classic seriation task, where, as described in Chapter II, children were presented with a set of different-sized uprights sticks or rods with which the child is invited to "make a staircase." The children were then invited to place a removed element (stick) into its correct place in in the ordered set. Piaget claimed that this is achieved at the same time as spontaneous seriation, though detailed group data are not provided in his accounts (Piaget, 1952a; Piaget & Inhelder, 1974). Preoperational children—below the age of 6–7 years, and who could only seriate in a trial-and-error fashion—were reported to succeed on insertion tests in a global fashion provided there was a rough correspondence between the small and large items on either side of the target item. Characteristically they could not withstand any contraction, expansion or reversal of one the series—as with number conservation.

Although Piaget's claimed age of success was confirmed in subsequent investigations (Elkind, 1968), the data were reported in terms of stages of success rather than broken down by age and task factors. Elkind nevertheless reported that by Stage III (his oldest participants were aged 6 years) the children were capable of rapid and errorless insertions. Since then, there has been almost no further research into this simple insertion task. Though perhaps surprising in itself, this is unsurprising in the light of the fact that size seriation itself has been a relatively neglected area, as we reviewed in Chapter I. One exception is a modified replication of the insertion task has been reported by Blevins-Knabe (1987a), however, in which a complicated set of computerized tasks was presented to children aged between 3 and 6 years of age. Several subtasks were presented (six of which were two-stick tasks) and involved placing a correct end or middle stick. Two tasks presented six sticks, one of which required inserting an end stick. Other tasks (with two or seven sticks) actually gave the child the correct placement but asked them to choose the correct length out of two alternatives. The subtask closest to Piaget's was the six-stick internal insertion placement but in fact, this was not presented to most 5- and 6-year-olds for the perhaps curious reason that they had passed two-stick tasks. The reports of group success (50% or more passing) were, for 5- and 6-year-olds, being able to center-insert between two sticks, placing an end stick beside two sticks, choosing the correct length of a middle stick between two sticks and choosing the correct length for an end stick. By failing to explore beyond middle-sized in a 3-item set, there is, therefore, nothing in the insertion data by Blevins-Knabe to contradict Piaget's claim that it is difficult for children under the age of 7 to identify the internal ordinal positions within a single set of multiple items. Another rare study on item insertion was carried out by Pasnak et al. (2015) but it was primarily designed as an educational intervention aimed at improving sequence pattern recognition and core results on the insertion task are not reported.

Identifying an element. A second method of measuring ordinal size understanding was delivered in the context of one-to-one correspondence tasks as described in Chapter I. Typically, children would be presented with (e.g.) 10 wooden dolls differing regularly in height and a set of 10 toy walking sticks but representing a smaller size range and asking them to match the dolls to the sticks by various methods. This might be by arranging them so that each doll can find its stick (implicitly inviting seriation as a preliminary step) or simply identifying a doll and asking which stick it should take.

As part of this series of experiments, Piaget used one that demonstrated (to him) the interrelationship between seriation, ordination, and cardination. In this task, a staircase constructed by the child was disarranged when a doll was interrupted on a particular step when climbing to the top, and the child had work out how many steps it had climbed and how many more steps it had yet to climb to reach the top. For Piaget, success on this test indicated how far children could understand the relationship between ordinal and cardinal number:

> *the best evidence of their understanding of the relation between ordinal and cardinal number is to be found in the fact that once the steps already climbed have been reseriated and counted, the children feel no need to reseriate the remainder in order to discover how many steps remain to be climbed.* (Piaget, 1952b, p. 134)

The cardinal value of the "remainder" is described as "represented by the subtraction T-N or (A…T) – (A….N)" (p. 134) where N is the ordinal value and A and T represents the ends of the series.

As noted earlier, Piaget (1952b) was assuming that the ordinal position was encoded as a cardinal value where the "element n represents for the child both the nth position and the cardinal value n" (p. 114). To establish the ordinal and cardinal value of an item, therefore, it is necessary in these tasks to count from one or other end of the series as well as ensure that the items are seriated. Yet counting to establish an ordinal size is a topic that seemed to fall from favor along with the general interest in size seriation, despite the interest in counting behavior in the context of number conservation (Fuson, Secada, & Hall, 1983; Russac, 1978). Number conservation itself can be solved simply by the observation that nothing has been added or taken way, whereas ordinal size understanding as measured through correspondence **has** to involve spontaneous counting. All the more remarkable, therefore, that there has been so little subsequent research bearing on the ability to understand ordinal size using counting. An exception to this is a study by Kingma (1983b), which wardrobes had to be size-matched to the appropriately-sized puppet but as every wardrobe came to each puppet's shoulder, there was a nonordinal and noncounting basis for making the match. Counting as a skill in its own right, by contrast, is at the center of the vast burgeoning research on children's mathematical development as noted in Chapter I.

Ordinal Position Within a List

On the supposition that children do need to perform a spontaneous count to solve an ordinal size task, what do we know from the literature about this ability more generally? That is, when, using their count abilities, can children can accurately identify the ordinal position within a numbered list (e.g., 4th, 5th, etc.)? A small number of studies have compared the acquisition of the cardinal properties of numbers with their ordinal properties and these indicate a selective difficulty with the latter for children up to 5 years old. Using a variety of tasks, Fischer and Beckey (1990) found that whereas more than 80% of a group of ninety-seven 5-year-olds could count and make sets consisting of seven items, only 31% could point to the "third" in a row of toy cars and only 25% could name the yellow car as the "fifth." When asked to specifically order four cards in order of the number of dots they contained (two, four, five, and seven), only 23% succeeded. In a later study, Colomé and Noël (2012) used a variety of "give me" and "tell me" tests to study the relationship between cardinal and ordinal understanding of number with children aged between 3 and 5 years of age. For example, using a set of 10 toy cars queuing at a traffic light, they asked children to "give me the (e.g.) third car." Five-year-olds were 67% correct for smaller numbers, three and four, and 59% correct for larger ones (six and seven). Half the time the experimenters denoted the items by cardinal name ("car six") but half the time by the ordinal term ("sixth"). Although the authors point out that children rarely use ordinal words for larger numbers, there was no indication in their results that the children selectively failed to understand the ordinal terms. Whatever terms used, the children were worse at these ordinal tasks than the cardinal ones—in which a specific number was requested ("please put six cars in my garage"). The authors concluded that ordinality lags behind cardinality, at least in these tasks, and seems to be a gradual process that needs to be better characterized by new studies. New studies have not been forthcoming, however.

Ordinality and Seriation: How are they Connected?

The literature as far as it goes thus predicts that ordinal size calculation is as unlikely to be a spontaneous ability in 5-year-olds, as is seriation. If so, the question is "why?" On Piaget's account, it is because children of this age do not understand reversible operations. We, by contrast, having modeled spontaneous seriation would argue it is because the application of unidirectional search for a minimum distance interval is a prior and not yet emergent cognitive procedure. These are positions that are difficult to distinguish but we aim to do so by the end of this monograph, as they have important implications for cognitive growth. In the meantime, there is much we need to discover empirically about the development of ordinal size understanding as it has been so neglected in post-Piagetian research. Motivated by a concern to discover this purportedly related aspect of size seriation we examined ordinal size understanding in 5- and 7-year-olds using a variety of tasks, but based on the same stimuli as used in the touchscreen seriation tasks. This chapter describes our tasks and findings.

General Methodology

Our general methodology was as described in Chapter II for our size sequencing experiments. Training was used to obtain a measure of acquisition under conditions of touch-by-touch corrective feedback as before. In our earlier studies, a correct sequencing trial was error-free adherence to an ascending or descending sequence (of five or seven items) and was indicated by a fanfare and a man ascending a ladder. For the ordinal tasks, a correct trial was the selection of the appropriate ordinal position, and the man ascended the ladder following a trial block of five correct responses. The trial block usually represented all five ordinal positions. Further details are presented for each of the studies that follow.

The design of each task evolved partly as a consequence of what we discovered in the prior study. The first report describes a matching task in which a target ordinal size (e.g., second smallest) within a set had to be matched to the corresponding item from a set of samples.

Experiment 4

Rationale

The question addressed in this experiment was whether children could be trained on a touchscreen to specifically identify the ordinal sizes of the items used in the seriation tasks described in Chapter I. As with seriation training, we started with a modest set size of five items.

Method

Participants

These were 24 children drawn from the same population as the seriation studies. Twelve of the children (seven girls and five boys) were aged between

5;1 and 5;8 (*Md* = 5;3), and twelve (four girls and eight boys) were aged between 7;1 and 7;11 (*Md* = 7;7). As before, all of all children had English as their first language.

Task and Stimuli

The stimuli were the two size ranges of rectangular stimuli as used in Experiment 1. On this task, both size sets were presented in evenly spaced randomized arrays (with bottom edges aligned) one above the other, placed horizontally exactly as they had been for seriation training. Each samples array was of a uniform color but always different from the target array. The child was trained to select from the lower (target) display the item that was the ordinal size match of a blinking sample in the upper one, that is to match the middle-sized in one display to the middle-sized in the other on one trial; the second smallest to the second smallest on another, and so on. A correct touch was registered by a bleep; an incorrect touch by a buzz (only icons in the lower array were touch-sensitive). Following an error, the arrays remained on the screen and the sample continued to blink until a correct touch. The arrays then disappeared for 2 sec, and then reappeared in a new configuration and new colors, for the next trial and the next sample. See Figure 16 for an example.

Half the children were given the larger sizes as the sample set and the smaller sizes as the target set, and the reverse for the other half of each group. Every size was used as a sample once each within a randomized trial block of five. Criterion was set at 4/5 completely correct trial blocks within an upper limit of 40 trial blocks, and TSCR (now) refers to the trial on which the first successful trial block of 4/5 began.

A short pretraining task was given, using just two sizes and the relations bigger and smaller, until 4/5 correct choices had been made. Once the child started on the experiment proper, a display showing a man ascending/descending a five-rung ladder was used to encourage learning, whereupon he climbed a rung on correct completion of all five trials within a trial block. When criterion was met, the man reached an apple on a tree and a fanfare was played.

At the completion of the task, children were asked how many squares were in the top row and if they could describe their sizes.

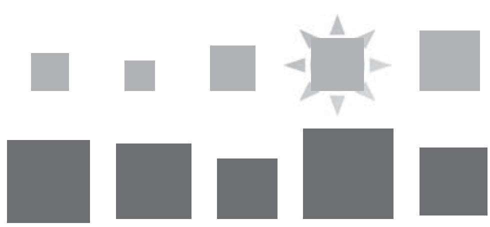

FIGURE 16.—An example of the layout for the MTS task using a 5-item set (Experiment 4).

Results

Analyses

The choice data were analyzed for age effects in terms of numbers of participants meeting the learning criteria, and the learning profiles for successful participants. RT analyses for successful participants are reported where relevant. Non-parametric tests were used where presumptions of normality of distribution were violated. Effect sizes were calculated using Cohen's d for F values and r for Wilcoxon's Z, where $r = (Z/\sum N)$.

Choice

Four of the twelve 5-year-olds met criterion, whereas all 7-year-olds did so. The mean TSCR for the successful 5-year-olds was 20.5 ($SD = 4.1$) and 7.4 ($SD = 6.6$) for 7-year-olds. If these data are collapsed into categories 1–20 and 21–40 trial blocks, the distributions are significantly different on a chi-squared test, $X^2(2) = 88.36$, $p = .001$, $r = 1.92$, a large effect.

The distribution of trials by age group and item position is shown in Figure 17 (left). One 7-year-old was deemed to be an outlier from the rest of the group in requiring more than 2 *SD*s from the mean and is depicted separately. A Kruskal–Wallis one-way ANOVA found no significant difference between the distributions for the successful 5-year-olds and the outlier 7-year-old, H (1, $N = 5$) $= 1.32$, $p = .25$.

A Scheirer–Ray–Hare two-way non-parametric ANOVA was used to compare the four successful 5-year-olds with the remaining 7-year-olds ($N = 11$). A large main large effect was found for age, $F(1,71) = 42.48$,

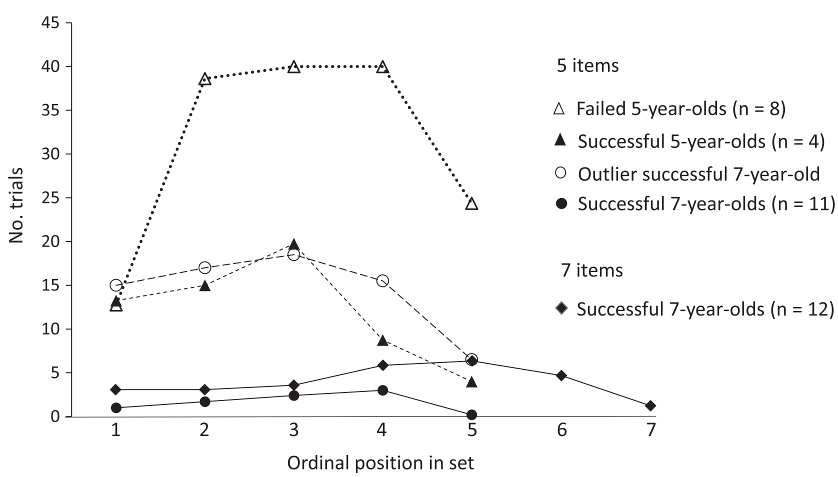

FIGURE 17.—Learning functions for the 5-item and 7-item Match-to-Sample tasks (Experiments 4 and 5) depicting the average trials to the start of criterion run across each item position, where position 1 represents the smallest item for both sets.

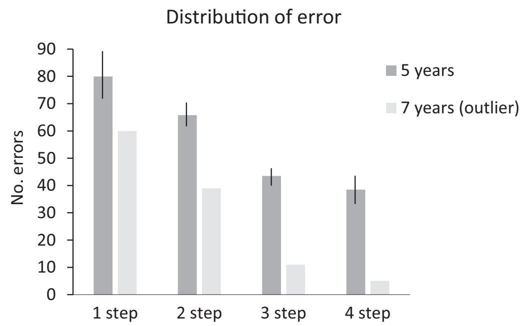

FIGURE 18.—Distribution of error across item position for Match-to-Sample task (Experiment 4). Standard error is represented by vertical lines.

$p = .010$, $d = 4.08$. There was no interaction between age and item position, $F(1,71) < 1$.

A more detailed analysis of the 5-year-olds' error was carried out to explore the source of the difficulty as far as possible and in particular to determine if it was restricted to the positions immediately adjacent to the target. The total error accruing to each position was calculated for the last 20 trial blocks (to reduce the effects of initial noise) and is shown in Figure 18, from which it can be seen that although it is highest around the adjacent items (one-step) there is a widespread distribution of errors for all participants for whom error was within measurable limits.

Verbal Report

The answers to questions were tabulated in a similar way to that described in Chapter II following the seriation task but with the omission of a question about order. They were sorted into (a) degree of unique specification, (b) number differentiated by linguistic label, and (c) number by answer (how many were there?). As before, labels that were not specific, (e.g., “a wee one and a small one” or “a middle-sized and a medium one”) were scored as (a) specifically identifying only one object uniquely, but two by number of labels (b). Thus, for example, “smallest and nearly the smallest” was scored as identifying two in both (a) and (b). The criteria for (a) took the context into account. If child used the indefinite article (e.g., “a big one”) but did not use the term “big” to apply to any other, it was scored as identifying one item uniquely.

As noted in Chapter II, scores deriving from the above experimenter judgments are essentially categorical rather than continuous, posthoc and somewhat open to subjective interpretation, but presented as an arithmetic mean across categories, they offer approximate idea of verbal accuracy. Scores can vary around the ideal mean of 5 for all categories. It can be seen from these averaged scores in Table 9 that successful 5-year-olds out-performed the failed participants on all counts but were still far from accurate.

TABLE 9
Mean score from verbal report of 5-item Match-to-Sample (Experiment 4)

Age	a	b	c
5 years failed ($N = 8$)	2.50	2.00	3.75
5 years successful ($N = 4$)	3.75	4.75	5.75
7 years ($N = 12$)	3.92	4.92	5.33

Note. a = unique identifiers; b = separate identifiers; c = number estimate.

Only one child, who was successful, scored 5 on all criteria. No child used numerical ordinal terms such as "second" or "third," but seven used the terms "middle," "medium," or "middle-sized." One unsuccessful child simply classified into "big ones and little ones." Two out of the eight failed children gave the correct number estimate as did two of the four successful children.

Two of the 7-year-old children used the terms "second" or "third," but both overestimated the count by one. Eight children from this group used the terms "medium" or "middle-sized." Only seven of the twelve children gave correct estimates of the number in each set. The incorrect number estimates varied from "four" to "seven." One child mentioned his estimate of two absolute sizes (7 and 10 cm). No child spontaneously mentioned counting the sizes.

Discussion

The main finding from this experiment was the extraordinary difficulty shown by most 5-year-olds (with one exception) as compared with the 7-year-olds, and also as compared to the relative ease with which children of this age learned to seriate (Chapter I). However, the two findings are completely reconcilable on the assumption that the MTS task presumes spontaneous seriation as a necessary if not sufficient condition. In order to find an ordinal match to the target in the MTS condition, a systematic seriation strategy would have to be deployed working from one or other end-point inward, together with an elementary count. The difficulty shown is entirely consistent with Piaget's findings from his correspondence tasks with children of 5–6 years of age: "he never fails to find the correct corresponding element for the two ends of the series. But as soon as one of the middle elements is chosen… the child is lost" (Piaget, 1952a, p. 112). This, according to Piaget, is because seriation is not yet operational. Without resorting to the circularity of Piaget's argument (whereby the seriation is not "operational" because it lacks this corollary skill) we would say that linear seriation is simply not a spontaneously deployed procedure at this age as clearly indicated by Experiment 1. Without such a readily available strategy, some other heuristic needs to be used. It is hard to determine what this is, given the very high spread of error across the set as a whole, and the best we can say is that children who struggle with this task seem to resort a more ad-hoc strategy. This is possibly influenced, moreover, by the fact that they were having to deal with two size ranges in identifying the sample and finding the match. The verbal labels used by the children to describe the items is consistent with the confusion surrounding the internal items, insofar as they were able to fairly accurately identify the end-points but with high lack of specificity regarding the other three—even for those who succeeded. Although more affluent, the linguistic descriptions offered by 7-year-olds were far less specific than suggested by their highly accurate performance, and so extrapolations from verbal protocols have to be circumspect.

Experiment 5

Rationale

The disparity between 5- and 7-year-olds in the seriation studies described in Chapter II was enhanced by extending the set by just two items, a factor that made almost no difference to the performance of older children. The original intention here was, in a similar manner, to test all children on an expanded set of items now using the matching task, but the length of training and difficulty that the MTS task induced in 5-year-olds rendered that an impractical objective. Seven-year-olds alone were given a 7-item set using the rod-like stimuli for the 7-item seriation task of Experiment 3. They were asked about the number of items in the top row and if they could name their sizes after completion.

Method

Participants and Procedure

Selected from the same population as above, the participants were twelve children (four girls and eight boys) aged between 7;4 and 8;0 (*Md* = 7;6). As before, all of all children had English as their first language.

The method and procedures were exactly as described for Experiment 4, except that the rod-like stimuli as depicted in Figure 1B (Chapter II) were used. From the seriation transfer tests described in Chapter II with another group of 7-year-olds, we had no reason to believe that the new stimuli would be a source of discrimination difficulty.

Results

Choice

All children reached criterion with an average of 8.4 ($SD = 4.5$) trial blocks to the start of the criterion run. There were no outliers in the data set.

Whereas the performance by 7-year-olds showed evidence of learning, there was insufficient error data to make a meaningful calculation of learning difficulty by ordinal position, as around half the data points were a zero (spontaneously correct). The error distribution across items is shown in Figure 17.

An unpaired t test showed that performance by 7-year-olds in terms of TSCR was not significantly different from the performance of 7-year-olds in Experiment 4, $t(22) = 0.42$, $p = .678$.

Verbal Report

The verbal posttest was tabulated as for Experiment 4 and showed high similarity to the scores for 7-year-olds following Experiment 4, with a mean category score of 3.75, 4.92, and 6.00 for categories a, b, and c, respectively. The most obvious point to note is that the language used by 7-year-olds does not reflect their level of expertise on the task. Only one child described the set in a way in which a listener would be able to reconstruct it completely. Seven children gave the correct number estimate. The incorrect estimates ranged from “five” to “nine,” although five children spontaneously mentioned counting. Six children used numerical identifiers such as “second biggest” or “third smallest.” Seven children specifically mentioned the middle-sized item; four naming it as such, the remaining three as the “fourth smallest.”

Discussion (Experiments 4 and 5)

The results from Experiment 5 confirm that there is a radical shift in ordinal matching ability by the age of 7. The hypothesis generated from the modeling in Chapter III was that this is supported by a spontaneous deployment of a seriation strategy, starting with one or other end-point. The marked difference in the performance of 7-year-olds was not fully reflected however in their ability to verbalize the five or seven size relations or to correctly estimate the number of items in the set.

The results are consistent with Piaget’s tests of ordinal and cardinal correspondence but in the context of our program, they raise the question of why it was so difficult to train the 5-year-olds to make the correct match when they could be relatively easily trained to seriate five items. An obvious issue is whether the 5-year-olds were simply confused by the matching task. The seriation involves a cumulative procedure in which learning can be built up gradually, whereas the MTS task has no such structure. The data from studies such as those of Blevins-Knabe (1987a) and Siegel (1972) suggest that understanding ordinal sizes within small sets of two or three items might precede that of more extended understanding involving four or more items. The next experiment was devoted to exploring ordinal matching under more gradualist training.

Experiment 6

Rationale

The question being addressed here is what lies at the heart of the difference between monotonic sequential and ordinal size comprehension and, in particular, whether accidental features of our tasks were responsible for the difference in performance by 5-year-olds. Monotonicity, proceeding from biggest to smallest or vice-versa, is an assumed aspect of expertise in each, although it would be open to the individual as to which end to start the search for middle-sized for example. But what of the learner who has yet to achieve that solution? The monotonic size sequence can enlist this strategy by accumulating correct responses over time. Once the correct first position is learned, the participant can concentrate on the next correct and so on. There is no such cumulative cue in our ordinal matching task, where all items have to be identified in a random democratic fashion across trial blocks. However, should the difficulty on ordinal learning disappear under more gradual and cumulative training, then it would indicate that there is no in principle difference between learning to seriate and learning to ordinate but rather a task advantage applying to the former because of its very nature. In this study, 5-year-olds were trained on ordinal matching in a cumulative fashion starting with the two size relations bigger and smaller.

Method

Participants

Fourteen 5-year-old children (seven girls and seven boys) were selected from the same population as before. They were aged between 5;0 and 5;11 ($Md = 5;2$). All children had English as their first language.

Task and Stimuli

Using the same rectangular stimuli as in Experiment 4 (Figure 16), the task was presented across four phases. In the first phase, two adjacent sizes from an end-point were represented the target stimuli bigger and smaller, but using the same intervals as in Experiment 4. We shall call this P2 to avoid confusion between phase number and number of items presented. As before, a corresponding pair from the other size range was used as the sample display and, as before, the selection of the size range to be used as the sample display was counterbalanced across participants. The choice of end-point was also counterbalanced, such that for half the children the bigger item would be 2nd smallest from the set as a whole (the smaller being the smallest one); for the other half it would be the biggest (the smaller now being the 2nd biggest). In P3, one stimulus was added on the previous two, at one interval distance, creating a “middle-sized” selection rule; in P4, a fourth item was

TABLE 10

Percentage of successful participants and mean (SD) trials required for four the phases of Experiment 6

Phase	2	3	4	5	*5 Items (Expt. 4)*
%	100	100	71	50	*33*
TSCR	3.8 (5.3)	14.7 (10.5)	16.0 (9.6)	12.3 (8.8)	*20.5 (4.7)*

Note. TSCR = trials to start of criterion run.

added so the ordinal selection rules were now “biggest,” “smallest,” and “second (or next) biggest” and “second (or next) smallest.” Finally, in P5, all stimuli were used, thus converging on the MTS task used in Experiment 4A.

For each phase, the task was presented in trial blocks (of two, three, four, then five trials each) with the ordinal rules randomized within trial blocks and until a criterion of 4/5 completely trial blocks could be met within an upper limit of 40 trial blocks. On failure to meet criterion, the child exited the task and was not given the next phase. TSCR represents the number of trial blocks to the start of criterion run.

Results

Choice

All children succeeded in reaching criterion on P2 and P3; four failed on P4, and of the ten who went on to P5, seven succeeded. Table 10 shows the mean number of trials blocks given to all participants within a phase whether or not they succeeded, and mean TSCR for successful participants. Comparison data from Experiment 4 are shown.

Figure 19 shows the error distributions across all participants attempting each phase. P5 is compared with the error data from the four successful MTS participants in Experiment 4. The varying number of participants across set size makes it difficult to statistically analyze the effect of item position across each substage of the study. However, the distributions of error in the final stage are similar to those from Experiment 4 and both are different from an even distribution, showing most error on the middle three items.

Discussion

In this experiment, the seriation and ordinal task requirements were converged by introducing a cumulative structure in the ordinal task. This did not remove the relative difficulty attaching to the ordinal task, however. Although fewer children failed by the final phase than in the corresponding 5-item task in Experiment 4, only half the sample here succeeded within the training limits. The error profiles from that final phase were highly similar to those from Experiment 4 in showing most error on the middle items.

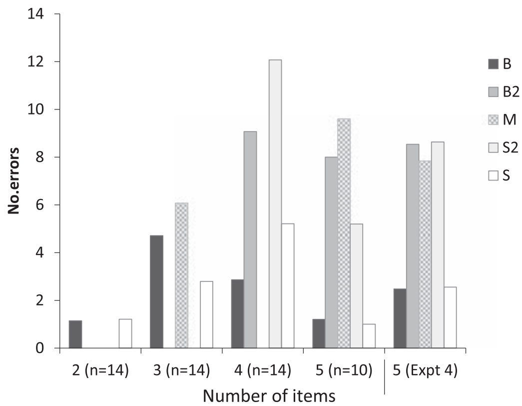

FIGURE 19.—Performance by 5-year-olds on the four phases of the cumulative MTS task (Experiment 6) depicting trial of last error across item position and compared with the performance from Experiment 4.

Taken together the MTS tasks show a level of difficulty that is, in terms of successful participants, greater than that shown by the 5-year-old participants in Experiment 1 with the same trial limits. Although there was no expectation that ordinal abilities would be anything other than restricted in this age group, it is important to be sure that extraneous aspects of the task are not responsible for their poor performance.

It is in the nature of testing for a mental code representing an ordinal size value that it is not simply based on the absolute properties of the stimuli. In the matching task, an ordinal value derived from the sample display has to be matched to one from a different range of sizes (see also Blevins-Knabe, 1987a). Although this preserves the essential element of disallowing an absolute size value from being used for correct matching, it could be argued to be quite confusing to have both ranges present in an overall display of 10 items, which are otherwise similar. (In Piaget's correspondence tasks the two sets were usually different objects such as dolls and sticks.) Although a non-relational match is unlikely for the positions, biggest and smallest, difficulty in identifying inner ordinal positions could throw children onto an absolute match attempt under these circumstances.

The next study in this chapter uses a different, single-set, measure of ordinal understanding based on a color conditional training paradigm that had been successfully used with monkeys (McGonigle & Chalmers, 2002).

Experiment 7

Rationale

In this experiment we wished to explore the development of ordinal size identification without, this time, using a correspondence or matching technique. Other investigators (Siegel, 1972) have isolated single sizes per participant. Here we wished to explore the way in which children could separately identify each and every one of the sizes used in the sequencing tasks described in Chapter II. The method used was based on training an association between five different colors and the five ordinal positions, biggest, second biggest, middle-sized, second smallest, and smallest. The ordinal sizes were not specific to one set of sizes to ensure (again) that the ordinal calculations were not based on absolute properties of the stimuli.

Method

Participants

Fifteen 5-year-old children (seven girls and eight boys) and twelve 7-year-olds (four girls and eight boys) were selected from the same population as before. The younger group was aged between 5;1 and 5;9 (*Md* 5;5); the older group between 7;1 and 7;11 (*Md* = 7;7).

Stimuli

The stimuli were the same rectangular shapes as those used in previous experiments displayed in bottom aligned single horizontal rows as in the seriation tasks. Unlike previous tasks, where a single color was allocated to each child, this time a different color was allocated to a different ordinal rule. The allocation of five colors to five rules across participants was performed by allocating each child to one of a set of 5×5 Latin square designs, in which each size rule was randomly allocated to one particular color (from a parent set of ten). The two size ranges were deployed randomly within colors across trials.

Design and Procedure

The experiment was conducted in two phases. In the first, the participants were given five trials per item before proceeding to the next ordinal position. The positions were presented in a random rotation with the constraint that each was presented once within a total of five trial blocks. At the start of each new trial block, the child was instructed to touch the stimulus that “went with” the color. Criterion was achieving 4/5 completely correct trials for each ordinal position (with two correct from each size range) within an upper limit of 16 trial blocks on any one position. Training was discontinued for this phase if performance on one or more ordinal positions failed to meet this criterion. Immediately following Phase 1, Phase 2 was presented with the instruction that the colors would now change each time and that the child should still try to remember which color went with each size. Here each ordinal position was presented once each in a random alternation within trial blocks of five. The learning criterion was a completely correct performance on 4/5 trial blocks with an upper limit of 50 trial blocks (50 trials per ordinal position). A criterion run had to include two trials from each size range. On both phases, feedback as to progress was shown as before with a man ascending one rung of a ladder after a correct trial block and picking an apple from a tree on meeting criterion. Owing to the ambiguity and difficulty in classifying previous linguistic data from 5-year-olds, 7-year-olds only were asked about the number, colors, and sizes of the stimuli at the end of the experiment.

Results (Phase 1)

Choice

Three 5-year-olds failed to meet the learning criterion on this phase (requiring at least 16 trial blocks on at least one ordinal position); the remaining 12 participants succeeded within a mean of 8 trial blocks ($SD = 3.1$). All 7-year-olds met criterion within a mean of 1.75 trial blocks ($SD = .62$). The number of trial blocks required by successful participants in each age group was significantly different on a Kruskal–Wallis one-way ANOVA, $H(1, N = 24) = 17.04$, $p < .001$, $d = 3.28$, a large effect.

The first trial out of a block of five was computed separately from the remaining trials in that block, on the grounds that the first trial also carried the highest memory load (i.e., remembering the color-size association). Figure 20 shows the error distributions from the first trial of each block versus the mean derived from the subsequent four trials. Five-year-olds have been divided into all ($n = 15$) and successful ($n = 12$). Only successful 5-year-olds were included in the comparison with 7-year-olds. All children were better on Trials 2 to 5, but 7-year-olds were virtually error-free on these trials. For Trial 1, a Scheirer–Ray–Hare two-way non-parametric ANOVA showed a large main effect for age, $F(1,106) = 146.72$, $p < .001$, $d = 5.13$. No interaction was found between

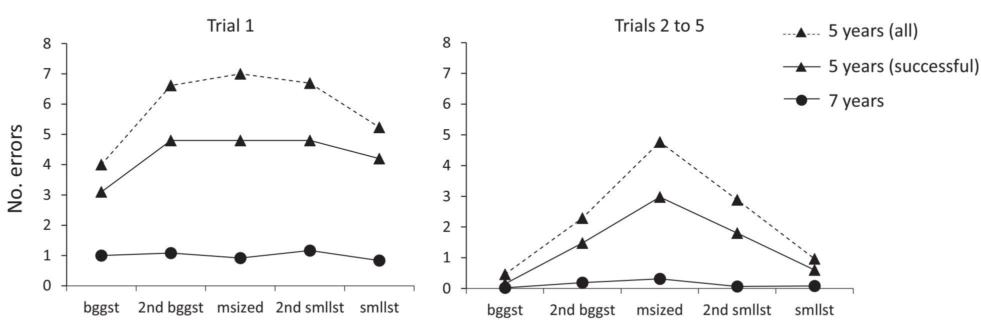

FIGURE 20.—Learning functions for Phase 1 of the color conditional task (Experiment 7) depicting the average trials to the start of criterion run across each item position for both age groups.

age and item position, $F(1,106) < 1$. Errors for Trials 2 to 5, likewise, showed a large main effect for age, $F(1,106) = 101.31$, $p <$,001, $d = 4.29$. No interaction was found between age and item position, $F(1,106) < 1$.

It appears from Figure 20 that the error by 5-year-olds is less sensitive to ordinal position on the first trial as compared with the subsequent four trials. This was confirmed using an unpaired Wald–Wolfowitz Standardized Runs Statistic. Trial 1 shows randomness, $r = -1.43$, $p = .08$, whereas Trials 2–5 show evidence of a pattern, $r = -4.85$, $p < .001$). Figure 20 shows that this pattern reflects highest uncertainty on the middle item.

Results (Phase 2)

Choice

Two of the 5-year-old children who failed Phase 1 refused to continue and of the remaining 13 who did, five failed to meet the learning criteria on this phase. The successful participants took an average of 21.2 trials to the start of criterion run ($SD = 14.8$). One 7-year-old did not complete the phase. Of the remaining 11 participants who did, all succeeded within a mean of 19.9 TSCR ($SD = 12.4$).

Figure 21 (left) shows the degree and distribution of error for all 5- and 7-year-olds ($N = 13$ and 11, respectively) and also for successful 5-year old participants ($N = 8$). A Scheirer–Ray–Hare two-way non-parametric ANOVA carried out on the data from successful participants showed no main effect for age, $F(1,91) < 1$, nor an interaction between age and position: $F(1,91) < 1$.

Reaction Times

Given the parity (for the first time) across ages in terms of errors committed by children who succeeded on Phase 2 of the color conditional task,

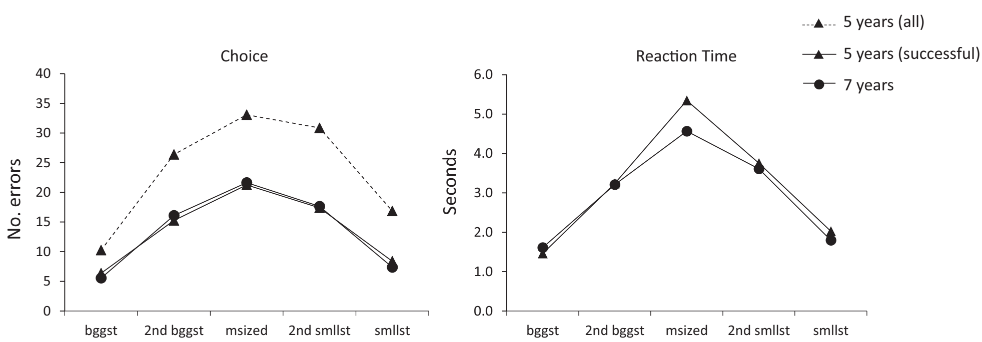

FIGURE 21.—Learning functions for Phase 2 of the color conditional task (Experiment 7) depicting (A) the average trials to the start of criterion run and (B) mean reaction time (RT) (sec) for correct trials across each item position and for both age groups.

RTs were examined for any deeper sources of difference. Mean RTs from successful participants were calculated for the last four correct trials and subject to outlier replacement (any score more than 2 *SD*s from the mean was replaced by the original mean); 6% of cells were replaced for each age group. The resulting functions are shown in Figure 21 (right) from which it can be seen that the functions were highly similar. A Scheirer–Ray–Hare two-way non-parametric ANOVA found no effect for age: $F(1,91) < 1$, nor any interaction between age and position: $F(1,91) < 1$.

Verbal Report

All but two of the twelve 7-year-old children were able to correctly verbalize color associations to the objects, although the size of the objects would not all have been unambiguously identified by a listener (e.g., "yellow: medium sized; blue third"). The number estimate was highly accurate with only one participant estimating it as "six." Eight children used numerical ordinal terms such as "second" or "third" (e.g.) "biggest" and five used those terms in such a way that would uniquely identify the full set. The score for unique specification by these participants was 4.3 out of a possible 5.0. There was no favored order with which they were mentioned; four described them monotonically; four by mentioning the end items first and the remaining four in random order. One child only used numbers in a monotonic description: "green: smallest; white: second smallest; brown—three; blue—four; pink— five."

MTS Posttest

With eight 5-year-old children succeeding by Phase 2 of this study, we were curious as to whether they would experience the same level of difficulty on the MTS as the participants in Experiment 4. They were presented with the task as described in Experiment 4 as soon as possible after completion of the Phase 2 Color Conditional. Task transfer levels were very high with a mean TSCR of 5.6 ($SD = 2.7$).

Discussion

Once again, ordinal identification by size was again shown to be extremely difficult for 5-year-olds even without a matching requirement. This was evinced even in Phase 1 where the color code for each ordinal selection rule remained constant for five trials. It was also evident in the fact that around only half the group succeeded on Phase 2. The error distributions showed highest difficulty on the middle-sized item for all participants. The age difference in training performance on Phase 1 was as evident as it was for seriation training (Experiment 1) and Match-to-Sample training (Experiment 4). Somewhat surprisingly, however, the two groups converged to some extent during Phase 2. Although the 7-year-olds succeeded as a group, they showed a level (and distribution) of difficulty similar to the successful 5-year-olds. Unlike the seriation performance in Experiment 1 (Chapter II), furthermore, their RTs were not significantly different from the successful younger children on Phase 2. This, together with their error on this phase indicates a possible constraint on how available the codes for ordinal position (whatever they may be) are for the older children. One possible reason for this is, because of their short training on Phase 1, the memory requirement during Phase 2 was relatively enhanced for the older age group, simply due to reduced initial exposure to the color associations.

The color conditional task therefore now raises the question of why 7-year-olds cease to show the qualitative advantage over 5-year-olds they have shown in every other size relational task we have described thus far, including Phase 1 of this study. The possibility just raised is that the second phase of this study imposes a relatively heavier WM on 7-year-olds due to reduced exposure to the color associations in Phase 1. Alternatively, the ordinal capability shown by the older group may be such that it does not easily permit such associations to be made under the conditions employed in Phase 2. The verbal protocols from the 7-year-old group indicated that by the end of training, they had reasonable access to ordinal descriptions. Although around half the group fell short of uniquely specifying the set, this could simply be due to difficulty in verbalizing information that is stored primarily in some nonverbal way. The question is how is that information stored?

We have utilized the concept of LTM in the computational modeling of sequence learning and shown how it can start to store unitized slots allowing for spontaneous seriation. So far this has fitted the empirical picture on ordinal identification. However, we now have evidence that 7-year-olds may have a constraint on how they may store ordinal size relations in LTM, despite their evident ability to compute them during a sequencing task. This implication does not repudiate the model in itself, but it indicates that our explanatory model may only be applicable to situations where spontaneous seriation (the necessary precursor to slot formation) is enabled by the task. Before pursuing the implications of this possibility, we take stock of what we can conclude from our experiments on sequential and ordinal size understanding thus far.

Summary of Age Differences Across Studies

So far, we have examined the empirical basis for the claim that a shift occurs in sequential and ordinal size understanding across the ages of 5–7 years. We have strongly endorsed this claim, but with two new caveats arising from comparing and contrasting across related training tasks using the same stimuli. The first is that ordinal size identification is considerably harder to train in 5-year-olds than seriation (and marginally harder even for 7-year-old spontaneous seriators). These results oblige us to treat ordinal size understanding as secondary and arguably emergent from the serial processing of size relations rather than as a codependent ability. The second is that expertise in identifying ordinal sizes values in the context of sequencing does not extend to being able to encode them as a set of independent codes in LTM that can be retrieved at random and applied to a given item outside the act of sequencing.

Although each study is based on a small sample of the relevant ages, each one used a depth of training and measurement that goes beyond the one-off success criteria of the classic Piagetian methodology. We also checked the poor performance registered by the younger children across a variety of conditions, but the level of training difficulty remained fairly consistent. Initial difficulties that might have arisen from aspects of the matching paradigm were mitigated by (lengthy) training on a color conditional task, as also shown by the high level of transfer but this task itself yielded high levels of difficulty before successful performance was achieved and even then only by a subset of younger participants.

The tasks were more modest in terms of set-size/processing load than the traditional one, in that we used 5- and 7-item sets rather than the 8- to 10-item sets more commonly employed. Does this raise the possibility that 7-year-olds may not have shown the spontaneity they evinced in seriating and matching correctly had the sets been larger? Apart from the fact that no suggestion arises in the literature that set size would affect older children, the modeling carried out to explain seriation spontaneity in Chapter III predicts a natural extension to larger sets. The main point is that difficulty by 5-year-olds has been documented with very small sets well within their counting range. Seriation ability does appear to change dramatically between these ages but ordinal identification even more so.

New Questions Arising

Assuming that we have stumbled upon a problem with the accessibility of ordinal codes in 7-year-olds, we now have to consider what that might be, and in what sense it is different from the way that ordinal sizes were codified to solve a matching task. We have already speculated that ordinal position may be quickly and expertly calculated using a spontaneous monotonic search strategy from either end-point of the series together with a possible stop-rule or count applied to the search. So far, this is consistent with the results from 7-year-olds in Phase 1. Although the first trial of each trial block puts some strain on their memory for the correct color-size association, once retrieved the ordinal position in WM can now be used in order to solve Trials 2–5 without having to retrieve anything else from a WM store other than some sort of count or tally based on performing a principled search. In Phase 2, however, a much greater strain is put on remembering all the ordinal values and accessing them at random from LTM. Here it seems that an arbitrary symbol like a color is not easily used as a place marker in memory for ordinal position for reasons that are not clear. Seven-year-olds took fewer than four trials to learn an arbitrary color string in Experiment 1 and so learning a list assigning each color (a symbol) to each successive position would seem a task well within their WM capability.

One resolution of the paradox would arise if the codes representing ordinal values are not stand-alone entities (at this age) in the way that color names are, but are tied to online visuomotor search-and-stop procedures that require cueing from the visual field. Any codification remembered as a tally or a count would be stored temporarily until a response was made. In those terms, the MTS tasks provides a visible sample array, cueing a count to be made to the flashing object from the nearest end-point (as in MTS). The count is remembered until a similar visual search procedure is carried out on the target array. On this view, the match is **not** made on the basis of an abstracted or symbolic rule such as "find second biggest" but simply on a count-and-stop procedure cued directly from the samples array. Similarly, Phase 1 of the color conditional task provides a type of "pointing" to the correct stimulus (and count) on the first trial and no other tally or count has to be remembered until the next trial block. Phase 2, however, does not have the feedback from the immediately previous trial to provide the count value. Here the only rapid solution would be to rely exclusively on separate "codes" for ordinal size that exist in LTM in order to construct the full list of mappings. That they can do so by the end of training shows that this is at least achievable by some means for this size of set for 7-year-olds and even around half of the 5-year-olds. However, the convergence across age in terms of relatively lengthy training and similarity in RT profiles suggests that there is at the very least a lack of availability of stand-alone ordinal codes for all these children that requires further scrutiny. The divergence across groups for Phase 1, by contrast, suggests that principle search can indeed be augmented by a search-and-stop count or tally enabling ordinal identification by 7-year-olds at least under some conditions.

Before pursuing these possibilities further, we must acknowledge that memory for color associations is a possible confound in this task. We now turn to a different and more direct means of evaluating memory for ordinal sizes. The final study in our program describes a single set sequencing task—not monotonic seriation as described in Chapter I, but non-monotonic ordering of ordinal codes.

Experiment 8

Rationale

This study tested the ability of 5- and 7-year-olds to store ordinal codes in a fixed random order in LTM. The task was a sequencing one, run in exactly the same way as the serial monotonic tasks described in Chapter I. Here, however, the sequence could not be computed on line by seeking the next biggest/smallest from one end. It could be learned only by forming a list of ordinal sizes in the form of five distinct codes, whether as numbers 1–5 or as labels, "biggest," "second biggest," and so forth.

Method

Participants

Drawn from the same population as the other experiments, there were twelve 5-year-olds (seven girls and five boys) aged between 5;0 and 5;10 (*Md* = 5.02). The older group (four girls and eight boys) were aged between 7;4 and 7;11 (*Md* = 7;8). All children had English as their first language.

Stimuli and Task

The stimuli and screen presentation were identical to that used for the 5-item monotonic seriation condition described in Chapter II (Experiment 1). Two size ranges, as used in all the above tasks, were presented across trials in random alternation. The single difference from Experiment 1 was that the stimuli had to be touched in a fixed random sequence. There were several ways of constructing such sequences and allocating them to participants. To allow at least some comparison to be made within participants in terms of "level" of non-monotonicity, two sequences were selected; one closer than the other to the monotonic version. Neither sequence started with an end-point. The "easier" sequence started with a "next to" end-point, had a mean of 1.75 interval distances between the items and only one change of direction. This sequence was second biggest, middle-sized, smallest, second smallest, and biggest, hereafter called 4 3 1 2 5 where 1 denotes smallest, 5 biggest. The harder sequence—3 2 5 1 4—started with middle-sized, had a mean interval distance of 2.75 and three changes of direction. Half the participants were allocated to one; half to the other. Whatever the outcome of training on the first sequence, the other one was presented as a posttest so that all participants attempted both sequences.

Procedure

The training procedure was the same as for the seriation task described in Chapter I. In summary, this involved a sequence pretraining episode in which children learned to touch two shapes in a particular order by listening to the bleep/buzz feedback. They also learned that the intertrial image of a man ascending a ladder to reach an apple indicated success and progress to criterion. Training on the 5-item size sequence then followed, and children were told that they now had to learn to touch the sizes in a particular order. As it became quickly evident that this was far harder for all participants than seriation training, the upper limit of training was set at 100 trials as long as the participants appeared to be happy to continue trying. Training thus occurred over consecutive days for some participants. After completion on each sequence participants were asked in what order they were to touch the squares and how many there were. They were not corrected or pressed for further information.

Results

Choice

Four out of the twelve 5-year-olds failed to meet criterion within 100 trials on the harder sequence (one refused to continue after 70 trials). One failed and one refused to continue after 50 trials on the easier sequence. The overall mean TSCR (*SD*) for successful participants was 34.3 (19.9) for the easier sequence ($N = 10$) and 38.6 (19.5) for the harder sequence ($N = 8$). All 7-year-olds succeeded with a mean (*SD*) of 14.8 (10.7) for the easier sequence and 20.9 (15.9) for the harder. The age difference for successful participants (when combined across sequence type) was significant on a Kruskal–Wallis one-way ANOVA, $H(1, N = 22) = 6.28$, $p = .01$. This was a large effect, $d = 1.12$. There were no significant effects of sequence for either age group or when groups were combined (averaged). Paired-sample *t* tests for the successful 5-year-olds showed no effect of task order, $t(14) = .25$, $p = .82$, but a large effect for 7-year-olds: $t(22) = 4.9$, $p < .001$, $d = 1.41$.

Similar results were obtained when analyzed within each condition. A Scheirer–Ray–Hare two-way non-parametric ANOVA with a large main effect was found for age, $F(1,96) = 41.95$, $p < .001$, $d = 2.91$, for the easier condition, but no interaction between age and item position, $F(1,96) < 1$. For the harder condition a large main effect for age was found: $F(1,96) = 19.23$, $p < .001$, $d = 1.97$ but no interaction between age and item position: $F(1,96) < 1$. The error distributions are shown in Figure 22.

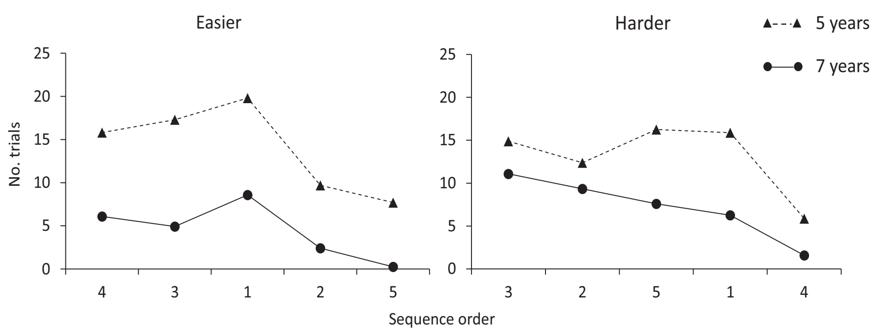

Figure 22.—Learning functions for 5-item non-monotonic sequencing (Experiment 8) depicting the average trials to the start of criterion run across each item position and across two series; the "easier" being closer to a monotonic one than the "harder."

TABLE 11

MEAN SCORE FROM VERBAL REPORT OF 5-ITEM NON-MONOTONIC SEQUENCE (EXPERIMENT 8)

	5 Years				7 Years			
Task	a	b	c	d	a	b	c	d
Easier	0.08	1.08	5.18	0.25	0.67	3.75	5.25	0.75
Harder	0.08	1.54	5.45	0.42	0.50	4.08	5.08	0.67

Note. a = order correct (max = 1.0); b = unique identifiers; c = number estimate; d = correct description of 1st item (max = 1.0)

Verbal Report

The analysis of verbal descriptions of the stimuli was based on the reports of all children irrespective of success on the task and focused on (a) whether or not a listener could work out the sequence that had just been presented to the child (yes = 1, no = 0); and (b) the mean score based on the number of items that were uniquely identified during the description of the order. For example, whereas “little bit bigger; little bit smaller; big then tiny then small big one” would be “no” in (a), it would gain a score of 3 in (b) as “tiny,” “big,” and “small big one” were reasonably unambiguous descriptions of the smallest, largest and second largest in their correct positions in the set. The value in (c) was the mean number estimated by children to be in the set. These estimates ranged from “four” to “six” for both groups and two 5-year-olds said they didn’t know. As the sequence started on an inner position, the term used to describe it was scored in (d) in terms of unique identification, for example, “the one next to biggest” (easier sequence) or “the medium one” (harder sequence). This was a given a score of correct (1) or incorrect (0). The data from both age groups are shown in Table 11.

Given the particular demand to remember each position separately from every other in this task—which we thought might induce a numeric code—we scored the verbal descriptions under (b) in Table 11 under specific categories. These were: incorrect; relational (e.g., “middle-sized” or “one below biggest”); and numerical relational (e.g., “third biggest” or “second smallest”). These data are shown in Figure 23 from which it can be seen that there was large shift toward the numeric relational codes for 7-year-olds though it still represented only around half of the total descriptions. No child used a single real number.

Discussion

Although results from the non-monotonic sequence showed a difference between 5- and 7-year-olds, this could not be described as indicating a discontinuous shift in ability in the way that age differences in size sequencing were manifested. Although they demonstrated significantly faster learning,

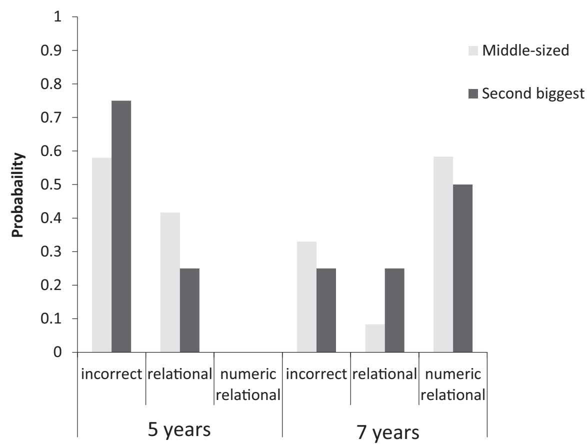

FIGURE 23.—Frequency (by percentage) of linguistic label used to describe the first item in the non-monotonic 5-item sequence (2nd biggest and middle-sized) for both age groups.

the older children were far from spontaneous in correctly adopting the arbitrarily ordered sequence of sizes. Of note was the fact that they took a mean of 11 trials to even learn the first item in the sequence (the middle-sized one). They did, however, show a strong effect of practice suggesting that they were initially affected by the unfamiliarity of the task.

Experiment 9

Experiments 7 and 8 have introduced an issue that we have not encountered either in reports from previous research on ordinal size understanding, or from our previous experiments. This is the issue of having to recall ordinal sizes from memory as distinct from computing them in situ. The issue became evident when we tried to ensure that younger children were not being overwhelmed by large stimulus arrays and switched away from a matching paradigm to a color conditional and then to a non-monotonic sequencing task. In so doing, the paradigm shift has also revealed a possible constraint on how older children succeed in their ordinal size abilities. To be clearer about the degree and severity of this constraint, a new group of 7-year-old participants was given 7-item non-monotonic task using the same rod-like stimuli as used in 7-item monotonic sequencing task (Experiment 2) described in Chapter II. On the premise that the longer sequence might engender much greater difficulty, a subgroup of 7-year-olds ($N = 8$) were given the 5-item non-monotonic sequence first (using the rod-like stimuli) and then transferred after learning to a longer sequence of seven items. This is called Experiment 9A. The difficulty already encountered by the 5-year-olds on the 5-item task disallowed the use of younger participants on longer sequences.

Method (9A)

Participants

Participants were eight 7-year-olds (five boys and three girls) ranging in age from 7;5 to 8;1. with a median age of 7;6.

Stimuli and Task

To allow seven items to be placed on the screen, these were the rod-like stimuli used in Experiment 2 (Chapter II). The sequences were once again 43125 ("easier") and 32514 ("harder"). However, after learning the first sequence the same participants were given two new stimuli of the same interval distance added at the "large" end. Thus the "small" range stimuli were (in height) 10–30 mm and the large range were 30–50 mm and then 30–60 mm. The transfer sequence for the easier task was now 6512743, preserving the start item as second biggest and the interval distances and changes in direction for the first five items in the set with a new mean interval distance of 2.5 and two changes of direction. Similarly, the new "harder" sequence (4371625) also preserved the start item (middle-sized) and the adjacency relationships for the first five items with a new mean interval distance across all seven items as 3.83 with a total of five changes in direction. Based on the 7-year-old performance in Experiment 8, 50 trials were allowed for the initial 5-item sequence and a further 40 for the transfer phase.

Following completion of the transfer phase, children were asked how the sequence was different from the first (5-item one) and to specify any difference in the number of items.

Results (9A)

Choice

One participant failed to meet the learning criterion for the 5-stimulus "harder" sequence but passed on the transfer phase, whereas two participants failed on the transfer on this sequence. The mean (and *SD*) TSCR for the 5-item easier task was 16.8 (2.4) for successful participants and for the harder task was 17.4 (6.8). The corresponding TSCR data for the 7-item transfer phase are and 20.75 (5.7) for the easier task and 28 (8.5) for the harder.

Verbal Report

All children noted that there were two more items, but the question provoked answers that were difficult to interpret as some appeared to be describing the new order; others the new sizes. However, six children mentioned a "fourth" biggest or smallest, and one mentioned a "fifth biggest" when describing the previous set, but the order described for the new set only went up to this "fifth biggest" and not beyond.

Summary (9A)

The pilot study indicated that 7-item non-monotonic sequencing, although difficult, is within the capability of 7-year-olds at least under transfer conditions. To obtain a fuller comparison with 7-item MTS and monotonic seriation, a new group of 7-year-old participants were given the 7-item task from scratch.

Method (9B)

Participants

Participants were twelve children (four girls and eight boys) aged between 7;5 and 8;1 (*Md* = 7;7).

Stimuli and Procedure

The two 7-item sequences as described for Experiment 9A were randomly but equally allocated to the participants. On completion (successfully or not) of the first sequence, the second one was presented. Training was conducted across consecutive sessions if necessary. Participants were allowed to continue for an upper limit of 100 trials if they were willing to keep trying. Children were asked about the order and number of sizes after successful completion of each sequence provided there was time left after the successful session.

Results

Choice

One child failed to meet criterion after 100 trials on the harder sequence; another refused to continue after 43 trials. Three children gave up on the easier sequence after 50, 52, and 83 trials, respectively. The mean (*SD*) TSCR for the remaining children were 28.0 (11.6) and 26.5 (14.5) for the easier and harder sequences. Paired samples t tests found task order to be significant, $t(22) = 2.77$, $p = .02$, $d = 0.79$, an intermediate effect. A Scheirer–Ray–Hare two-way non-parametric ANOVA found a large effect of task in favor of the easier over the harder, $F(1,136) = 19.84$, $p < .001$, $d = 2.16$, but no interaction between task and item position, $F(1,136) < 1$.

The learning functions for all successful participants are shown in Figure 24.

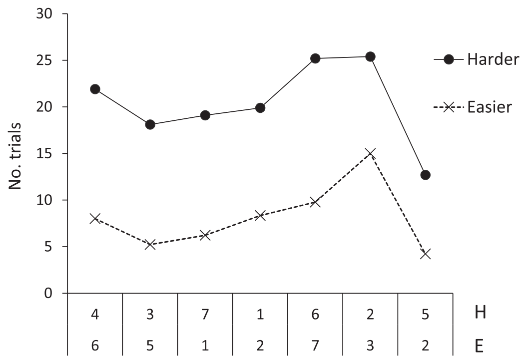

FIGURE 24.—Learning functions for 7-item non-monotonic sequencing (Experiment 9) depicting the average trials to the start of criterion run across each item position and across two series; the "easier" being closer to a monotonic one than the "harder."

Verbal Report

Owing to the unwillingness of one participant to describe the sequence, plus the four failed attempts, there were 17 protocols in total (eight for the easier sequence, nine for the harder). The scores are shown in Table 12. Only two descriptions out of the total would have permitted a listener to identify the order, and the levels of specificity were only at maximum for four participants (on the easier sequence). The incorrect number estimates varied from "six" to "ten." All children except one used numerical relational terms but one just said there "big ones and little one—four big ones." All other children correctly identified the second biggest (highest, tallest) start end-point and described it in those terms, though one said it was "one just shorter than the biggest." Seven of the eight correctly identified the start end-point as medium in the hard series, although only one named it as such. The remainder called it "fourth smallest" ($N = 3$) or "fourth biggest/tallest" ($N = 4$). One participant was unsure and said "fourth tallest—smallest I mean." Two participants described the last two items in the series as outliers, that is (for middle-sized and second smallest), "there were two left—biggest then smallest" and (for 2nd biggest, 2nd smallest, and 3rd biggest) "there were three left—biggest out of the three, second smallest, then the one that was left."

TABLE 12

MEAN SCORE FOR 7-YEAR-OLDS DURING VERBAL REPORT OF 7-ITEM NON-MONOTONIC SEQUENCE (EXPERIMENT 9B)

Task	a	b	c
Easier ($N = 8$)	0.25	6.00	7.12
Harder ($N = 9$)	0.00	3.56	8.56

Note. a = order correct (max = 1.0); b = unique identifiers; c = number estimate

Discussion

The expanded (7-item) non-monotonic task proved very hard for 7-year-old children when given from scratch. Our prediction that one version might be easier than another if it adhered more closely to a monotonic series was given some support, but this is a factor that would require a very much larger study in order to de-confound ordinal position with the intervals between items and the changes in direction. In the current study, moreover, there appeared to be an effect of practice, as if a heuristic for solving the task was becoming available at least to some participants. Despite the many questions that this task now raises, the results show unequivocally that, although 7-year-olds may be able to compute ordinal sizes on-line, encoding these as a fixed list in LTM is another matter. As the adult investigators and their colleagues were genuinely surprised at how hard children found this task, we decided to carry out a brief but systematic check as to how intrinsically hard the non-monotonic task might be even for adults. There were no precedents at all in the literature for a task of exactly this sort and so we carried out a short study with adult students and colleagues from our lab.

Experiment 10

Rationale

This was short evaluation of the intrinsic difficulty of the nonmonotonic task using adult participants.

Method

Participants and Design

These were eight students and staff (three female and five male) at our lab ranging in age from 21 to 40 years in age. Although educated to University level, none were involved in the ongoing research and only one was a psychologist. Bearing in mind that their abilities on computerized task might be higher than the in the average adult population, we attempted to gain a comparison regarding the relative difficulty of the non-monotonic task Accordingly, it was presented within a set of all the 7-item tasks as given to children. For four participants the order was Match-to-Sample, Non-monotonic sequencing, Monotonic sequencing, and Color sequencing. For the remaining four the order was Color, Monotonic, Non-monotonic, and MTS. Four were given the “harder;” four the “easier” Nonmonotonic task and all participants were asked about the order and number of items after this task. The procedures for all tasks were identical to those described for children.

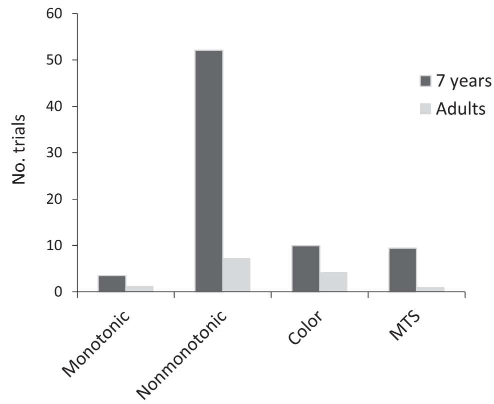

FIGURE 25.—Summary of performance on sequencing and ordinal tasks by 7-year-old children and adults showing mean trials to the start of criterion run.

Results (Adults and 7-Year-Olds Compared)

Choice

All adults solved all tasks without difficulty and, specifically, for the nonmonotonic task, they started criterion run in an average of 7.3 trials ($SD = 3.4$). The nonmonotonic task was therefore nearly eight times easier for adults than for successful 7-year-olds in terms of learning criteria. To place these findings in context and for illustrative purposes only, Figure 25 shows the performance in terms of mean TSCR across all four tasks given to adults compared with the comparable data from 7- year-olds reported earlier. (Error bars would make this graph highly unclear and are not depicted.)

Reaction Times

As the error level was too low in adults to see if the level of nonmonotonicity affected them (there was mean difference of 0.3 TSCR), we compared their RTs once the sequence had been acquired to see if this continued to affect performance even in skilled learners. RTs from correct sequences during the criterion run (8 per ordinal position for 7-year-olds and 4 for adults) were collated and subject to replacement of outlier cells, where any single RT that was more than 2 *SD*s from the mean for that column (ordinal position) was replaced by the original mean. Around 20% were replaced distributed across participants. Similar outlier replacement in adults resulted in 26% replacement. The data from the first and second presentation was combined for 7-year-olds. A comparison of functions for children and adults is given in Figure 26.

Large main effects were found for age for both the easy and hard conditions using a Scheirer–Ray–Hare two-way non-parametric ANOVA, $F(1{,}710) = 24.92$, $p < .001$, $d = .99$ for the easy condition and $F(1{,}780) = 28.83$, $p < .001$, $d = 1.05$ for the harder. Both groups showed large effects of task in

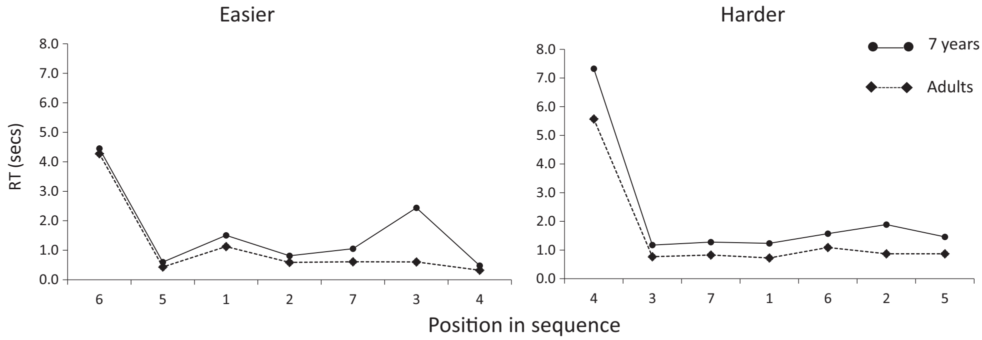

FIGURE 26.—Reaction time (RT) functions for 7-year-olds and adults on an easier and harder 7-item nonmonotonic sequencing task depicting the mean RT (sec) for the last (8) successful trials across item position.

favor of the easier task, $F(1,1053) = 25.26$, $p < .001$ for 7-year-olds, and $F(1,437) = 9.7$, $p = .002$ for adults. The functions are shown in Figure 26. However, as Figure 26 shows, much of this effect can be attributed to the difference at the first position. This was confirmed by t tests comparing the RT values at each position for both easy and hard conditions for each age group, which showed that for the 7-year-olds, the harder condition was approximately 3 sec slower on average, $t\ (150) = -5.56$, $p < .001$, $r = 0.90$, a large effect, and for the adults, the harder condition was approximately 2 sec slower on average, $t(62) = -4.90$, $p < .001$, $r = 1.23$, a large effect. The t tests for the remaining positions for 7-year-olds and adults showed a mixture of significant and insignificant differences, with RT differences varying between 0 and 1 second.

Verbal Report (Adults)

Five out of the eight participants described the sequence accurately, all gave the correct number estimate and the mean score for linguistic specification out of a maximum of 7 was 6.0 Although they were all scientifically trained, only two participants converted the sequence into numbers; the remainder used the same numerical relational terms as the 7-year-olds, including the middle-sized one that was described as such by only one participant (the rest used "fourth largest" or "fourth smallest"). However, one participant said he counted to find (number) four each time but did not report the sequence in terms of numbers. Two participants (like two of the 7-year-olds) described the last two items in the sequence separately from the rest: "of the two remaining it was the larger than smaller" and "out of the two left—big then little—they were both kind of middly."

Discussion

It was not our purpose to carry out a full and formal comparison between 7-year-olds and adults on the nonmonotonic task but rather to ensure children had not been given a task that is in principle very difficult to solve. The data from the adults disconfirmed this possibility in showing rapid learning, but neither the verbal reports nor the RTs suggested that there was any radical difference across age groups in terms of how the sequence was being executed once learned. The adults were as sensitive to task difficulty as were the children and as shown by their RTs. In other words, both groups appeared to execute the sequences in a similar way—a finding we shall reconsider in Chapter V. In the meantime, we must conclude that, whatever changes may occur later, for 7-year-olds, ordinal size calculation abilities do not easily extend to the codification required to support the learning of a non-monotonic size sequence.

Empirical Assay: Summary and Discussion

In this chapter, we have explored the development of ordinal size understanding across the age span during which a dramatic change in seriation abilities takes place. We have tested ordinal size understanding using a correspondence method, a color association method and finally a method involving having to remember how ordinal sizes were sequenced within an arbitrary list. All these tasks showed high levels of difficulty for 5-year-old children for small 5-item sets and even for sets of four items (Experiment 6). All tasks showed greater difficulty than reported in Chapter II for size sequencing. Although aspects of matching and color conditional training could account for some of the discrepancy, it is nevertheless the case that 5-year-olds did not show the ease of ordinal identification shown by 7-year-olds under the same conditions—at least in the matching task and Phase 1 of the color association task.

So far, we can say that the improvement in ordinal identification that we found to occur by 7 years of age is in agreement with Piaget's findings of robust ordinal size correspondence for children of that age. However, in pursuing possible task artifacts we introduced a memory requirement that generated the unexpected finding that 7-year-olds were now no longer proficient. This has an immediate implication from the conclusions that Piaget drew from his correspondence and insertion tasks. For Piaget, the connected relationships that appear at around 7 years of age are between spontaneous seriation, effective counting, and ordinal and cardinal understanding. On his account, identification of the correct matching item would equate to assessing each item in the sequence as a potential count to item N, as well as encoding the count as a cardinal and ordinal value. In a situation such as a matching task, these are hard to disambiguate. When we removed the matching factor, however, it became apparent that 7-year-olds were not necessarily converting the ordinal code to a cardinal value that would allow an easy mapping to a color (such as 2 = green) or a disordered sequence to be remembered (43125). This puts a spotlight on how the 7-year-olds do **not** solve ordinal identification tasks and so we consider this now in relation to the results from Experiments 7–10.

LTM and the Encoding of Ordinal Values

In the first phase of the Color Conditional task (Experiment 7), older children rapidly identified the correct position after the first trial on which they learned (or were reminded of) which color went with which stimulus. In Phase 2, when the trials were randomly alternated, their performance was no longer different from successful 5-year-olds, suggesting that an important support for their ordinal abilities was now missing.

One such support is the immediately prior prompting of the correct position. In all of Piaget's tasks, in our own Match to Sample task, and in Trials 2–5 of our color conditional task, the correct position (count) within the set has been singled out and identified first. The participant can carry that count forward when attempting the match or, in the case of the color conditional task, the subsequent trials where the color cue remains the same. In Phase 2 of the color conditional task, there is no such cue, and on each and every trial, the correct count for the presented color has to be retrieved from memory from a set of alternatives. In reviewing the relevant research, we can see that this last feature is unusual for tests of ordinal size comprehension. In size discrimination learning, different size rules are not randomly interleaved but kept constant until criterion is met. In Piaget's ordinal correspondence tasks, one particular item is singled out for matching. This item furthermore can be found by seriating the other items in the array. The seriation modeling described in Chapter III is in fact quite consistent with this conclusion. The principled search that was envisaged to be enabled by spontaneous seriation task could presumably be accompanied by some sort of tally whose size value only needs to stay in WM long enough to find the corresponding tally in the other array. This hypothesis is operationalized in the model at the end of this chapter.

However, as far as the color conditional task is concerned, having to remember a set of arbitrary color associations in addition to the ordinal codes was a possible confound. We tested memory for ordinal codes more directly without this confound in our nonmonotonic sequencing tasks (Experiments 8 and 9), which had no such requirement. Here we encountered even greater difficulty, not only by 5-year-olds but also by 7-year-olds, especially for 7-item sets. We now had a strong indication not only of what expert seriation implies for 7-year-olds but also what is **not** implied by it. These children did not appear to have any LTM or permanent memory for ordinal codes that can exist independently of the on-line procedures for computing these values. The results from adults showed that it is possible to learn this task quite rapidly in principle (though there was little support from their data that the ordinal positions were automatically codified as numbers). There are many reasons why adults found this easy (and we return to this issue in Chapter V).

We now suppose that the only option available to the 7-year-old when trying to learn a non-monotonic sequence as fixed list, is to use the same (monotonic) search strategy as if they were seriating. If so, every ordinal position would become associated with a specific search procedure based on a monotonic core where each series is treated as a greater or less departure from monotonicity. Short-cuts for such a procedure could involve mini-sequences such that, for example, the sequence 43125 would be coded as a monotonic downward sequence from biggest (but where the first item is not selected), then a switch to an upwards mini sequence from smallest, finally followed by choosing biggest. Such a strategy which would explain the effect of relative disorder (which manipulated adjacency and number of switches) on the performance in the nonmonotonic tasks. It would also explain why even the first item in the sequence was so inadequately described. The expertise gained across tasks would be a matter of becoming readier to apply such an ad-hoc strategy on to a new sequence. Having to remember a disordered sequence thus seems to lay bare the fact that serial monotonic procedures for ordinal calculation by children of 7 years old are highly expert but are also the only means by which they can encode and remember the ordinal values.

Taking our studies on ordinal understanding together we can now reinforce the picture with which we concluded Chapter II, in confirming the phenomenon of a shift in ordinal size understanding between 5 and 7 years of age. As for the doubt raised at the end of that chapter regarding Piaget's account of the shift, we now submit that it is not easily explicable as the deployment of a reversible atemporal "operation." This is for two reasons. First, we were able to explain principle seriation and ordinal identification without recourse to this concept, and using, by contrast, a concept of serial unidirectionality. Second, if the 7-year-olds had access to the ordinal information in logical and atemporal terms, the non-monotonic task should have posed no real difficulty. In fact, it is a moot point as to whether such a logic would suggest any less privileged status to a non-monotonic as opposed to a monotonic order—apart from perhaps greater familiarity with the latter. Certainly, one nonmonotonic order should be equivalent to any other.

As for what **does** seem to be controlling choice in the older children, seriating the items prior to choice appears to be the means by which a specific ordinal code is generated. Both WM and LTM clearly have a role in this. The formal modeling that follows makes explicit how these processes could work to explain our results.

A Computational Model of Ordinal Competence

In Chapter III we presented three models which illustrated our general hypothesis as to how developmental progression between trial and error and principled size sequencing happens. We argued that a trial-and-error learning process creates a probabilistic serial order ranking within LTM (heuristic search model), which when combined with an increase in WM power (transitional model) allows a discontinuous change to spontaneous size seriation powered by a "select smallest difference" rule (principled search model). Specifically, we argued that the combination of *Beta* probability distribution stability and link stability jointly create a slot in LTM representing a separate unit for every item in a series. To this general hypothesis we now add the proposal that it is the use of these slots to store the cumulative tallying of the links that allows the agent to solve ordinal matching problems.

This extension to ordinal competence needs a new model, but crucially using the same representations that were formed during the transitional model presented in Chapter III (*Beta* distributions and the links between them). In Figure 27, we propose an architecture for an *ordinal search model*, in which an agent is presented with a matching to sample task aligning to the MTS task presentation in Experiments 4 and 5 in this chapter.

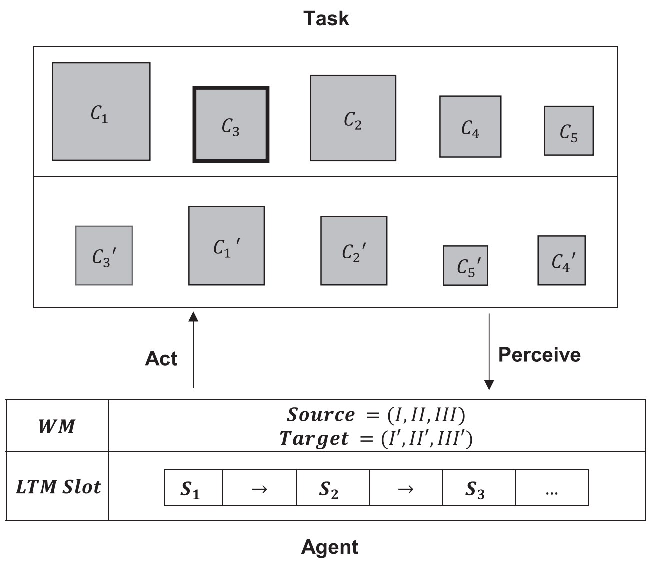

FIGURE 27.—Task and agent architecture overview in the ordinal model simulation. A task change from one that demands serial to one that demands ordinal competence forces the agent to leverage the slots (*LTM Slot*) previously established in long-term memory (LTM) via the transitional model. These slots are used by working memory (WM) to facilitate a count and tally of the number of times a smallest size difference encounter has occurred. In this scenario, stimulus C_3' within the bottom array contains the target stimulus. The agent must recognize the target stimulus as analogous to the source stimulus C_3 (flashing on the screen, indicated by the bold border) within the top array. The principled search model routines are (re)used to select the source stimulus, with the addition of a cumulative tally being made until the source is reached. When the source stimulus is reached within the top array, the cumulative tally is stored, indicating the stopping condition. The agent then interrogates the bottom array, again cumulatively tallying up the smallest size difference stimulus encounters, until the target stimulus, C_3', is reached.

In Figure 27, source array stimulus C_3 flashes, indicating to the agent to scan to it from the end (big or small) closest to it, tallying the stimuli encountered on the way. On reaching C_3, the agent scans the target array from the end selected within the source array scanning, and again tallies up the stimuli, the stopping condition being the equivalent number of tallies that were recorded for C_3, the target being C_3'.

The agent searches of source and target stimulus arrays using a "select smallest difference" rule, which was used within the principled search model in Chapter III. However, in this task scenario, which involves stopping during a search of a source array and finding a corresponding stimulus in a different array, merely seriating spontaneously is not enough. Such an ordinal identification task requires deeper information processing in terms of searching and recording a count of stimuli. The slot placeholders in LTM are updated from WM to store this count. The ordinal requirements of the MTS task provoke a competence that was dormant but not required in a spontaneous seriation scenario. Specifically, slots formed on the agent having reached threshold levels of rank order precision and set element relational invariance (each ordinal position now being seen as a discrete unit) can now be leveraged as holders in LTM to store the "stop" position until the equivalent position is found in the target array.

The tallying mechanism can be understood as follows. On each trial, we begin with a pair of empty lists in WM, one of which represents the source stimulus array, *Source* = (?), and one the target stimulus array, *Target* = (?). On each trial, a flashing stimulus in the source array indicates the ordinal position that must be selected in the analogous target stimulus array.

When an agent scans the source array from one of the end-points, and encounters a stimulus that is not in the required ordinal position, it marks this encounter in the source stimulus array in the source list (e.g., for the first encounter, $Source = (I)$). The further the correct stimulus is from an end-point in the source array, the more marks will appear in the source list (e.g., for a middle item within a 7-item stimulus array, we have $Source = (I, II, III, IIII)$). When the agent eventually encounters the correct stimulus in the source array, the last marker in the source list now defines the count of scans the agent must make from the end-point of the target stimulus array to identify correctly the target stimulus.

When an agent scans the target array from one of the end-points, and encounters a stimulus that is not in the required ordinal position (in this example $IIII$), it marks this encounter in the target stimulus array in the target list (e.g., for the first encounter, $Target = (I')$). On a tally mark match between source and target lists (e.g., $Target = (I', II', III', IIII')$ and $Source = (I, II, III, IIII)$) we have an indicator that the agent has reached the required ordinal position in the target stimulus array.

This mechanism is driven by knowledge in *LTM* that each stimulus in the source, and target arrays have corresponding slots. Slots have the effect of framing each perceived stimulus as a discrete set, the "smallest interval" encounters between stimuli thus being countable. A slot contains temporal markers that are stored in WM but only until the correct slot is found in the target array.

Ordinal Search Model Simulation

The aim of this simulation is to provide an answer to question 3 stated at the start of Chapter II, regarding the emergence of ordinal competence from spontaneous size sequencing skill. As within the principled search model, the agent has the competence for spontaneous, error-free sequencing via a recursive "select smallest size difference" rule. This routine (*PRINCIPLED SELECT*, as detailed in Chapter III) is augmented with the tallying mechanism. The general methodology is identical to that for the simulations reported in Chapter III. As with spontaneous seriation there is no validation data from ordinal matching by 7-year-olds due to the sparseness of their error. The validation is based on the ability of the model to simulate highly successful ordinal matching. Simulations of 40 trials each were executed representing the model. The task representation was analogous to that provided to the 7-year-old child participants using the 7-item sets (Experiment 5) and applied the same training and feedback. The simulation pseudo-code is as follows:

```
simulation ← 10
trial ← 40
n ← | C_n |
Slots_n ← LoadSlotsFromTransitionalModel
Mode ← Source
for simulation simulations
  for trial trials
    SizeDifference ← SCAN (C_n)
    selectedInWM ← PRINCIPLED SELECT (SizeDifference)
  If Slots_selectedInWM = TRUE then
  If selectedInWM = nextStimulusFromEnd then
      inhibited ← INHIBIT (selectedInWM)
   If selectedInWM != flashingTargetStimulus then
        TALLY (selectedInWM, Mode)
  If selectedInWM = flashingTargetStimulus then
    If Mode = Source then
    Mode ← Target
    If Mode = Target then
        selectedOnScreen ← TARGET SELECT (selectedInWM)
     If TALLY MATCH (Source, Target) = FALSE then
          errors ← errors ∪ selectedOnScreen
```

```
else
    errors ← errors ∪ selectedInWM
next trial
next simulation
```

The simulation parameters are identical to the principled search model in Chapter III, in which the stimulus variance is set to 1.5. This means there will be a small amount of uncertainty on scanning the stimuli.

Results

All agents completed the tasks spontaneously with few errors (six in total), the result of a small amount of stimulus size confusion on selecting the minimal size difference between referent and target within the source and target arrays. The ordinal search model has thus shown precisely how an agent equipped with a rule derived from the transitional model's operation, with the additional assumption of a tallying routines, can achieve spontaneous, error-free ordinal matching.

Discussion (Ordinal Simulation)

The ordinal search model uses the representations in LTM created by the transitional model, which indicate to it that it can use a "select smallest difference" rule. Provoked by the ordinal task demands, the agent "shifts gear" and both leverages the power of the slots in LTM and its information in WM in which to store the results of tallying. The model represents a 7-year-old child that can seriate spontaneously, using only online sources of information to do so, but simultaneously use the slots in LTM and WM to augment this process with cumulative tallying. The "reality matching" limitations, as well as the faithfulness and plausibility arguments, made in Chapter III also apply to this model, so we focus here on the psychological evidence for the proposed tallying mechanism.

There is strong evidence that both children and nonhuman primates can count the elements of a sequence within an ordering task, should they be required to do so by the experimental design. For example, McGonigle and Chalmers (2006) trained *Cebus apella* monkeys to solve categorization tasks which required the counting of items in each of the categories. The monkeys were trained on a core spine, "touch the square, then circle, then triangle" (*ABC*), and then they transferred to an extended version, "touch all the squares, then all the circles, then all the triangles" (*AAABBBCCC*). As all the shapes were of the same size and color within each category, the monkeys appear to have been cumulatively tallying the number of stimuli in each to ensure no backtracking or redundant touches. A cumulative tally is, therefore, a faithful and plausible way of representing the information processing in play here, especially as the *C. Apella* monkeys were culture-free agents in the sense of having no numerical or linguistic knowledge, akin to our computational ones.

It is important to note that such a tally does not represent a stand-alone integer or cardinal value. Tied to a temporal procedure, it explains how a 7-year-old child can readily match within a set of different sizes without (necessarily) being able to deploy each tally as a stand-alone code, allowing an arbitrary sequence of ordinal values to be easily remembered. Although adults appear to remember the sequence relatively easily, the effect of levels of non-monotonicity suggests that they too use a monotonic sequence as a default structure for encoding ordinal values, as we also see in our last chapter.

V. Sequential and Ordinal Size Understanding: A New Characterization

Questions Addressed in This Monograph

We began our monograph by seeking out Piaget's arguments regarding, in particular, the understanding of the logical properties of discrete sets at the age of concrete operations at around the age of 7. Focusing on size seriation and ordinal size identification and how they bear on the emergence of the concept of a unit, we argued that key questions still remained to be addressed despite decades of subsequent developmental research on the broader areas that surround this topic. Those questions and others now arising can be summarized as follows:

1. Is there an important discontinuity in this particular domain of logicomathematical understanding at around the age of 7? Our answer to this will be in the affirmative (but we also note the limitations of our empirical studies in this section).
2. How can the discontinuity best be characterized at an empirical level and in terms of a computational model? In this chapter, we summarize our answers to these questions and go on to ask:
3. How does the resulting empirical and modeling characterization compare and contrast with the characterizations of others? As we have adopted an explicit training approach using touchscreens, we shall also ask:
4. What are the implications for possible educational tools? Finally, we go somewhat beyond the remit of the monograph and speculate on:
5. What happens after the age of 7 in this and related domains?

In Chapter II we explained why we have approached the issue of this developmental discontinuity through the relatively cultural-free task of size seriation and related ordinal size comprehension. Using a variety of sequential and ordinal tasks presented to small groups of 5- and 7-year-olds, we are now in a position to offer some definitive answers.

Is There an Important Discontinuity in This Domain of Logicomathematical Understanding?

Our short answer to this question is yes. Taking Piaget's size seriation and ordinal correspondence tasks as our point of departure, we have both reinforced and amplified the existing evidence that children undergo a radical shift in how they perceive, process and manage the information in sets containing multiple items.

Size Sequencing

Concentrating first on size seriation, we used a touchscreen paradigm to eliminate ad-hoc features accruing to the select-and-place paradigm first devised by Piaget and Szeminska (1941). With this procedure, we were able to measure size-sequencing abilities using a training paradigm, from which it was evident that 5-year-olds required trial-and-error learning to become completely successful even on 5-item sets. Children at this age showed considerable variability in the training effort required, as well as a high degree of sensitivity to a set size increase from five to seven items. The 7-year-olds, by contrast, were almost completely spontaneous in the application of a sequencing rule for both set sizes. We compared size sequence training with the training of an arbitrary color sequence in order to identify any aspects of the learning that were specific to ordering size relations. Highly similar to list-learning techniques used in memory research, this arbitrary sequencing task allowed us to obtain a stand-alone measure of WM. It showed that the age difference in size sequencing was accompanied by an improvement in WM from 5–7 years of age, raising issues about the extent to which age-related changes in WM can explain seriation development. We return to this important issue later. In the meantime, we review the reliability of our findings in terms of consistency with other research.

Consistency With Findings From Seriation Research

We commented earlier (Chapters I and II) on the lack of any direct challenge to Piaget's claim of a sudden change in seriation performance at around the age of 7. As noted earlier, this is due to the fact that many of the Piagetian follow-up studies were primarily concerned with cross-correlations across his tasks, reporting the strength of these intertask relationships rather than the core task results themselves (Achenbach & Weisz, 1975; Dodwell, 1960; Tomlinson-Keasey et al., 1979). Here we consider how many experimental studies have explicitly confirmed the transition to spontaneous seriation apart from our own. With a large variation in age, stimuli, and task presentation, it is difficult to pin down data that are directly comparable with those we have generated on our touchscreen version, but as far as we can tell, our results are completely consistent with those from studies investigating seriation development with real objects.

Elkind (1968) carried out a replication of Piaget's tests with children aged between 4 and 7 years. This study was aimed primarily at exploring variability arising from test materials. Elkind noted an effect of stimulus type which, unlike our own study, showed better performance with two-dimensional than one-dimensional differences, but he concluded that the study gave "no warrant for assuming the effects of the combination of any one material and any one test varies with age level" (p. 65). Elkind did not break down the accuracy of children's performance by age, but endorsed Piaget's description of the main stages of development, showing that children at stage III (around 7 years) construct a series of nine elements quickly and in a principled errorless manner. Little (1972) carried out a study primarily aimed at looking at IQ effects within a longitudinal study on some Piagetian tasks with children aged 4.5–5 years, and then again when the children were aged 6.5–7 years. One of the tasks was to seriate 10 Montessori cylinders. Little does not quantify the results either in terms of number of trial and error, or spontaneously correct sequences, but she categorized responses into three levels. The first indicated no real comprehension of the task; level 2 showed "intuitive reasoning," and level 3 showed "concrete logic." Using these criteria there was a shift in seriation performance across age for average and superior intelligence levels, though more so for the higher IQ group. An average IQ group (IQ = 95–104) were predominantly at level 2 performance when tested at around the age of 5 (80%), and that proportion dropped only by about 15% when tested two years later. A superior intelligence group (IQ = 115+), showed greater improvement and were predominantly level 3 performers (83%) by the age of 6.5–7. Assuming that level 3 encompasses what Piaget meant by operational seriation, we can tentatively conclude from this study that children with an intermediate IQ (that she does not report) will also fall into this category by the age of 7. Further conclusions are difficult to draw from the highly qualitatively way in which these data are reported.

A follow-up of Little's study (Tomlinson-Keasey et al., 1979) included a promising breakdown of seriation into 9- then 4-item sets for failed participants, followed by transfer to 7-, 9-, and then 14-item sets. Unfortunately, this titration of performance was used only to predict performance other Piagetian tasks within a longitudinal study. Consequently, for seriation, there is no reported breakdown of performance by age, and we learn simply that children with an average age of 6 years were 59% "concrete operational."

Research by Kingma (Kingma, 1983a, 1983b, 1984b) has provided potentially more instructive data from the classic task in that he used a version in which 10 upright tubes with 0.5 cm difference had to be arranged in order, employing five seriated puppets as a model. He also used a task using only four squares, and a weight seriation task using five cubes. All of these tasks were combined to produce a composite measure of "single seriation." Dividing his younger children by 6-month age intervals and the older ones by 1 year, the percentage correct scores for this single seriation condition for children aged 5, 5.5, 6, and 7 years were: 15, 47, 55, and 91 respectively. This is a strong replication of Piaget's original findings, despite the difficulty in breaking the data down any further by set size or object type. Similarly, although aimed primarily at looking at series recognition rather than construction, Blevins-Knabe (1987b) reported that a level of above 50% "operational" seriation was recorded only for her oldest group (average age: 6 years and 5 months).

In short, we can conclude from the literature that we are not contradicting any evidence revoking the discontinuity that Piaget originally observed.

Ordinal Size Understanding

Here we took Piaget's correspondence tasks as a first means of exploring this ability but in the context of exactly the same stimuli and similar training procedures to those we had used in the size-sequencing tasks. Discontinuity across age groups was extremely evident (again), but against a backdrop of very much poorer performance by 5-year-olds than they had shown in the sequencing tasks, such that two-thirds of the group failed to reach criterion by a total of 200 trials. The high training effort concerned the inner three items in particular, and error was not confined to immediately adjacent items but spread across the other inner items in the set. We explored possible task artifacts by introducing the task in easy to hard steps, moving from 2-item to 5-item sets, but this (still) resulted in nearly half the group failing by the 5-item stage. We then moved to a task that did not use ordinal correspondence, but the learning of associations between colors and size rules. Although it produced a slight improvement in the performance of 5-year-olds, once again we found an age-related discontinuity. When given the color associations within a trial block of five, 7-year-olds were just below chance on the first trial of each new trial block (where guessing was the only option), but nearly at floor level on the subsequent trials within the block, whereas the 5-year-olds took many more trials, and some failed altogether.

Moving to the second phase of our color conditional task, this phase, and all subsequent tasks we report, required that participants could access all the ordinal sizes from a more permanent LTM store, either by remembering the color associations in single random alternation as required by Phase 2 of the color conditional task, or, in the non-monotonic sequencing tasks, by remembering an arbitrary list of ordinal codes. Although these subsequent tasks were still subject to age effects, they ceased to produce the discontinuity observed in all prior experiments, in that 7-year-olds were now showing measurable difficulty especially on an extended set size. As we concluded in Chapter IV, this indicates that the LTM of the series as utilized by these children was of a procedural nature representing a unidirectional search to a given item from one or other end of the series. The prerequisite for the procedure appeared to be that the given item needed itself to be cued in advance, and also by a serial unidirectional search. We consider later the extent to which these conclusions point to a possible constraint on the ordinal capabilities of 7-year-olds. In the meantime, we consider the consistency of all our findings on ordinal understanding with regard to previous research.

Consistency With Other Findings on Ordinal Size Understanding

As mentioned in Chapters I and IV, similar studies looking specifically at ordinal size understanding are rare. Once again, we shall try to isolate those that have used age groups and tasks that we can compare with those used here. The evidence on ordinal identification comes from two classes of research: discrimination learning, and the Piagetian correspondence and insertion tasks using seriated sets of sizes. The discrimination learning data derive mainly from the field of transposition research spawned by the relational/absolute controversy of the behaviorist tradition. Some of these studies extended the classic binary discrimination to 3-item sets to see whether the ordinal rule middle-sized would be generalized to new values. Each transposition study thus includes an initial discrimination learning episode involving three stimuli. As far as children under the age of 7 are concerned, this has been recorded as showing either the same or greater difficulty levels as we found when we reduced the set size to three. Zeiler and Friedrichs (1969) reported a training range of from 7 to 18 trials to criterion combined across all three relations (biggest, smallest and middle-sized) and also combined across ages five to seven. In comparison, we recorded an average around 15 trials for 5-year-olds. Stevenson and McBee (1958), by contrast, reported an average performance of less than 70% correct by 6-year-olds on 3-item sets by Trials 31–40. One difference across these two studies is that children in the Stevenson and McBee study were allocated to only one rule per child suggesting that they may have lacked the impetus to discriminate arising from having to attend to all the relations in the sets given. As noted in Chapter IV, allocating a single rule per stimulus set is also a possible reason for the very high levels of difficulty found by Siegel (1972) for children up to the age of 6.

As for studies using children around the age of 7, most report evidence of the more advanced ordinal abilities in this age group and generally citing low levels of error similar to those we obtained here. We recorded an average of around 6.5 trials for all the rules in a 5-item set and around 8.5 for 7-item sets. Zeiler and Gardner (1966) found that 7- to 8-year-olds, when trained only on the middle-sized relation, took only 2.5 trials to start their criterion run. Yeh (1970), however, found that children as old as 10, (also trained only on "middle-sized") took around seven trials before they started the criterion run (of 10 consecutively correct choices—the stricter criterion being one possible reason for the discrepancy). Similarly, Siegel (1972) found that the reduction in error by the age of 7 came down to around 10 trials to the start of a criterion run—and here the criterion was set at 9/10 correct responses. Although it appears from these data that we can situate our results within the learning range shown in other studies, it has to be remembered that we trained identification of all five (and then seven) items—a task demand well in advance of any other study we know of. This showed very high robustness of identification abilities in 7-year-olds, placing them at a considerable distance from children just 2 years younger in terms of our own measures.

Explaining Inconsistencies in the Ordinal Size Understanding Literature

From the foregoing, it can be seen that research on ordinal size understanding casts some doubt on whether ordinal abilities are the reliable index of the age of concrete operations as Piaget implied. Whereas seriation research appears to confirm the sudden appearance of the ability to order fairly large numbers of items, no such age-related success clearly emerges clearly from post-Piagetian research in that there is high variability across studies, and there are some reports of considerable difficulty even by older children in simple discrimination tasks. We have noted the conditions in which we ourselves found ordinal abilities to be compromised and have argued it to be strongly related to the opportunities for seriating the items in question. In the light of this analysis, we can now explain these inconsistencies in the research literature. Were we to summarize the reason behind the high variability across different studies, we would say it reflects the extent to which children were provoked by the task into a serial search across all the items in order to find the correct ordinal value. This is true of both older and younger children as we review next.

In line with the suggestion that ordinal abilities can at least be facilitated in younger children by inviting a serial search, a follow-up study of Siegel's study, found that 5-year-olds showed an improved performance on identifying the "next smallest" in 4-item sets, if they were encouraged to point to the smallest item first (Gollin et al., 1974). A failure to spontaneously search the sample array, however, could explain the difficulties on our own matching and color conditional tasks with younger children. But it would also explain the positive transfer to the matching task that we found for children who had by then successfully learned the size rules in the color conditional task. In short, ordinal size abilities do not seem to be easily trained by simply asking children to discriminate sizes. If it can be trained at all in younger children, it is by establishing a search rule to allow them to locate the correct size within the relevant set. The rather extensive variability in ordinal identification in younger children across different investigations seems to depend a great deal on how far a serial strategy is induced by aspects of the task.

An explanation of the discrepant results with older children again resides in how we now believe rapid learning on ordinal matching is enabled in the first place. We have shown in our modeling that it would require spontaneous seriation of the sample array for that simulation to work, while also attending to the cumulative tally provided by that seriation (an additional task factor). Although we would not expect ordinal matching to be as error-free as seriation itself, we would expect on our model that it would be enabled by a task in which seriation to the appropriate item is directly solicited. This is exactly what the matching task does for the child who seriates spontaneously, that is, by highlighting the appropriate match and keeping it available while the corresponding item is found. No such prompts or cues to seriate are available in discrimination learning tasks. On this argument, correspondence and item insertion tasks with seriation opportunities should yield stronger performance with older children than the discrimination learning studies.

As we noted in Chapter IV, there are almost no studies on either ordinal correspondence or item insertion with older participants but if we turn to the earlier studies of Piaget and Elkind, it would seem that these two tasks do secure more robust performance by children aged 7 than the discrimination tasks mentioned above. When children are given sets that they have just seriated (Elkind, 1968; Piaget, 1952a), as in our matching task, 7-year-olds have little difficulty in identifying the ordinal position of internal items.

In summary, our empirical data on serial and ordinal size learning up to seven items, with its specific focus on age differences, helps to considerably expand and explain the most directly relevant literature. However, our data must also be scrutinized in terms of reliability and robustness and we consider these issues next.

Limitations of the Empirical Studies

As an exploratory program without many precedents in the literature, our experiments were all of a probing nature designed to tease apart various theoretical possibilities regarding the extent (and limitations) of children's capabilities in the area of size seriation and related skills. This was achieved by directly following up the implications of a previous experiment when it seemed appropriate. The investment on a series of small-sample studies however results in certain overall limitations that we comment on next.

Sample Size and Statistical Power

In our program of experiments, we spread testing resources across a series of connected studies which led to selecting small samples for each. As noted in Chapter IV, however, this raises the question of whether the small participant samples compromise our conclusions regarding discontinuity of expertise. Apart from consistency with other studies on this issue (see above), our own modeling puts us in a position to suggest that the discontinuity will be an inevitable outcome of age and insight through learning, and would always be found in tests of this sort whatever the sample size. Indeed, other modeling accounts have been offered without an informing database at all other than the qualitative descriptions provided by Piaget (Mareschal & Shultz, 1999). Modeling apart, however, it is always the case that the margin of error accruing to small samples can lead to false inferences, but it is also the case that strong effects should survive sampling error and should not require large samples to be easily detected. Although we did not pose every possible age effect in terms of formal hypotheses, furthermore, every main expected age difference tested in every experiment was statistically significant. The specific comparisons that we made with color sequence learning both in the data and in the computational model, also testify to the precision and specificity of the effects we did and did not expect to see in the size tasks.

What we were unable to offer, however, was statistical power based on strong parametric tests, but this was as much to do with the very discontinuity we were measuring as sample size per se. The price of showing discontinuity in a developmental investigation such as this is that the differences are nonlinear. When success is high and error at floor levels, there is no further basis for comparison in terms of choice. However, the corollary of the discontinuity we expected is that, before they converge on a single principled solution, younger children will vary in terms of how they tackle the task. This is consistent with the high variability that we obtained across our 5-year-olds in terms of overall learning speed but by the same token, it has to constrain any firm assertions regarding how the average 5-year-old behaves in such tasks and this would have to await further study with much larger samples.

Sample Composition

A similar point must be made with regard to participant characteristics, particularly with regard to IQ. We did not extend the testing to including IQ batteries, and indeed to draw any conclusions from this, we would, in any case, have needed very much larger samples. The school which we used had entrance criteria that allowed us to assume that none of the children were borderline intelligence (confirmed later with the school administration). The relationship between Piagetian tasks and IQ has been noted in other studies. We can only say that if our sampling was high-IQ biased, then the performance of 5-year-olds across a wider population may have been even more variable than we found here.

Limited Set Size

Another possible limitation on our claim of discontinuity is the relatively small sets of stimuli deployed, that is in using 5- and 7-item sets rather than the 8/10 normally employed. Does this raise the possibility that 7-year-olds may not have shown the spontaneity they evinced in seriating and matching correctly had the sets been larger? This is unlikely. The classic test results showing seriation expertise by 7 years usually are with larger sets of eight to ten objects as noted above. In fact, a more serious complaint might have arisen had we used of stimulus sets that were arguably too large. Having arrays as small as five items, allowed us to clearly demonstrate the limitations on seriation and especially ordinal identification abilities in the younger children. Although it has been noted (Koslowski, 1980) that sets as small as five might actually induce an unhelpful classification strategy in the classic seriation task (identify the biggest and smallest and place the middle ones accordingly), this criticism could not apply to our 5-item seriation task as we specifically trained a sequential response.

Transfer of Learning

A more serious restriction on our overall conclusions is the absence of any transfer of learning to tasks of a similar type. Training was used in this program as an in-depth means of measuring performance rather than as a means of accelerating development. Indeed, when our program was designed and timetabled, we were of course innocent of the information we now have. Nevertheless, on our own ensuing argument that exposure to relational aspects of the task lead to change and insight, then transfer to new problems could have provided a valuable test of this. Although the computational model was devised to account for the specific case of size relations, it contains features that are sufficiently general to apply to other types of dimensional difference. This could have been tested under transfer conditions and we acknowledge this as a limitation on the study as a whole. Model theoretic testing apart, long-term transfer of learning could be a major educational outcome of further study of size sequencing skills in school-aged children, and we return to this issue later.

How Can the Discontinuity Best Be Characterized?

Assuming the legitimacy of our findings, at least within these limitations, how can they be best characterized? Our characterization has been based on the circumstantial case arising from the empirical studies and also in much greater detail, on the basis of the computational modeling.

The Empirical Characterization

In terms of both seriation and ordinal size matching, the performance of the older children can be described as the ready application of a principled monotonic search to a set of discrete sizes. This is executed as if there is a preexisting understanding that every size represents a separate unit with a unique ordinal value. The younger children performed as if they were having to learn that a rough classification of items into small ones and big ones would not suffice, and that they had to learn to serially differentiate each one from the other. When the task simply solicited such a strategy implicitly (as in the matching task), the younger children were at a considerable disadvantage relative to their older peers. The inability for most participants to reach criterion on the three middle items of a 5-item set suggested that they were attempting some kind of absolute match across the sets rather than seeking ordinal correspondence through counting. Insofar as they succeeded at all, the length of training required in Experiments 4 and 6 (where training was graduated across increasing set expansion) suggested they did not approach the task with a sequential strategy in mind.

In short, all the empirical studies reported here indicate a shift in strategy from weak item differentiation to principled serial search, though the resulting solution does not accord with how the operational achievement is normally characterized, as we review below. We acknowledge, however, that the case is circumstantial and, as we have noted, interpretation could possibly be subject to claims of empirical weaknesses. A much stronger and more detailed explanation of how the age shift could occur came in the form of a computational model informed in the first instance by learning data from the size and color tasks given to 5-year-olds.

The Modeling Characterization

We characterized the empirical data computationally via a series of models presented in Chapter III (size sequencing) and Chapter IV (ordinal matching). Our models are motivated by the economics of information processing (Ballard et al., 1997; McFarland & Bösser, 1993) and architecturally aligned to simulated, embodied artificial agents (Parisi & Schlesinger, 2002) interacting with a simluated task environment. Our models show how LTM structures representing order may dynamically change over time via an active process of perception, learning, and action, this process buttressed by an increase in WM. The LTM representations changes after each correct action, from those allowing broad classification, to those allowing rankings of growing precision, and finally to those allowing the parsing of sets of stimuli into discrete units.

In Chapter III we argued that a Bayesian trial-and-error learning process creates a probabilistic serial order ranking within LTM (heuristic search model), which when combined with an increase in WM (transitional model), allows discoveries to be made about the utility of using the smallest available interstimulus relationships. This discovery allows an iterative "select smallest difference" rule to operate (principled search model). Specifically, we argued that certain LTM representations are seeded from a set of *Beta* continuous probability distributions with end-point biases. *Beta* probability distribution stability combined with interstimulus link stability create a slot in LTM representing a placeholder for the smallest-interval calculation. The modeling in Chapter III has three connected parts. First, it simulates a 5-year-old child who can sequence by trial and error (heuristic search model). The next (transitional) model characterizes a child between the ages of 5 and 7, whose WM has matured to a threshold level of power such that it augments trial-and-error routines with new discoveries. These discoveries concern the invariant properties of the search, such as the repeated application of a "smallest interval" rule. This rule becomes represented in the third (principled search) model in the form of an emergent algorithm that drives search in a monotonic direction. The principled search model is tested in a simulation that produces nearly error-free seriation as found in 7-year-olds. The simulation of progress to spontaneous seriation shows a distinct discontinuity of process from a learning heuristic designed to minimize error to an algorithm designed to control search.

In Chapter IV, we argued that the information-processing mechanism established in the operation of the principled search model in Chapter III is sufficient to represent ordinal competence, with the additional assumption of a cumulative tallying mechanism. The principled search model was thus put to work in the ordinal search model. We used the representations in LTM created by the transitional model in Chapter III (*Beta* distributions and the relational links between them), which allowed the "select smallest difference" rule operation. Provoked by ordinal task demands, slots that are available in LTM can facilitate the results of tallying in WM, a dormant competence not pressed into play in the principled search model. The Chapter IV model thus simulates a 7-year-old child that can seriate spontaneously, using only online sources of information to do so, but can now utilize the slots in LTM to augment this process with cumulative tallying in WM in order to carry over an ordinal match from samples to a target array. There is no suggestion that the slots represent anything other than a tally linked to a serial search strategy. Although they could in principle hold search-independent semantic information in the form a symbolic code like "middle-sized," this is counter-suggested by our results from the non-monotonic tasks. For the time being, we have to consider this possibility as representing an entirely different (or additional) solution that would require dedicated research and modeling using older children and adults.

Capturing EF Factors in the Model

Our models simulated children perceiving, acting upon, and learning from sets of objects. This was achieved via a simulated perceptual-learning-motor loop, and so architecturally the models touch upon the EF concepts of WM, LTM, RI, and planning (Miyake & Shah, 1999; Pennington et al., 1996). Each of these was made explicit in the model but varied according to task and level. The heuristic model had a default level of WM, allowing scores to be assembled and actions to be made probabilistically via the evolving *Beta* distribution rankings. The speed with which the agent evolved these distribution rankings was decreased by an increase to the *softmax* temperature.

Explaining more rapid learning by some younger children, however, did not in itself explain transitioning to principled search. For this, the transitional model had an enhanced WM in the format of a variable, *WMAvailability*, a probability value between 0 and 1 indicating the likelihood of WM remaining available for longer (Ashby et al., 2005), thus allowing the utility of attending to interstimulus relations to be discovered. The manipulation of *WMAvailability* within the transitional model allowed us to simulate individual differences in an agent's ability to reach LTM slot threshold values (i.e., the *Beta* distribution variance and interstimulus link strength thresholds). This, in turn, affected its ability to transition to the principled and ordinal search models and the usage of a "select smallest size difference" rule and cumulative tallying. This rule is effectively a plan, but it would be tautological to suggest that improvements in planning caused it to emerge.

As for RI, all models incorporate a routine that prevents, but does not stop completely, stimuli that have been selected from being re-selected by multiplying candidate values by very small constants. RI was manipulated selectively for the size versus color task as our parameter-tuning exercise discovered that the size condition model required a smaller inhibitory constant than the color condition model when simulating differential patterns of backwards errors (see Chapter III). In principle, these variables could be adjusted to capture variation across individuals relating to impulsivity. But there is no reason to believe that changes in RI could effect the changes that are brought about by the key interactions between memory resources and the physical environment. Our EF implementation contrasts with that of Houdé and Borst (2015) who suggest cognitive advance can occur through improved inhibition, allowing ineffective strategies to be overcome. We suggest that RI is an important learning constraint but one that works in conjunction with specific action routines and does not, by itself, explain the transitions across levels whereby new routines become available.

Thus, the heuristic, transitional, and principled search models each implement EF concepts somewhat differently, but arguably there is a developmental interdependency between the models and the EF factors themselves. The key change in level cannot in fact be explained by any single EF factor independently of the task. Rather, endogenous EF factors such as RI and WM interact with information sources external to the agent to procure (in the case of seriation) a replacement of a learning heuristic with a reliable algorithm. For example, when extra WM is made available to the agent at a transitional level it combines with the ranking being formed in LTM to enable the discovery of the "select smallest size difference" rule within the size condition. Performing the selection is an information-processing routine that uses information extracted exclusively from the task environment. The information utilized in the selection is not a given but depends on a threshold level of WM and LTM capacity. It also depends on the emergence of slots, which are assignations of each item to a single unique placeholder in memory—essentially the representation of a unit.

Finally, we do not see principled seriation as requiring greater complexity than that required by trial-and-error learning. The dynamic created by learning actually suggests a tendency for the agent to offload information processing whenever possible onto the external environment, as opposed to relying on the growth of ever more complex LTM structures. The cost of deploying extra memory resource is paid off by the discovery of an algorithm that reduces rather than embraces informational complexity. This is not to deny that some sorting and conditional reasoning tasks have levels of complexity that can only be conquered by appropriate inhibition and activation of certain high-level rules (Zelazo et al., 2003), but in the specific context of discrete set understanding, the modeling suggests that most generative and effective rule is to reduce potential complexity by exploiting redundancy in the physical environment.

Limitations on the Model Characterization

The models we have presented are serial as opposed to parallel in design and implementation, in that they are not made up of distinct modules which interact via messages, as per the COGENT cognitive architecture specified by (Yule et al., 2013). Furthermore, although having neuroscientific touchpoints (Ashby et al., 2005; Bogacz, 2017), they were not explicitly informed by neural structures. The agent and task representations were simplified as far as possible; perception and action selection mechanisms did not use simulated cameras or actuators, and the additional WM representation was collapsed into one variable per list element; inhibition and learning mechanisms used constants and difference equations, respectively. There are thus some limitations in terms faithfulness and plausibility. We emphasize, however, that our Bayesian ranking approach [Chapter III, Equation (4)] is potentially neurally coherent, despite requiring biological validation. As Jensen et al. (2015) put it in a different context:

> *a wealth of empirical evidence, particularly symbolic distance effects, suggest that a linear state space is essential. Additionally, massed trials will disrupt stimulus orderings unless some form of implicit updating is employed. At present, the literature is mute as to whether either an appropriately designed POMDP or the betasort algorithm can be instantiated in the brains of tilapia and tree shrews.* (p. 21)

However, the simplified nature of the models arguably makes them more transparent, and we emphasize that they should be considered as blueprint designs for more realistic models, especially as their validity is restricted to error distribution matching.

Is Our Combination of Cognitive Modeling Methodologies Problematic?

Our model capitalized on the strengths of production systems (explicitly representing knowledge as rules), connectionist and dynamical systems (representing learning and development), Bayesian cognitive models (explicitly representing knowledge as graphical structures), cognitive architectural approaches (explicitly defined components, e.g., storage buffers for WM and LTM), housed within a traditional procedural simulation. This methodological combination was a pragmatic decision to allow us to align the model's architecture, constituent components, and dynamics to the empirical phenomena we are trying to understand. In doing so, we allowed a sensible mapping between what Boden (1996) refers to as the *explicandum* (the phenomenon to be explained), and *explicans* (the explanation). This is a mapping that could not have been achieved by adopting any single computational methodology.

However, this synthesis may seem to contradict the position of theorists who advocate a purely dynamical system view of cognitive development (Di Paolo, Barandiaran, Beaton, & Buhrmann, 2014) and who eschew accounts of cognition that include hand-crafted (as opposed to evolved) knowledge structures, as they generally have a poor evolutionary story. Pollack (2014) summarizes this position:

> *As we've seen, mindless intelligence abounds in nature, through processes that channel mathematical ideals into physical processes that can appear optimally designed yet arise through and operate via exquisite iteration. The hypothesis for how intelligence arises in nature is that dynamical processes, driven by accumulated data gathered through iterated and often random-seeming processes, can become more intelligent than a smart adult human, yet continue to operate on principles that don't rely on symbols and logical reasoning. The proof lies not only in Markovian situations where a greedy sequential-choice algorithm driven by values converged under Bellman's equation, but also in the reliability, complexity, and low cost of biologically produced machines.* (p. 289)

Our response to this is twofold. First, the representations we propose in our models are not symbolic or logical in the traditional information-processing sense (Van Gelder, 1995). However, they can be considered to be explicit representations. For example, once unitized slots are formed, the requirement to create such slots can represent the necessary properties of any size series and be used to direct search in future tasks. Furthermore, they are grounded environmentally in real-world size differences (see Chapter III) in such a way that a specific common criticism of more traditional AI models does not apply. The "symbol grounding problem," which is the problem of how symbols get their meaning (Harnad, 1990) is avoided. Indeed, Pollack's (2014) and Van Gelder's (1995) dynamical systems accounts admit the necessity of representations, albeit not those of a symbolic or logical type. We thus see our approach as wholly consistent with the dynamical systems approach. Second, we suggest that once symbolic systems are in use by the complex biological system in question, computational formalisms that are different to those underlying dynamical systems are needed. There is a danger of mixing up explicandum and the explicans (Boden, 1996) in the case of representational growth in the child. However, this danger aside, the emergence of symbolic systems in evolution and development remain as phenomena that science has to explain. We suggest that, when externalized, symbolic cultural artifacts such as language and logic are involved, it makes sense to include computational formalisms that are representationally better suited to these problem domains. To this end, we see any future computational models of non-monotonic sequencing tasks as aligning to formalisms that incorporate the representation and manipulation of labels and symbols such as Logic Programming (Stenning & Lambalgen, 2012).

In short, a considered combination of computational models seems appropriate to explain cognitive growth, of which dynamical systems theory is surely an integral component.

Two other limitations of the model are first, that our model does not exhaustively address all aspects of the child's behavior, and second, that our model has not been tested for its utility for designing control mechanisms for intelligent systems.

As for the first, the perceptual scanning algorithm we propose for the agent's extraction of information from the stimulus array (see Figure 6) could be verified against reality by measuring eye movements made by the child while executing size-sequencing tasks with eye-tracking technology. The perceptual scanning pattern we propose could also be mapped on to RT data we have not used for validation purposes to date.

As for the second, the equations and data structures proposed should allow a suitably designed mobile robot to acquire serial ordering competence. Contemporary cognitive robotic architectures, such as the iCub humanoid robot used to investigate finger counting (De La Cruz et al., 2014; Metta et al., 2010) and the Khepera robotic vehicle used to characterize embodied serial order using dynamic field theory (Sandamirskaya & Schoner, 2010) could enhance our models with a parallel architecture and an actual (as opposed to simulated) real-time perception, learning, and action loop.

More generally, current neuroscientific theorizing on the *free energy* framework for modeling perception and learning (Ballard et al., 1997; Bogacz, 2017; Clark, 2016) is especially relevant to our position. The basic idea here is threefold. First, thermodynamic free energy is a measure of the energy available for a complex biological system to do useful work. Second, the closer a probabilistic model maintained by such a system is to reality, the less free energy is in play. Third, such systems are motivated to keep unexpected events posed by the environment to a minimum, and to do this they seek to minimize free energy. The result is a system that maintains a more accurate, and thus more adaptive, world model, with overall uncertainty (termed prediction error) reducing over time. Our characterization as to how (Beta distributed) probabilistic representations of serial order gradually increase in precision (and so reduce in uncertainty) over time is a potential developmental psychological touchpoint for this framework.

This framework also potentially aligns to the engineering of intelligent systems, such as simulated and robotic agents, motivated only by a need to minimize entropy. Further discussion of this area can be found in Clark (2016), and a hands-on tutorial introducing the free energy framework within an experimental psychological context can be found in Bogacz (2017).

How Does the Resulting Characterization Compare and Contrast With Those of Others?

It is not easy to situate our account with any existing theory as it the characterization we offer occupies a space of argument and empirical discovery that does not quite exist elsewhere as we discuss next.

Executive Functioning

Although we have made extensive use of concepts deriving from the neuropsychological research into EF, the breadth of ongoing debate in this area is, for the main part, far too wide to locate our findings easily within that field. Used ever more widely in education and intervention, EF is still the subject of much debate about how it should be defined (McKenna, Rushe, & Woodcock, 2017). In the main, we can say that our analysis aligns more with the idea of a fractionated EF approach in which the is a complex interplay of components as the child develops as opposed to a single factor model that depends on the maturation of a single EF resource such as WM capacity. But even within more fractionated approaches, many issues remain unanswered (Miyake et al., 2000). Specifically, questions have been raised about the separability and interconnectedness of its putative components including the usefulness of the idea of a central executive (Wasserman & Wasserman, 2013). A recent meta-analysis of neuroimaging studies (McKenna et al., 2017) confirmed that the partly isolable component of information updating was itself related to increases in WM and the authors comment that:

> *A focus for future research may be to assess the development of both dimensions of updating during childhood. And examine if there is a temporal link between improvements in WM capacity and the advancement of the executive component of updating and updating-specific abilities.* (p. 12)

Given that we have specifically considered the relationship between WM, the updating of information and emergence of a new plan or routine (planning being another EF component related to frontal lobe functioning), we would appear, on the face of it, to fit well with this requirement. But conventional tests of EF rarely directly assess the interconnectedness of components within the tasks themselves; rather they assess these through isolated individual component tests such the n-back task (a simple recall task) for WM retrieval, and the Stroop test (overcoming a conflict between meaning of a written color word and its font color) for RI. We would certainly endorse the view that separable EF factors are a good guiding principle (Diamond, 2013; Miyake & Shah, 1999), but until such time as imaging studies are carried out on our size-sequencing tasks, we cannot use neuropsychology of EF as it stands to further inform our position. A further concern would be the general lack of attention to visual scanning as a contributor to updating in EF especially as the age of 6 years has been particularly characterized as showing changes in visual-motor search and planning routines (Welsh, 1991).

Embodied Cognition

At its most general, our position is close to what is described as *embodied cognition,* in which we hope to have gone somewhat against the tide, as noted by (Marshall, 2016), of computational approaches neglecting the role of bodily action. He notes that computational approaches are gradually improving in terms of acknowledging the links between perception and action in explaining how the world comes to have meaning for the agent. He also endorses newer multilevel Bayesian approaches but is concerned that these too fail to give full consideration of the “role of the fully embodied person in relation to processes of thought and reasoning” (p. 246). In this monograph, we have tried to bring back the acting and perceiving child into understanding seriation development. We also feel we can defend ourselves against a general criticism that Marshall makes of the downplaying of constructivism in the new information-processing theories that arose when Piaget’s influence lessened. Though very concerned with the moment-to-moment processing of information, our approach differs markedly from the information-processing theories that look to fixed maturational factors and formal measures of task complexity as a key to understanding cognitive development. In the following sections, we expand on our own interactionist position and in so doing, identify the broad theoretical claims from other approaches that we support and also those that we repudiate—at least as they apply to the understanding of discrete sets.

A Discontinuity in the Acquisition of Necessary Knowledge

What we have conserved from the classic and similar neo-Piagetian accounts is that there is a discontinuity in behavior with discrete sets of items varying in size at around the age of 7. This is manifest in the way older children set about ordering differing sizes or finding ordinal matches across sets. The spontaneity and correctness of older children in seriation and ordinal matching tasks regardless of set expansion contrasts with the high variability of speed of learning and accuracy in younger children across individuals and set expansion. Apparently based on seeking the smallest divisible unit before commencing a search, this understanding of unitization could be described as a universal logic on which all children seem to converge. What we have not conserved is the idea that it is prompted by a grasp of the logical properties of discrete sets. It is important to note that this did not arise by an effort to directly refute Piaget so much as to reconsider what the emergence of seriation represents in terms of cognitive control and development.

Piaget's Agenda Versus Ours

As described in Chapter I, Piaget was motivated to find the origins of logical operations, and there can be no dispute over the logical properties of discrete sets as he defined them. On his end-state defined view, precursor behaviors are about discovering the interconnectedness of these properties by acting on objects. Our own agenda was not defined by this particular characterization, but rather by simply considering how and why children learn to differentiate items to procure size seriation and ordinal expertise. The "how" we have argued to arise from the emergence of a unique algorithmic solution that may be subject to certain executive factors in its implementation such as WM. The "why" is based on the need for all intelligent systems to manage information in the environment in the most adaptive and efficient way (see Chapter III on the theoretical assumptions behind the modeling). We now elaborate on how this directly contrasts with the implications of an operational account.

An Informational Rather Than a Logical Structural Approach

A key difference between our own and the logic-inspired account of Piaget lies in a task analysis based on its information-processing requirements as opposed to its logical structure. This difference creates two very different scenarios for how we conceptualize cognitive advance in the domain of managing size relationships. For example, a seriation task using n elements provides the same potential informational load to a 5-year-old as to a 7-year-old, but we have argued that the older child develops a strategy that reduces the potential amount of information in a multiple item set of sizes by monotonic organization. On being faced with a 5-item size-related set, we hypothesize that the 5-year-old child acquires and maintains in LTM a rank order representation ($Beta_1 \gg Beta_2 \gg Beta_3 \gg Beta_4 \gg Beta_5$) via perceptual learning. The 7-year-old child also maintains this ranking, but crucially it is augmented with a set of "smallness interval" relations linking adjacent set elements ($Beta_1 r Beta_2\ r\ Beta_3\ r\ Beta_4 r Beta_5$). Thus, on being presented with the same 5-item size-related set, the older child will follow the same pattern of pairwise comparisons, but as opposed to using probabilistic guesswork, a smallest interval rule is now used. The older child now knows that if the smallest size difference is selected from the set, the correct answer is guaranteed. They can now select the minimum from a list of size differences after having scanned the set of candidates, whereas previously, each candidate has to be scored, a much more complex calculation. In sum, a lean, efficient representation that emerges via perceptual learning in the 5-year-old allows more efficient behavior in the 7-year-old.

If, by contrast, we consider an alternative consistent with Piagetian relational reversibility, we can see immediately that this scenario is much more costly to the system in information management terms. Take the example of a 7-year-old child seriating a 5-item size-related set, $\{A > B > C > D > E\}$. At each point in the set, at least four mental (as opposed to perceptual) comparisons will have to be made. With A as the target, we have $\{A > B, A > C, A > D, A > E\}$; with B, $\{B < A, B > C, B > D, B > E\}$; with C, $\{C < A, C < B, C > D, C > E\}$; with D, $\{D < A, D < B, D < C, D > E\}$ and with E, $\{E < A, E < B, E < C, E < D\}$. The comparisons, aside from being psychologically very complex due to their bidirectional nature, are combinatorially much more complex than our perceptual learning account. Whereas 10 unidirectional, perceptual comparisons are necessary to allow a rank order to form in our account, this scenario shows that 20 bidirectional, mental comparisons are necessary to form such an order. This can be described as *omnidirectional* access to each item as opposed to unidirectional access. Relational reversibility is thus not required to explain an increase in the total amount of information gained, and indeed it is anathema to that outcome. We expand on this below with regard to traditional seriation and transitivity.

Relational Reversibility

What we specifically reject from the classic account and some neo-Piagetian theories is the idea that changes in discrete set understanding are brought about by relational reversibility, related directly to a coordinate logic in the mind of the concrete operational child. Piaget's contention was that the serial, ordinal, and cardinal properties of relationally connected items become understood through a mental coordination prompted by the discovery of the reversibility of an asymmetric relation (see Chapters I, II, and IV). This endured (and still endures) in many post-Piagetian accounts. Taking a classic Piagetian view, Elkind (1968), for example, says:

> *To construct a series systematically, as third-stage children do, it is necessary to attribute to each element s > r and s < t because (to form a series) each element is chosen so that it is smaller than each element that follows it and larger than each element that precedes it.* (p. 70)

The concept of integration and coordination of relations was cited subsequently by several others in the context of seriation (Moore, 1979), and especially transitivity, where a transitive choice was taken as evidence of the logical coordination of pairwise relations (Bryant & Trabasso, 1971; Russell, 1981), or a “higher” level (Rank 4) of understanding than one based on temporal chaining of trained premises (Halford, 1993).

We dispute the case for reversibility in the case of seriation and ordinal size understanding on the grounds that it is neither necessary nor efficient. We have illustrated how the expert seriator must choose an element that represents the smallest difference from the preceding element. Although this requires scanning the likely alternatives in a forward search, it does not require placing or knowing anything about the next following element until that element is itself chosen. This results in the highly efficient skill of principled monotonic search every time a discrete set of elements has to be organized and identified on an individual basis. When identifying a specific ordinal position, in larger sets, the scanning can proceed from either end of the series depending on an initial identification of the nearest end-point, but this still amounts to a unidirectional rather than a bidirectional processing of the relational information for the purposes of identifying any given element.

We are not alone in eschewing the concept of reversibility in a seriation context; others have simply dropped the concept or failed to find any direct evidence for it (Blevins-Knabe, 1987a; Chapman, 1988; Fürth, 1969; Leiser & Gillierion, 1990; Schliemann, 1983). This has been particularly true of modeling accounts. In an attempt to cohere Piaget’s theoretical model with an information-processing account of seriation, Leiser and Gillierion (1990) set out with the intention of producing algorithms that would capture both the observed behavior (with children aged between 6 and 10) and also Piagetian structural concepts. But despite their strong initial allegiance to the classic Piagetian framework, Leiser and Gillierion (1990, pp. 169–182) found little support at the end of their study for characterizing seriation development through the emergence of reciprocal relational reversibility within an order structure. Although they found a propensity in all their participants to use an extremum method during series construction task (wherein the largest element is found and placed alongside the previous largest element), they did not find concomitant evidence of even any understanding of the need to keep repeating the operation even in their oldest participants. As the authors put it, “somewhat surprisingly, inventing the procedure does not imply understanding it” (Leiser & Gillierion, 1990, p. 175). This was a finding echoed by Baylor et al. (1973) who modeled the behavior of three children on size and weight seriation tasks but did not find a role for reversibility in their computer simulations and concluded, “These are aspects of Piagetian theory that have not yet found adequate non-trivial representations in (the) information processing models” (p. 195).

As for transitivity, there are several difficulties in evaluating a coordinate model of asymmetric relations in this context. First, for those promoting this hypothesis, it requires constructing test situations that rule out the possibility of nonlogical solutions. As we reviewed in Chapter II, the literature on whether spatial, temporal, or perceptual supports were in fact effectively ruled out in transitivity training tasks has been a long-standing debate

(Kallio, 1982; Perner & Mansbridge, 1983; Schnall & Gattis, 1998; Thayer & Collyer, 1978; Wright et al., 2011; Youniss & Murray, 1970). The ample evidence arising was that temporal—and even spatial—encoding of a set of trained relations is the default means of learning the premises in transitivity tasks by children (and even adults, as we review later). This is not simply a methodological point. It raises the questions as to why, in any case, would a human reasoner adopt a more complex omnidirectional conceptualization requiring large increments in memory resource.

The idea of omnidirectionality, moreover, does not readily scale up to the case of seriation, where access to every possible relation would put a very high demand on memory resource as formulated above. An obvious rebuff to this is that seriation is by nature a temporal activity that may not require an atemporal apprehension of the relational structure of the set. If this were true, then it is not clear why it takes at least as long to become a manifest ability in school-aged children as genuine transitive reasoning. In short, and as we see later in the case of adult transitive reasoning, the application of monotonic search can readily explain how transitive deductions are made without the assumption of further complexity.

Verbal Justifications Suggesting Reversibility and Logical Awareness

One important element in the arguments supporting reversibility has been the nature of verbally retrieved information after test responses. Justifications given by older children showing the active retrieval of both premises in transitivity tests—or the adjacent relations in seriation tests—could suggest an act of coordination (Elkind, 1968; Wright et al., 2011). But these are after the fact, and do not imply that this information is recruited to solve the problem in the first place. Nevertheless, we do have to ask what linguistic or symbolic access actually may be telling us about underlying processes of which the child may not be directly aware. This has broad implications for the relationship between thought and language going well beyond the remit of this monograph, so we shall confine ourselves to specific observations regarding our own evaluation of verbal reports. We did not ask children to justify their responses but simply to describe what they remembered about the about the sequential and ordinal tasks. Although the degree of linguistic specificity was definitely more advanced for 7-year-olds than for the younger children, it seriously underestimated the accuracy of the performance in all tasks given. This was also true of the eight adults we tested by way of comparison. The lack of a distinct relationship between verbalizing and performance indicates that the children are reflecting on what they know and remember—not the processes by which they came to that knowledge. On that basis, it would be relatively easy for a child who has already seriated three items, (e.g., yellow > red > blue) in a transitivity test, to retrieve both relevant premises to produce a justification answer that could imply relational coordination such as "because the yellow stick is longer than the red stick and the red stick is longer than the blue one." Unsurprisingly, therefore, Glick and Wapner (1968) found no consistent association between correct transitive choices and verbal justifications for participants between 8 and 18 years of age.

In short, we have no evidence that spontaneous seriation, and seriation as a determiner of ordinal position, involves or requires reversibility of thought, nor any awareness (even by 7-year-olds) that the relations thus computed form a fully coordinated logical structure. The point we wish to stress however is that this in no way compromises the unique and powerful informational status of the solution deployed by spontaneous seriators when confronted with discrete sets. Indeed, as the necessary substrate for all mathematics based on the properties of numbers, it is as arguably as important to human knowledge as the deductive inference.

The Role of Perception in Conceptual Understanding

The foregoing observations lead us to what we see as our final departure from Piagetian and many post-Piagetian accounts. Transitivity research was popular because it appeared to be conceptual rather than perceptual and this reflected the mindset of many developmentalists then and now. Piaget believed that to make the transition to operational thinking in transitivity, seriation, and related domains, reliance on direct perception must be relinquished, a view echoed in neo-Piagetian terms by, for example, Halford (1993) and Karmiloff-Smith (1992). This supposed emancipation from perceptual information is not required by our own characterization, and it is actually counter-suggested by the conditions under which our 7-year-olds succeeded. Showing little or no error during the normal perceptual conditions of seriation tasks, the expert performance by 7-year-olds collapsed when the supporting perceptual array was removed. This support was provided by a target item being highlighted on every trial within a sample that could be seriated through visual interrogation. It was also provided when the target item was identified on the first trial of a single set in the color conditional task, and the child simply had to perform a similar interrogation and count on the subsequent four trials to select it. These are the conditions that also prevail in all the Piagetian tasks of seriation and ordinal and cardinal correspondence.

We have no argument with the idea that children have to acquire a concept of a series if they are going to execute a principled search when presented with a set of jumbled elements. What we take issue with this is the idea that in order to form this guiding concept children need to be "freed" from perception. Our analysis tells us that the advance to regarding every member of a set as occupying a unique class of its own is an act of perceptual learning. In support of this view is the fact that blind and partially sighted children have been reported as lagging behind in size but not weight seriation by up to 3 years (Friedman & Pasnak, 1973; Hatwell, 1985;

Lebron-Rodriguez & Pasnak, 1977) but we note these with caution as others have failed to find a difference (Lister, Leach, Ballinger, & Simpson, 1996).

The modeling approach in this monograph reinforces the importance of perception as a causal ingredient in knowledge gain. Although there is a very large literature on how cognition can exert a top-down influence on perception (see Firestone & Scholl, 2016, for a review), the converse is rarely argued for. The computational modeling here offers a special window on a bottom-up influence as to how an agent may, through perceptual learning, grow the representations that allow it to conceptualize the world. It has shown that the outcome of learning is not just a better-differentiated set of items, but a perceptual rule that can guide future searches and provoke ever leaner representations. The perceptual field is an essential causal element in this, as well as in guiding the search thereafter.

Perceptual Learning in Cognitive Development

In reinstating the importance of perceptual activity in a traditionally cognitive domain, we note that this goes against a rising trend. From the early 1970s onward, Pick (1992) has noted that interest in perceptual learning has "waned" (p. 791). But we must also clarify what we mean by perception in this context. We have used the term perceptual "differentiation" to describe the progression of sequential and ordinal understanding from 5–7 years of age, but here we have to be clear that we are not describing an improvement in acuity. Although there are some developmental changes in visual acuity up to 9 years of age (Semenov, Chernova, & Bondarko, 2000), no study of perceptual discrimination has suggested that children of school age cannot detect a size difference with the interval ranges used here. As noted in Chapter IV, however, size discrimination ability in children up to 10 years of age can vary greatly as a function of how the discrimination tests are presented. We noted in Chapter IV that presenting them in a serial context was likely to improve performance. May and MacPherson (1971) showed, moreover, how exposure to variation in the minimum size interval across items to be discriminated dramatically reduced errors in two-choice size tests in children up to 9 years of age. Such task factors affecting discrimination learning were considered extensively in the light of attentional and reinforcement mechanisms arising from the behaviorist tradition (Reese, 1968), which then became outdated as discrimination research itself dropped out of fashion. Based broadly on the idea that responses have to be appropriately tuned to relevant environmental stimulation, this does not, in any case, capture the role we give to perceptual activity here. As detailed in the computational model, it would be more accurate to say that the environmental stimulation from a set of sizes itself becomes "tuned" to the act of serial scanning. Differences do not become more noticeable; they become more informative to the perceiver during perceptual learning. Searching for the relevant difference starts to control the visual scan in a cycle of feedback-informed behavior. The end result is a planned search starting with the biggest/smallest element, next biggest/smallest, and so on.

Although she considered perceptual differentiation with regard to distinctive features of objects rather the stimulus relations, our characterization of perceptual learning is not dissimilar to that of Eleanor Gibson. In developmental studies with letter-like forms as stimuli, Gibson, Gibson, Pick, and Osser (1962) showed how children learn to attend to dimensions of are relevant to the class at hand such as rotational differences (that would convert a "b" to a "d") but not slant or tilt (a legitimate source of variation in writing). Applied to size differences, we could say, similarly, that children learn the ecological importance of detecting the smallest interval in a set to enable them to cope with the total amount of information that that such a set could potentially convey.

Conceptual Development from Perceptual Activity

By situating perceptual activity at the heart of essential information gain, we eschew the sharp distinction between perception and conception that Piaget and some others have drawn. Our view of perception thus shares much in common with the approach of Goldstone and Barsalou (1998) who have argued that a rather false dichotomy between perception and cognition has understated the role of perception in "grounding" conceptual distinctions and subscribe to their view that: "many mechanisms typically associated with abstract thought are also present in perception" (p. 231). Described in the context of classification and similarity judgments, their thesis can also be seen to be directly applicable to how we have considered perception to work in order to procure real representational change in seriation, and why we repudiate Piaget's rejection of perception in the cognitive achievements of sequential and ordinal size understanding:

> *Conceptual end states do not imply an absence of perceptual origins. Even if the end-state of a concept were free of perceptual information, perceptual processing may have been required to build it.* (p. 244)

We submit that a concept of a series would be one such example.

Perception and Transactionalism

Though clearly different from Piaget's concept of "operationality," in a broader sense our position endorses Piaget's claim for a "genetic epistemology" of logicomathematical knowledge. Transactionalism is an extremely appropriate way to cast the way in which children come to grasp the fact that all sets can be unitized through the personal experience of interacting with actual objects through sight or touch. From a starting point where any two objects can be differentiated, sets of multiple objects can—and apparently do—reside in the minds of young children without being seen as logically discrete or orderable entities. This is a hard concept to grasp for the adult mind, but the robustness of difficulties in seriation—and especially ordinal identification—recorded here and elsewhere testifies to the fact that sets of divisible entities do not automatically impose their numeric properties on the human mind. Making those properties discoverable by interrogation and selection is, at the very least, a highly plausible explanation of how every individual comes to construct that knowledge. We realize that there are much wider arguments here regarding the external reality and existential truth of mathematics itself and its evolutionary history (Lakoff & Nunez, 2000). We are not in a position to make any bold claims about mathematics as a cultural construct, save to say that if one were seeking a universal primitive for such a construct, as Piaget argued over half a century ago, it can be found in the developing behavior of the young child.

Maturation as a Sole Cause of Discontinuity

The foregoing should make it clear why we do not endorse the view that discontinuity is simply brought about by the ineluctable growth of WM capacity. In our view the growth and role of WM (and associated LTM stores) has to be seen as more than providing a conceptual space for the child to move into, but rather as an intrinsic feature of the child/environment interaction. The modeling has detailed how this would work in two distinct ways. First, it operates within a trial-and-error level when participants are actively learning to discriminate the sizes, contributing to the high individual variability associated with this stage. However, it also calculates a threshold level that all systems would require to meet to be in a position to benefit from this learning, suggesting a possible minimum age of acquisition. Individual differences across children would be expected to have long-term consequences for how quickly the critical threshold level is reached for discovering the universal properties of a discrete set, with consequential changes on the contents of LTM.

The acquisition of a new (unvarying) skill has the effect of democratizing cognitive advance in this area, creating something of a level playing field in which LTM contents guiding sequencing and search become universal. This is not to deny, however, that the individual differences noted at much younger ages in terms of a sense of approximate number (ANS), for example, won't have long-term implications for later math achievement. Research investigating these individual differences in terms of both behavioral (Odic et al., 2016) and neural measures (Hyde, Simon, Berteletti, & Mou, 2017) points to such a correlation, though not without a certain confounding with math education itself as observed in Chapter I. Individual variation, notwithstanding, all children in most cultures need to acquire insights in the necessary rules underwriting number understanding. This leads us to the issue of the optimum learning environment for supporting the feedback loop between visual interrogation and action during learning, as we consider next.

The Implications for Possible Educational Tools

We noted in Chapter II, that many of the specific seriation training studies were aimed at looking at transfer to other Piagetian tasks or general effects on IQ (e.g., Kingma, 1984b). A few researchers have, however, looked more precisely at the effects of sequential training on mathematics scores, and some have endorsed the benefits of such training on children's math achievements in school, albeit cautiously (Pasnak et al., 2015). Children are becoming increasingly exposed to counting, arithmetic and the cardinal meaning of count words during the preschool and early school years, and there has been considerable interest in which ability trained in preschool will best predict later arithmetic problem-solving (Chu, vanMarle, Rouder, & Geary, 2018; Geary, vanMarle, Chu, Hoard, & Nugent, 2018; Geary, vanMarle, Chu, Rouder, et al., 2018). Some have specifically noted the predictive value of seriation abilities in preschool children for later arithmetic ability or number-line (Kingma, 1984c; Stock, Desoete, & Roeyers, 2007, 2009).

We would now submit that it is time to be much more precise about the nature of educationally-oriented training as far as discrete set understanding is concerned, and the age at which it is likely to have greatest effect. Beneficial effects have been reported by Pasnak et al. (2015) with 6-year-old children using what he describes as "sequentiation." This was a multifarious set of tasks where children have to try to determine the overall pattern or sequence in letters, numbers, clock faces, and rotated objects. The sequencing itself was of the item insertion variety where children had to find the correct place for a missing item or complete the sequence in the manner of the "Pattern Understanding" subtest in the Kaufman Assessment battery for children (Kaufman & Kaufman, 1983). It is hard to tell which specific types of material or instruction benefitted the children's performance on mathematics scales, but it does encourage exposure to sequential behavior as an educational tool. If we were to follow this further into the construction of an educational game, the implications of our analysis become relevant. We have noted strong consistency between our age-related results and studies using the classic seriation task, and we ourselves found a similar pattern of set-size-related difficulty in 5-year-olds when training on the classic select-and-place task. This does not, however, necessarily implicate a strong educational benefit of ordering with sticks or Montessori blocks. We have already commented in Chapter II on how the select-and-place method might both help and hinder the process of seriation in the younger child. As long as there are no visible consequences of an incorrect choice, the temporal operations of perceptual tuning can apply to this task as they do in our touchscreen tasks. What happens, however, when a selection results in an incorrect placement and the array under construction is disordered? The inability for a child to correct a seriation mistake is a well-known (almost defining) aspect of preoperational behavior, and it is easy to see how this could occur. Correction implies that the participant can perform the very act of perceptual differentiation that they are still acquiring. Conversely, as also noted by Young (1976) there should also be enabling aspects of spatial seriation, where correct choices become translated into an ordered visual array, thereby reinforcing the correctness of the choice. A testable hypothesis from this conjecture was reported in McGonigle and Chalmers (2001, p. 274). Sequencing both as an activity in itself but also in relation to creating a visual array is a field wide-open to the testing and development of educational tools.

As for timing, many have noted that training can be wasted on children who are too young or not cognitively ready for it (Kingma, 1986; Pasnak, Hansbarger, Dodson, Hart, & Blaha, 1996). Any approach that endorses the role of maturation would caution against the efficacy of imposing on children who lack the appropriate processing capacity, tasks that are too complex for their age (Case et al., 1993). Our identification of a threshold-critical WM factor is consistent with this view. This does not mean, however, that development should just be allowed to take its course without intervention or educational tools aimed at cognitive advance. Although the right type of experience at the right time could enable transition to new ordinal and cardinal awareness, there is no evidence that size seriation in and of itself has been turned into an educational tool. An obvious exception is the use of Montessori blocks in preschool, but although the value of Montessori education has been reviewed (Lillard & Else-Quest, 2006) we can find no data on the specific efficacy of the blocks-based teaching for serial abilities. Simple sequential touchscreen training on sizes (and other dimensions) with (as yet unexplored) possibilities for optimum visuo-spatial feedback seems to be a candidate case for developing a targeted teaching tool for children of around 6 years of age.

What Happens After the Age of 7?

Our conclusions regarding the nature of so-called concrete operational shift at around 7 years have implications for how we might consider what happens next. This is clearly too vast a subject to embrace within a monograph dedicated primarily to the 2 years between 5 and 7, but we can be fairly definitive about what we think is unlikely to be a direct successor to what we have found. The clues come from our nonmonotonic task. The contrast with expert adult performance makes it tempting to conclude that we have uncovered a constraint in the child's grasp of ordinality and that further developmental advance will automatically take place once that constraint is overcome. There are two obvious constraints that could apply. One is WM resource; the other is the way in which the tally is encoded in memory in the first place. If it were to be seen as some sort of restriction, however, the first is hard to reconcile with the fact that 7-year-olds showed that they were able to retain a list of seven colors after an average of fewer than 10 trials. As for the second, we could say that to effectively remember the nonmonotonic sequence, a new form of representation needs to form in the LTM slots that we think represent unique units. For this, each tally would need to go onto to be represented at a new symbolic level, either as an abstracted cardinal value (a number) or as a stand-alone ordinal description such as "third biggest." A set-independent value—an integer or ordinal value—should allow a quick resolution of the nonmonotonic case, if that is how the sequence becomes encoded. In the first case it would allow the creation of a string such as 43125; in the second as, for example, "second biggest, middle-sized, smallest, second smallest, biggest." Although only a minority of our eight adults reported converting the sequence to numbers, something along these lines could account for the relative ease with which an arbitrary list of ordinal sizes can be encoded by older participants. A contraindication to this argument, however is that fact that the adults showed a similar sensitivity in their RTs to the easy or difficulty of the sequence as defined by its sequential departure from monotonicity, that is, by the interval relations across the ordinal positions, and not by the relative difficulty of each item independently with regard to its distance from the nearest end-point. (Apart from the slowness in finding middle-sized as the start point for the harder sequence as shown by both age groups, this second measure of difficulty appears to produce quite different functions across the sequence. We have not presented these functions here, however, as a proper comparison of predictions arising from a symbol-based slot encoding in LTM would require further appropriate modeling and more experimentation with the nonmonotonic case.)

From the data we have to hand, we have to conclude that if there is any sense in which adults used some form of symbol encoding to remember the sequence from trial to trial, it would seem that this augmented memory for the series rather than replaced the way they actually executed the task. Either way, in exploring cognitive advance at least within this domain, we must caution against the proposition that new solutions represent a direct and private cognitive advance in the way that children (we believe) come to understand the necessary unitization of discrete sets by around the age of 7. That is to say, we do not see symbol-level knowledge based on numbers or semantic descriptions of ordinal values such as "third biggest," as a more abstract version of a core procedure in the sense of either the R-R account of Karmiloff-Smith or Piaget's concept of reflective abstraction. Both of these ideas conserve the fundamental properties of an earlier solution, in the form of re-representing a procedure in the former case (Karmiloff-Smith, 1992), or consciously reflecting on previous operations in the latter (see Campbell & Bickhard, 1986, p. 91). Although the private discoveries giving rise to series representation may be the foundation for understanding what numeric and ordinal codes refer to in the real world, the connection between these levels of understanding is a vast area of enquiry in its own right relating among other things to the nature of symbolic meaning and consciousness itself (Fodor, 1975; Harnad, 1990; Searle, 1980). Whatever symbolic resources may be imported into size sequencing, they are bound to be interwoven with diverse influences from school and the broader culture, some of which will be explicitly taught and not necessarily directly abstracted from private knowledge. Most of arithmetic teaching falls into this category in the sense that primary intuitions eventually have to be relinquished in favor of formal symbol manipulation. If we must speculate on how and why our adult participants succeeded on the nonmonotonic tasks whereas the 7-year-olds failed, we would say it is because the adults used a combination of an intuitive monotonic search together with the use of formal and perhaps linguistic symbols such as numbers and ordinal positions that helped in some way to store the sequence in LTM. What we would dispute is that such symbols are somehow abstracted representations of the search procedures, and we propose instead that they have a new and different role in preserving cognitive economy.

We would agree with Clark (2008) that once such symbols are externalized as shared, cultural artifacts, they too allow a radical reduction in the complexity of information processing that an agent must carry out to solve a sequencing problem. Describing such symbols as "tags" Clark noted, "once fluent in the use of tags, complex properties and relations in the perceptual array are, in effect, artificially reconstituted as simple inspectable wholes. The effect is to reduce the descriptive complexity of the scene" (p. 46). So, competence in a nonmonotonic task may represent developing expertise in the use of externally represented symbolic systems, using them to structure LTM in some way. Knowing when this competence emerges after the age of 7, however, an area for future research and further modeling.

On our interpretation then, the 7-year-old's expertise represents a natural end-state of sequential and ordinal size understanding which in and of itself never gets surpassed or replaced by a better understanding or better procedures for information management. On this scenario, by forcing the management of independent size codes in the random alternation phase of our color conditional task and, in the nonmonotonic tasks, we have strayed into a cognitively different domain. To solve these tasks without difficulty, the processes appear to be connected to seriation expertise but in a deeply complex way. Our surmise is that rapid solutions to these problems will depend on additional tools for memorizing the sequences such as cultural exposure to the use of symbols as place-markers for represented elements. But although this may make the task of remembering non-monotonic tasks easier, it may have little effect on the core sequential procedures required to actually execute them. The suggestion from our adult data is that monotonicity is the cognitive default solution for dealing with certain mathematical and logical problems for adults and children alike. In the next section, we show how this conclusion can be supported from other relevant research

The Prevalence of Monotonicity in Human Cognition

In this monograph, we have confined ourselves to insights that can derive exclusively from private experience with the physical world. These insights are not dependent on formal tuition, but on the functionality of the brain as an information reducing device.

It is a reasonable supposition, therefore, that such economic tools for information management will persist as default strategies into adolescence and adulthood even in domains that can, in principle, be resolved through learned systems of formal logic. Indeed much of the work on human adult reasoning since the 1970s has explored the many ways in which adult reasoners depart from logical rules (Johnson-Laird & Steedman, 1978; Kahneman, Slovic, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). Without pursuing the very many threads into nonlogical strategies for problem-solving used by human adults and adolescents in areas as diverse as conditional, spatial, and class-inclusion reasoning, we illustrate this point using two examples specifically relating to monotonicity. The first is that children's mathematical advance in the school years shows a strong dependency on spatio-temporal monotonicity even after the age of 7. The other is that exactly that sort of dependency tends to characterize how adults solve related logicomathematical problems.

Spatio-Temporal Procedures in Children in the Domain of Number

An arguably contradictory case regarding our suggested relationship between monotonic sequencing and the domain of number is the order-relevance principle in counting. Order irrelevance was considered by Gelman (1972) to be one of the basic principles of number understanding and it refers to the fact that the total cardinal value of a set is independent of the order in which the items are counted. As Colomé and Noël (2012) point out, however, the (spatial) order of a count is in fact highly relevant to understanding ordinal position. In testing this proposition with children aged between 3 and 5 years of age, they used a cartoon character counting to a target toy car queuing at a traffic light and the children were to judge the accuracy of the count. By 5 years of age, the children tended to correctly reject trials where the character took an unconventional spatial path for counting the set from the last car but also causing them to wrongly reject the count of the total number of cars queuing. Of relevance here is that a similar observation was made in a study Kamawar et al. (2010) using much older participants. Children aged from 5–10 years old took part in a game in which an animated frog named Hoppy counted squares arranged in rows on a computer screen. The child had to monitor Hoppy to see if he made a mistake. Using small numbers ranging from 3 to 13, Hoppy was either correct or incorrect, but sometimes he also counted in an unconventional manner instead of following a grid-based trajectory. Around 30% of children, including some as old as 10 or 11, categorized unconventional counts as incorrect and the authors conclude that "early adoption of very stringent and reliable counting procedures is an effective way for children to develop excellent skills" (p. 143).

Spatio-Temporal Procedures in Adults

A second example of the primacy of a monotonic solution, this time in a logical rather than numerical domain, comes from the realm of transitive reasoning. Consider the three-term series task described in Chapter I. When posed to adults linguistically, this task presents two premises and then poses an inferential question, for example, "Bill is shorter than Mike; Bill is taller than John. Who is the tallest of the three?"

Researchers using this three-term series problem have consistently identified considerable difficulties shown by adolescents and adults when the premises were presented non-monotonically or in a *heterotropic* order as per the example just given (Clark, 1969; Glick & Wapner, 1968; Hunter, 1957; Huttenlocher, 1968). There was a strong consensus that participants would convert disordered premises (e.g., $B < A$; $B > C$) into the sequence $A > B > C$ before answering the question. The idea that they did this by means of a spatial paralogical device was given strong support by Trabasso and colleagues when following up on the original Bryant and Trabasso study (Bryant & Trabasso, 1971). Using 6- and 9-year-olds and adults on a six-term training series, Trabasso and colleagues (Riley & Trabasso, 1974; Trabasso & Riley, 1975; Trabasso, Riley, & Wilson, 1975) found that children (and adults) consistently showed a "Symbolic Distance Effect," after Moyer (1973). That is, the RTs for answering test questions were faster than for the adjacent items on which they had been trained. This suggested that the premises were effectively seriated during training rather than combined using logical deduction. On the spatial model, it is thought that the participants read off the answer directly; the more discriminable the items on a mental linear scale the faster the answer.

In summary, we suggest that ordering monotonically is the default means by which human adults and older children process relational size information, because it has maximum predictability and highest possible redundancy in the informational sense. Although spawned by private discovery, we believe, it has obvious manifestations in the form of cultural devices such as rulers, thermometers, musical scales, and the notational systems of mathematics. Whatever happens after the age of 7, the sequential and ordinal skills that appear in children of this age provide the core psychological foundation without which such cultural devices could never have been invented.

Returning to the definitions of the logical properties of discrete sets with which we started this monograph—ordinality, cardinality, and unitization—we now have a clear take-home message. These three properties do not seem to be equal partners in an operational coalition. If there is a primary element it is unitization, and even this we see as owing its emergence to something that Piaget tended to treat as a symptom rather than a cause of operational development—principled monotonic sequencing. We have argued that it is in the very activity of ordering that the economy of individual-item identification becomes available. This activity makes ordinal identification a secondary—and to some extent an optional—skill. It is not always necessary to compute the ordinal position of each item in a set unless it is demanded by the task in hand, and it is possible to be capable of expert seriation without this computation playing any part. Task requirements notwithstanding, there is, we believe, a single foundational ability behind the way children can start to behave logically with discrete sets of objects differing in size, and we would claim that this ability lies in the act of sequencing them.

This message alone, however, lacks psychological depth or insight into that foundational understanding. It is here that we see the computational modeling as providing depth and detail by capturing transitional processes that can never be evident simply in the progression from failure to success, and by reifying the changes in LTM that we can only infer from the behavior alone. With the benefit of a task analysis based on simulation, furthermore, the tasks themselves gain explanatory power retrospectively, and for this reason we could envisage a role for our serial and ordinal tasks in a neuropsychological context. Highly relevant to the co-operative functions of the PFC, they would be simple to administer inside a scanner and could throw considerable light on moment-to-moment brain functioning during sequencing. In the meantime, we return to their importance of these tasks as empirical indicators of key long-term shifts in the organization of the brain across development. Gaining an even better understanding of this and other discontinuities in development remains a task for psychologists and cognitive modelers alike in the future alignment of the behavioral with the neuropsychological sciences.

Acknowledgments

We would like to thank the staff and children of George Heriot's (Junior) School, Edinburgh for their enthusiastic co-operation and support. We would also like to thank Jeff Johnson and Anthony Lucas-Smith from the Open University for their much-valued guidance with the computational model research presented in this paper, for which Theodore Zamenopoulos and Mark Steedman provided many insightful comments. Milton Garcia Borroto provided valued guidance with mathematical notation and SPSS, Marcin Skibicki with R, and Ben Barber with the proof in Endnote 5. The empirical work reported in this paper was supported by a grant from the ESRC of Great Britain (R000232488).

Notes

1. Bayesian cognitive model exploration as series of hand-on tutorials can be carried out via Lee (2013).

2. Van Geert (1994) provides hands-on tutorials for building dynamical systems of development, including a model of Kurt Fisher's theory of skill development.

3. Further exploration of Young's seriation model can be carried out via Scott and Nicolson (1991) who provide guidelines for a Prolog implementation. Young's model of subtraction competence in the child, alongside many other models of cognition, are available via Yule et al. (2013). As noted above, the production systems formalism has been adopted into contemporary cognitive architectures ACT-R and SOAR) (Miłkowski, 2013), and incorporate various mechanisms for representing learning and plasticity. Anderson (2007) provides theoretical and practical guidance for ACT-R implementations.

4. Exploration of the main connectionist models of cognitive development can be carried out via hands-on labs in Plunkett and Elman (1997). Quinlan (2003) provides a thorough overview of connectionist models of cognitive development, with a biological as well as psychological focus.

5. It can be proven mathematically that given enough pairwise comparisons from a set, and a way of persisting the results of these comparisons, that a rank order will always form.Assume a set $S = \{e_1, e_2, e_3 \ldots e_n\}$, and let $X_i(t)$ be the weight of each of e_i at time t, assuming a set of discrete time steps for t. Assume $X_i(t) \subset \mathbb{Z}$. At each discrete time step t, a random pair of set elements are selected according to the relation $1 \leqslant i < j \leqslant n$. At this point, 1 is added to the weight of set element j and 1 is subtracted from the weight of set element i. Considering the i set element only, as the j set element procedure is equivalent:

$$X_i(t+1) = \begin{cases} X_i(t), & \text{if } i \text{ was not chosen at time } t+1, \\ X_i(t)+1, & \text{if } i \text{ and a smaller } j \text{ were chosen at time } t+1, \\ X_i(t)-1, & \text{if } i \text{ and a larger } j \text{ were chosen at time } t+1. \end{cases}$$

Assume that each $X_i(t)$ is a random walk on the integer number line $\mathbb{Z}$, with different probabilities of moving up, moving down or staying the same. The expectation of $X_i(t)$ takes the form $c(((i-1)/(n-1)) - (1/2))t$ for some constant c, and so the weights $X_i(t)$ are expected to be in the correct order separated by numeric differences of an order t. However, the deviation from the expectation is around $\sqrt{t}$ with a high probability, according to Chernoff's inequality. Typically, the random fluctuations around the mean are not enough to break the ordering of the set, and so as t gets larger, the set elements will be highly likely to be ordered according to the relation $\{e_1 < e_2 < e_3 < \cdots < e_n\}$. This proof extends to the *Beta* ranking case, as each $X_i(t)$ is equivalent to the mean value $\alpha/(\alpha+\beta)$ for a particular *Beta* probability distribution, assuming the *Beta* distributions are updated in a way that allows a ranking of means to be established.

6. As noted previously, evolution appears to have discovered a behavioral program for economically scanning stimulus sets to facilitate rational choice. As a result of this pairwise procedure, the number of comparisons reduces from $n!$ to $n(n+1)/2$ (i.e., the triangular number of pairwise comparisons).

7. *Beta* distributions are defined by two parameters (α and β) that are real numbers determining the probability of a stimulus ranking along the unit (0–1) interval. *Beta* probability distributions were selected due to this simple, dual-parameter definition, their representing

more probability mass toward the end-points of the unit interval, and their conjugate prior property (prior and posterior distributions of the same *Beta* distribution format). These properties allow a representation of end-point biasing, a binary classification psychological constraint and a mathematically simple Bayesian updating routine. In order to represent a global ranking of, for example, $\{A > B > C\}$ along the *Beta* distribution unit interval, these counts must increase and/or decrease monotonically across the set. In the scenario where we have $\{A_\alpha, B_\alpha, C_\alpha\} = \left\{\frac{3}{3}, \frac{2}{3}, \frac{1}{3}\right\}$ and $\{A_\beta, B_\beta, C_\beta\} = \left\{\frac{1}{3}, \frac{2}{3}, \frac{3}{3}\right\}$, we have a global ranking of $\{A > B > C\}$ along the unit interval, where the mean values are: $A = 0.75$, $B = 0.5$, and $C = 0.25$. This accumulation of counts is equivalent to a Bayesian inference calculation in which prior and likelihood probabilities are combined to generate posterior probabilities.

8. A function used to convert a set of real values, vector $Score_n$, into action probabilities,

$$softmax(Score_n) = \frac{e^{Score_n/\tau}}{\sum_{i=1}^{n} e^{Score_i/\tau}}$$

where τ is a temperature parameter of value 1. Variability in *softmax* can be controlled via the temperature factor in this equation. This function is widely used in machine learning research to facilitate action selection. For high temperatures, all vector values have nearly the same probability, and for a low temperature, the probability of the action with the vector value tends to 1.

9. The term $Beta.\ \mu_{ccandidate}/Beta.\ \sigma^2{}_{candidate}$ is the precision of the candidate stimulus, where

$$Beta \cdot \mu_{ccandidate} = \frac{Beta \cdot \alpha}{Beta \cdot \alpha + Beta \cdot \beta} \text{ and } Beta \cdot \sigma^2_{candidate} = \frac{Beta \cdot \alpha \times Beta \cdot \beta}{(Beta \cdot \alpha + Beta \cdot \beta)^2 (Beta \cdot \alpha + Beta \cdot \beta + 1)},$$

and $Inhibition_{candidate}$ is a real valued inhibition level.

References

Achenbach, T. M., & Weisz, J. R. (1975). A longitudinal study of developmental synchrony between conceptual identity, seriation, and transitivity of color, number, and length. *Child Development*, **46**(4), 840–848. https://doi.org/10.2307/1128388

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195324259.001.0001

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, **54**, 695–701. https://doi.org/10.2307/1130057

Ashby, F. G., Ell, S. W., Valentin, V. V., & Casale, M. B. (2005). FROST: A distributed neurocomputational model of working memory maintenance. *Journal of Cognitive Neuroscience*, **17**(11), 1728–1743. https://doi.org/10.1162/089892905774589271

Asmuth, J., Morson, E. M., & Rips, L. J. (2018). Children's understanding of the natural numbers' structure. *Cognitive Science*, **42**(6), 1945–1973. https://doi.org/10.1111/cogs.12615

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation: II*, 89–195. https://doi.org/10.1016/S0079-7421%2808%2960422-3

Baddeley, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, **10**(1–3), 17–23. https://doi.org/10.1016/0010-0277%2881%2990020-2

Baddeley, A. D. (1984). The fractionation of human memory. *Psychological Medicine*, **14**(2), 259–264. https://doi.org/10.1017/S0033291700003536

Baddeley, A. D. (1992). Working memory. *Science*, **255**(5044), 556–559. https://doi.org/10.1126/science.1736359

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**(4), 723–742. https://doi.org/10.1017/S0140525X97001611

Barendse, E. M., Hendriks, M. P., Jansen, J. F., Backes, W. H., Hofman, P. A., Thoonen, G., et al. (2013). Working memory deficits in high-functioning adolescents with autism spectrum disorders: Neuropsychological and neuroimaging correlates. *Journal of Neurodevelopmental Disorders*, **5**(1), 1–11. https://doi.org/10.1186/1866-1955-5-14

Barnard, P. J. (1999). Interacting cognitive subsystems: Modeling working memory phenomena within a multiprocessor architecture. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance Development of active maintenance and executive control* (pp. 298–339). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Barth, H., Baron, A., Spelke, E., & Carey, S. (2009). Children's multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology*, **103**(4), 441–454. https://doi.org/10.1016/j.jecp.2009.01.014

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, **41**(4), 579–597. https://doi.org/10.1037/0012-1649.41.4.579

Baylor, G. W., Gascon, J., Lemoyne, G., & Pothier, N. (1973). An information processing model of some seriation tasks. *Canadian Psychologist/Psychologie Canadienne*, **14**(2), 167–196. https://doi.org/10.1037/h0082217

Baylor, G. W., & Lemoyne, G. (1975). Experiments in seriation with children: Towards an information processing explanation of the horizontal decalage. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, **7**(1), 4–29. https://doi.org/10.1037/h0081893

Beilin, H. (1999). Is Piaget truly an epistemologist? Commentary on Leslie Smith's "Epistemological principles for developmental psychology in Frege and Piaget". *New Ideas in Psychology*, **17**(2), 119–121. https://doi.org/10.1016/s0732-118x(99)00013-6

Berteletti, I., Lucangeli, D., & Zorzi, M. (2012). Representation of numerical and non-numerical order in children. *Cognition*, **124**(3), 304–313. https://doi.org/10.1016/j.cognition.2012.05.015

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, **81**(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bidell, T. R., & Fischer, K. W. (1992). Beyond the stage debate: Action, structure, and variability in Piagetian theory and research. *Philosophica*, **54**(2), 43–87.

Bingham-Newman, A. M., & Hooper, F. H. (1974). Classification and seriation instruction and logical task performance in the preschool. *American Educational Research Journal*, **11**(4), 379–393. https://doi.org/10.2307/1162791

Blevins-Knabe, B. (1987a). Development of the ability to insert into a series. *The Journal of Genetic Psychology*, **148**(4), 427–441. https://doi.org/10.1080/00221325.1987.10532482

Blevins-Knabe, B. (1987b). Remembering and constructing an order. *The Journal of Genetic Psychology*, **148**(4), 453–468. https://doi.org/10.1080/00221325.1987.10532484

M. Boden (Ed.). (1996). *The philosophy of artificial life*. London: Oxford University Press.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, **76**(Part B), 198–211. https://doi.org/10.1016/j.jmp.2015.11.003

Borghi, A. M., & Scorolli, C. (2009). Language comprehension and dominant hand motion simulation. *Human Movement Science*, **28**(1), 12–27. https://doi.org/10.1016/j.humov.2008.07.002

Borkowski, J. G., & Burke, J. E. (1996). Theories, models, and measurements of executive functioning: An information processing perspective. In G. R. Lyon, & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 235–261). Baltimore, MD: Paul H Brookes Publishing.

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2012). Inhibitory control in number-conservation and class-inclusion tasks: A neo-Piagetian inter-task priming study. *Cognitive Development*, **27**(3), 283–298. https://doi.org/10.1016/j.cogdev.2012.02.004

Bouwmeester, S., Vermunt, J. K., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. *Developmental Review*, **27**(1), 41–74. https://doi.org/10.1016/j.dr.2006.08.001

Bower, T. G. R., Broughton, J., & Moore, M. K. (1971). Development of the object concept as manifested in changes in the tracking behavior of infants between 7 and 20 weeks of age. *Journal of Experimental Child Psychology*, **11**(2), 182–193. https://doi.org/10.1016/0022-0965(71)90074-9

Braine, M. D. S. (1964). Development of a grasp of transitivity of length: A reply to Smedslund. *Child Development*, **35**, 799–810. https://doi.org/10.2307/1126505

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences*, **1**(2), 173–182. https://doi.org/10.1017/S0140525X00073842

Braitenberg, V. (1984). *Vehicles*. Cambridge, MA: MIT Press.

Breslow, L. (1981). Reevaluation of the literature on the development of transitive inferences. *Psychological Bulletin*, **89**(2), 325–351. https://doi.org/10.1037//0033-2909.89.2.325

Bringuier, J.-C. (1980). *Conversations with Jean Piaget*. Chicago, IL: University of Chicago Press.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, **19**(1), 1–15. https://doi.org/10.1037/h0044160

Bryant, P. E. (1972). The understanding of invariance by very young children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, **26**(1), 78–96. https://doi.org/10.1037/h0082418

Bryant, P. E. (1974). *Perception and understanding in young children: An experimental approach*. New York, NY: Basic Books.

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, **232**, 456–458. https://doi.org/10.1038/232456a0

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, **8**(9), 426–433. https://doi.org/10.1016/j.tics.2004.07.003

Campbell, R. L., & Bickhard, M. H. (1986). Knowing levels and developmental stages. *Contributions to Human Development*, **16**, 146. https://doi.org/10.1017/s0140525x00028697

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York, NY: Academic Press.

Case, R., Okamoto, Y., Henderson, B., & McKeough, A. (1993). Individual variability and consistency in cognitive development: New evidence for the existence of central conceptual structures. In R. Case, & W. Edelstein (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways* (pp. 1–100). Basel: Karger.

Chalmers, M., & McGonigle, B. O. (1984). Are children any more logical than monkeys on the five-term series problem? *Journal of Experimental Child Psychology*, **37**(2), 355–377.

Chalmers, M., & McGonigle, B. (1997). Capturing dynamic structuralism in the laboratory. In L. Smith, J. Dockrell, & P. Tomlinson (Eds.), *Piaget, Vygotsky and beyond* (pp. 183–200). London: Routledge. https://doi.org/10.4324/9780203360415

Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. Cambridge: Cambridge University Press.

Chapman, M., & Lindenberger, U. (1989). Concrete operations and attentional capacity. *Journal of Experimental Child Psychology*, **47**(2), 236–258. https://doi.org/10.1016/0022-0965%2889%2990031-3

Chapman, M., & McBride, M. L. (1992). Beyond competence and performance: Children's class inclusion strategies, superordinate class cues, and verbal justifications. *Developmental Psychology*, **28**(2), 319–327. https://doi.org/10.1037/0012-1649.28.2.319

Chen, Z., & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development*, **65**(2), 1–96. https://doi.org/10.1111/1540-5834.00084

Chu, F. W., vanMarle, K., Rouder, J., & Geary, D. C. (2018). Children's early understanding of number predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology*, **169**, 73–92. https://doi.org/10.1016/j.jecp.2017.12.010

Clark, A. (2008). *Supersizing the mind: Embodiment, action and cognitive extension*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, H. H. (1969). Influence of language on solving three-term series problems. *Journal of Experimental Psychology*, **82**, 205–215.

Clark, H. H. (1970). The primitive nature of childrens' relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 269–278). New York, NY: Wiley.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.

Cohen, D. (1983). *Piaget: Critique and reassessment*. London: Croom Helm.

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, **50**(6), 1640–1652. https://doi.org/10.1037/a0035901

Colomé, À., & Noël, M. -P. (2012). One first? Acquisition of the cardinal and ordinal uses of numbers in preschoolers. *Journal of Experimental Child Psychology*, **113**(2), 233–247. https://doi.org/10.1016/j.jecp.2012.03.005

Cooper, R. P. (2016). Executive functions and the generation of "random" sequential responses: A computational account. *Journal of Mathematical Psychology*, **73**, 153–168. https://doi.org/10.1016/j.jmp.2016.06.002

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, **17**(1), 11–22. https://doi.org/10.1111/desc.12108

Cowan, R., & Daniels, H. (1989). Children's use of counting and guidelines in judging relative number. *British Journal of Educational Psychology*, **59**(2), 200–210. https://doi.org/10.1111/j.2044-8279.1989.tb03091.x

Coxford, A. F. (1964). The effects of instruction on the stage placement of children in Piaget's seriation experiments. *The Arithmetic Teacher*, **11**(1), 4–9.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Special Issue: Advances in Developmental Cognitive Neuroscience*, **44**(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., & Cangelosi, A. (2014). Making fingers and words count in a cognitive robot. *Frontiers in Behavioral Neuroscience*, **8**, 13. https://doi.org/10.3389/fnbeh.2014.00013

Dehaene, S. (1997). *The number sense*. New York, NY: Oxford University Press. https://doi.org/10.1017/s0790966700006248

Delius, J, & Siemann, M. (1998). Transitive responding in animals and humans: exaptation rather than adaptation. *Behavioural Processes*, **42**, 107–137. https://doi.org/10.1016/s0376-6357(97)00072-7

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, **12**(1), 45–75. https://doi.org/10.1016/0273-2297%2892%2990003-K

Desrochers, S. (2008). From Piaget to specific Genevan developmental models. *Child Development Perspectives*, **2**(1), 7–12. https://doi.org/10.1111/j.1750-8606.2008.00034.x

Di Paolo, E. A., Barandiaran, X. E., Beaton, M., & Buhrmann, T. (2014). Learning to perceive in the sensorimotor approach: Piaget's theory of equilibration interpreted dynamically. *Frontiers in Human Neuroscience*, **8**, 551. https://doi.org/10.3389/fnhum.2014.00551

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, **64**, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dockrell, J., Campbell, R., & Neilson, I. (1980). Conservation accidents revisited. *International Journal of Behavioral Development*, **3**(4), 423–439. https://doi.org/10.1177/016502548000300405

Dodwell, P. C. (1960). Children's understanding of number and related concepts. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, **14**(3), 191–205.

Dodwell, P. C. (1968). Children's understanding of spatial concepts. In I. E. Sigel, & F. H. Hooper (Eds.), *Logical thinking in children* (pp. 118–240). New York, NY: Holt, Rinehart and Winston.

Donaldson, M. (1978). *Children's minds*. Glasgow: Fontana/Collins.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. Cambridge, MA: MIT Press.

Eames, D., Shorrocks, D., & Tomlinson, P. (1990). Naughty animals or naughty experimenters? Conservation accidents revisited with video-stimulated commentary. *British Journal of Developmental Psychology*, **8**(1), 25–37. https://doi.org/10.1111/j.2044-835X.1990.tb00819.x

Elkind, D. (1968). Discrimination, seriation and numeration of size and dimensional differences in young children. Piaget replication study VI. In I. E. Sigel, & F. H. Hooper (Eds.), *Logical thinking in children* (pp. 56–75). New York, NY: Holt, Rinehart and Winston.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, **39**, e229.

Fischer, F. E., & Beckey, R. D. (1990). Beginning kindergarteners' perception of number. *Perceptual and Motor Skills*, **70**(2), 419–425. https://doi.org/10.2466/PMS.70.2.419-425

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

Fodor, J. A. (1975). *The language of thought*. New York, NY: Thomas Y. Crowell.

Friedman, J., & Pasnak, R. (1973). Attainment of classification and seriation concepts by blind and sighted children. *Education of the Visually Handicapped*, **5**(2), 55–62.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, **60**(5), 1158–1171. https://doi.org/10.2307/1130790

Fürth, H. G. (1969). *Piaget and knowledge: Theoretical foundations*. Englewood Cliffs, NJ: Prentice Hall.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag Publishing.

Fuson, K. C., Secada, W. G., & Hall, J. W. (1983). Matching, counting, and conservation of numerical equivalence. *Child Development*, **54**(1), 91–97. https://doi.org/10.2307/1129865

Geary, D. C., vanMarle, K., Chu, F. W., Hoard, M. K., & Nugent, L. (2018). Predicting age of becoming a cardinal principle knower. *Journal of Educational Psychology, No-Specified*. https://doi.org/10.1037/edu0000277

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number-system knowledge. *Psychological Science*, **29**(2), 191–205. https://doi.org/10.1177/0956797617729817

Gelman, R. (1972). The nature and development of early number concepts. In H. W. Reese (Ed.), *Advances in child development and behavior* (**7**, pp. 115–167). New York, NY: Academic Press.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, **13**, 74–82. https://doi.org/10.1016/j.tics.2008.11.006

Gibson, E. P., Gibson, J. J., Pick, A. D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, **55**(6), 897–906.

Glick, J., & Wapner, S. (1968). Development of transitivity: Some findings and problems of analysis. *Child Development*, **39**(2), 621–638. https://doi.org/10.2307/1126972

Goldberg, E. (2017). *Executive functions in health and disease*. London: Academic Press.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, **65**(2-3), 231–262. https://doi.org/10.1016/S0010-0277%2897%2900047-4

Gollin, E. S., Moody, M., & Schadler, M. (1974). Relational learning of a size concept. *Developmental Psychology*, **10**(1), 101–107. https://doi.org/10.1037/h0035542

Greenaway, R., & Plaisted, K. (2005). Top-down attentional modulation in autistic spectrum disorders is stimulus-specific. *Psychological Science*, **16**(12), 987–994. https://doi.org/10.1111/j.1467-9280.2005.01648.x

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, **16**, 94–143. https://doi.org/10.1016/j.jecp.2014.09.008

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, **130**, 35–55. https://doi.org/10.1016/j.jecp.2014.09.008

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, **455**, 665–668. https://doi.org/10.1038/nature07246

Halford, G. S. (1989). Reflections on 25 years of Piagetian cognitive developmental psychology, 1963–1988. *Human Development*, **32**, 325–357. https://doi.org/10.1159/000276484

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., Andrews, G., Wilson, W. H., & Phillips, S. (2012). Computational models of relational processes in cognitive development. *Cognitive Development*, **27**(4), 481–499. https://doi.org/10.1016/j.cogdev.2012.08.003

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, **21**(06), 803–831. https://doi.org/10.1017/S0140525X98001769

Hanrahan, J., Yelin, M., & Rapagna, S. (1987). The effect of test materials and the order of presentation of the materials on young children's understanding of conservation of number. *British Journal of Educational Psychology*, **57**, 407–411. https://doi.org/10.1111/j.2044-8279.1987.tb00867.x

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, **42**, 335–346. https://doi.org/10.1016/0167-2789(90)90087-6-6

Harris, M. R., & McGonigle, B. O. (1994). A model of transitive choice. *The Quarterly Journal of Experimental Psychology*, **47**(3), 319–348.

Harris, P. L., & Nunez, M. (1996). Understanding of permission rules by preschool children. *Child Development*, **67**(4), 1572–1591. https://doi.org/10.2307/1131719

Hatwell, Y. (1985). *Piagetian reasoning and the blind*. New York: American Foundation for the Blind.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, **28**(4), 603–621. https://doi.org/10.1080/14640747608400587

O. Holland, & D. McFarland (Eds.). (2001). *Artificial ethology*. Oxford: Oxford University Press.

Houdé, O., & Borst, G. (2014). Measuring inhibitory control in children and adults: Brain imaging and mental chronometry. *Frontiers in Psychology*, **5**, 616. https://doi.org/10.3389/fpsyg.2014.00616

Houdé, O., & Borst, G. (2015). Evidence for an inhibitory-control theory of the reasoning brain. *Frontiers in Human Neuroscience*, **9**. https://doi.org/10.3389/fnhum.2015.00148

Houdé, O., & Guichart, E. (2001). Negative priming effect after inhibition of number/length interference in a Piaget-like task. *Developmental Science*, **4**(1), 119–123. https://doi.org/10.1111/1467-7687.00156

Hunter, I. M. L. (1957). The solving of three-term series problems. *British Journal of Psychology*, **48**, 286–298. https://doi.org/10.1111/j.2044-8295.1957.tb00627.x

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, **75**(6), 550–560. https://doi.org/10.1037/h0026748

Hyde, D. C., Simon, C. E., Berteletti, I., & Mou, Y. (2017). The relationship between non-verbal systems of number and counting development: A neural signatures approach. *Developmental Science*, **20**(6), 13. https://doi.org/10.1111/desc.12464

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Oxford: Routledge.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation (E.A Lunzer & D. Papert, Trans.)*. London: Routledge and Kegan Paul. (Original work published 1959).

Jensen, G., Muñoz, F., Alkan, Y., Ferrera, V. P., & Terrace, H. S. (2015). Implicit value updating explains transitive inference performance: The betasort model. *PLoS Computational Biology*, **11**(9) e1004523. https://doi.org/10.1371/journal.pcbi.1004523

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, **10**, 64–99.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development*, **79**(6), 1606–1624. https://doi.org/10.1111/j.1467-8624.2008.01211.x

Kahn, J. V., & Garrison, M. (1973). The effects of age, stimuli and order of presentation of the stimuli on conservation of number. *British Journal of Educational Psychology*, **43**(3), 249–256. https://doi.org/10.1111/j.2044-8279.1973.tb02265.x

D. Kahneman, S. P. Slovic, P. Slovic, & A. Tversky (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University press.

Kallio, K. D. (1982). Developmental change on a five-term transitive inference. *Journal of Experimental Child Psychology*, **33**(1), 142–164. https://doi.org/10.1016/0022-0965(82)90011-X

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman assessment battery for children*. Circle Pines, MN: American Guidelines Service.

Kamawar, D., LeFevre, J.-A., Bisanz, J., Fast, L., Skwarchuk, S.-L., Smith-Chant, B., & Penner-Wilger, M. (2010). Knowledge of counting principles: How relevant is order relevance? *Journal of Experimental Child Psychology*, **105**(1-2), 138–145. https://doi.org/10.1016/j.jecp.2009.08.004

Kendler, T. S. (1975). *Levels of cognitive development*. Mahwah, NJ: Lawrence Erlbaum.

Kercood, S., Grskovic, J. A., Banda, D., & Begeske, J. (2014). Working memory and autism: A review of literature. *Research in Autism Spectrum Disorders*, **8**(10), 1316–1332. https://doi.org/10.1016/j.rasd.2014.06.011

Kidd, J. K., Carlson, A. G., Gadzichowski, K. M., Boyer, C. E., Gallington, D. A., & Pasnak, R. (2013). Effects of patterning instruction on the academic achievement of 1st-grade children. *Journal of Research in Childhood Education*, **27**(2), 224–238. https://doi.org/10.1080/02568543.2013.766664

Kingma, J. (1983a). The development of seriation, conservation and multiple classification: A longitudinal study. *Genetic Psychology Monographs*, **108**, 43–67.

Kingma, J. (1983b). Seriation, correspondence, and transitivity. *Journal of Educational Psychology*, **75**(5), 763–771. https://doi.org/10.1037/0022-0663.75.5.763

Kingma, J. (1984a). Task sensitivity and the sequence of development in seriation, ordinal correspondence, and cardination. *Genetic Psychology Monographs*, **110**(2), 181–205. https://doi.org/10.1080/00221325.1984.9923433

Kingma, J. (1984b). Traditional intelligence, Piagetian tasks, and initial arithmetic in kindergarten and primary school grade one. *The Journal of Genetic Psychology*, **145**, 49–60. https://doi.org/10.1080/00221325.1984.10532250

Kingma, J. (1984c). Traditional intelligence, Piagetian Tasks, and Initial Arithemtic in kindergarten and primary-school grade-one. *Journal of Genetic Psychology*, **145**, 49–60.

Kingma, J. (1986). The range of seriation training effects in young kindergarten children. *Contemporary Educational Psychology*, **11**, 276–289. https://doi.org/10.1016/0361-476X(86)90023-8

Kingma, J. (1987). Training of seriation in young kindergarteners. *The Journal of Genetic Psychology*, **148**(2), 167–181. https://doi.org/10.1080/00221325.1987.9914546

Kingma, J., & Reuvekamp, J. (1984). The construction of a developmental scale for seriation. *Educational and Psychological Measurement*, **44**(1), 1–23. https://doi.org/10.1177/0013164484441001

Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F., & Salthouse, T. A. (1999). Models of working memory: Eight questions and some general issues. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 412–441). Cambridge: Cambridge University Press.

Klahr, D. (1992). Information processing approaches to cognitive development. In M. H. Bornstein, & M. E. Lamb (Eds.), *Developmental Psychology: An Advanced Textbook* (pp. 273–335). Hillsdale, NJ: Lawrence Erlbaum.

Kohler, W. (1925). *The mentality of Apes*. London: Routledge and Kegan Paul.

Kolb, D. A. (1983). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Koller, D., Friedman, N., & Bach, F. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press.

Koslowski, B. (1980). Quantitative and qualitative changes in the development of seriation. *Merrill-Palmer Quarterly*, **26**(4), 391–405.

Kuhn, D. (1992). Cognitive development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook*. Hillsdale, NJ: Lawrence Erlbaum.

Lakoff, G., & Nunez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Lawrenson, W., & Bryant, P. E. (1972). Absolute and relative codes in young children. *Journal of Child Psychology and Psychiatry*, **13**, 25–35. https://doi.org/10.1111/j.1469-7610.1972.tb01115.x

Lawson, G., Baron, J., & Siegel, L. (1974). The role of number and length cues in children's quantitative judgments. *Child Development*, **45**(3), 731–736. https://doi.org/10.2307/1127839

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, **105**(2), 395–438. https://doi.org/10.1016/j.cognition.2006.10.005

Lebron-Rodriguez, D. E., & Pasnak, R. (1977). Induction of intellectual gains in blind children. *Journal of Experimental Child Psychology*, **24**(3), 505–515.

Lee, M. D. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

LeFevre, J. A. (2016). Numerical cognition: Adding it up. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, **70**(1), 3–11.

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, **70**(1), 12–23. https://doi.org/10.1037/cep0000070

Leiser, D., & Gillierion, C. (1990). *Cognitive science and genetic epistemology*. New York: Plenum.

Liben, L. S. (1987). Information processing and Piagetian theory: Conflict and congruence. In L. S. Liben (Ed.), *Development and learning: Conflict or congruence?* (pp. 237–252). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Liben, L. S. (2008a). Continuities and discontinuities in children and scholarship. *Child Development*, **79**(6), 1600–1605. https://doi.org/10.1111/j.1467-8624.2008.01210.x

Liben, L. S. (2008b). Reflections on child development: The journal and the field. *Child Development*, **79**(6), 1597–1599. https://doi.org/10.1111/j.1467-8624.2008.01209.x

Lillard, A., & Else-Quest, N. (2006). THE EARLY YEARS: Evaluating montessori education. *Science*, **313**(5795), 1893–1894. https://doi.org/10.1126/science.1132362

Lister, C., Leach, C., Ballinger, S., & Simpson, L. (1996). Extent of similarity in concept development for visually impaired and sighted children. *Early Child Development and Care*, **117**, 21–28. https://doi.org/10.1080/0300443961170102

Little, A. (1972). A longitudinal study of cognitive development in young children. *Child Development*, **43**(3), 1024–1034. https://doi.org/10.2307/1127652

Lourenço, O., & Machado, A. (1996). In defense of Piaget's theory: A reply to 10 common criticisms. *Psychological Review*, **103**(1), 143–164. https://doi.org/10.1037//0033-295x.103.1.143

Lourenço, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, **21**(6), 873–881. https://doi.org/10.1177/0956797610370158.

Lourenço, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In S., Dehaene & E., Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 225–244). San Diego, CA: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-385948-8.00015-3

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). New York, NY: Cambridge University Press.

Mareschal, D., & Shultz, T. R. (1999). Development of children's seriation: A connectionist approach. *Connection Science*, **11**(2), 149–186. https://doi.org/10.1080/095400999116322

Mareschal, D., & Thomas, M. S. (2007). Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation*, **11**(2), 137–150. https://doi.org/10.1109/TEVC.2006.890232

Marshall, P. J. (2016). Embodiment and human development. *Child Development Perspectives*, **10**(4), 245–250. https://doi.org/10.1111/cdep.12190

May, R. B., & MacPherson, D. F. (1971). Size discrimination in children facilitated by changes in task difficulty. *Journal of Comparative and Physiological Psychology*, **75**(3), 453–458. https://doi.org/10.1037/h0030959

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, **1**, 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

McEvoy, J., & Mona O'Moore, A. (1991). Number conservation: A fair assessment of numerical understanding? *The Irish Journal of Psychology*, **12**(3), 325–337.

McFarland, D., & Bösser, T. (1993). *Intelligent behavior in animals and robots*. Cambridge, MA: MIT Press.

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, **3**(4), 341–350. https://doi.org/10.1016/0010-0277%2874%2990003-1

McGarrigle, J., Grieve, R., & Hughes, M. (1978). Interpreting inclusion: A contribution to the study of the child's cognitive and linguistic development. *Journal of Experimental Child Psychology*, **26**(3), 528–550. https://doi.org/10.1016/0022-0965%2878%2990131-5

McGonigle, B. O., & Chalmers, M. (1977). Are monkeys logical? *Nature*, **267**, 694–696.

McGonigle, B. O., & Chalmers, M. (1998). Rationality as optimised cognitive self-regulation. In M. Oaksford, & N. Chater (Eds.), *Rational models of cognition* (pp. 501–534). Oxford, UK: Oxford University.

McGonigle, B. O., & Chalmers, M. (2001). Spatial representation as cause and effect: Circular causality comes to cognition. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 247–277). Cambridge, MA: MIT Press.

McGonigle, B. O., & Chalmers, M. (2002). The growth of cognitive structure in monkeys and men. In S. B. Fountain, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behaviour: Behavioral, biological and computational perspectives* (pp. 269–314). Boston, MA: Kluwer Academic.

McGonigle, B. O., & Chalmers, M. (2006). Ordering and executive functioning as a window on the evolution and development of cognitive systems. *International Journal of Comparative Psychology*, **19**(2), 241–267.

McGonigle, B. O., & Jones, B. T. (1978). Levels of stimulus processing by the squirrel monkey: Relative and absolute judgements compared. *Perception*, **7**(6), 635–659. https://doi.org/10.1068/p070635

McGonigle-Chalmers, M., Bodner, K., Fox-Pitt, A., & Nicholson, L. (2008). Size sequencing as a window on executive control in children with autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders*, **38**(7), 1382–1390.

McKenna, R., Rushe, T., & Woodcock, K. A. (2017). Informing the structure of executive function in children: A meta-analysis of functional neuroimaging data. *Frontiers in Human Neuroscience*, **11**, 154. https://doi.org/10.3389/fnhum.2017.00154

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, **23**(8-9), 1125–1134. https://doi.org/10.1016/j.neunet.2010.08.010

Miłkowski, M. (2013). *Explaining the computational mind*. Cambridge, MA: MIT Press.

Minsky, M. (1988). A framework for representing knowledge. In A. Collins, & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 156–189). San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

A. Miyake, & P. Shah (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Moriguchi, Y., & Hiraki, K. (2011). Longitudinal development of prefrontal function during early childhood. *Developmental Cognitive Neuroscience*, **1**(2), 153–162. https://doi.org/10.1016/j.dcn.2010.12.004

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, **13**, 180–184. https://doi.org/10.3758/BF03214124

Moore, G. W. (1979). Transitive inferences within seriation problems assessed by explanations, judgments and strategies. *Child Development*, **50**(4), 1164–1172.

Mullane, J. C., & Corkum, P. V. (2007). The relationship between working memory, inhibition, and performance on the Wisconsin Card Sorting Test in children with and without ADHD. *Journal of Psychoeducational Assessment*, **25**(3), 211–221. https://doi.org/10.1177/0734282906297627

Neapolitan, D. M. (1991). *A micro-analysis of seriation skills* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh.

Neilson, I., Dockrell, J., & McKechnie, J. (1983). Justifying conservation: A reply to McGarrigle and Donaldson. *Cognition*, **15**(1-3), 277–291. https://doi.org/10.1016/0010-0277%2883%2990047-1

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Odic, D., Lisboa, J. V., Eisinger, R., Olivera, M. G., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychologica*, **163**, 17–26. https://doi.org/10.1016/j.actpsy

O'Hearn, K., Asato, M., Ordaz, S., & Luna, B. (2008). Neurodevelopment and executive function in autism. *Development and Psychopathology*, **20**, 1103–1132.

Okamoto, Y., & Case, R. (1996). The role of central conceptual structures in the development of children's thought: II. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development*, **61**(1-2), 27–58. https://doi.org/10.1111/j.1540-5834.1996.tb00536.x

Ozonoff, S. (1995). Reliability and validity of the Wisconsin-card sorting test in studies of autism. *Neuropsychology*, **9**, 491–500.

Page, R., Izquierdo, E., Saal, A., Codnia, J., & El Hasi, C. (2004). A response time model for judging order relationship between two symbolic stimuli. *Perception & Psychophysics*, **66**(2), 196–207. https://doi.org/10.3758/BF03194872

Papalia, D. E., & Hooper, F. H. (1971). A developmental comparison of identity and equivalence conservations. *Journal of Experimental Child Psychology*, **12**(3), 347–361. https://doi.org/10.1016/0022-0965%2871%2990030-0

Parisi, D., & Schlesinger, M. (2002). Artificial life and Piaget. *Cognitive Development*, **17**(3-4), 1301–1321. https://doi.org/10.1016/S0885-2014(02)00119-3

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, **32**(4), 301–345. https://doi.org/10.1016/0001-6918%2870%2990108-3

Pascual-Leone, J. (1976). On learning and development, Piagetian style: II. A critical historical analysis of Geneva's research programme. *Canadian Psychological Review/Psychologie Canadienne*, **17**(4), 289–297. https://doi.org/10.1037/h0081846

Pasnak, R., Hansbarger, A., Dodson, S. L., Hart, J. B., & Blaha, J. (1996). Differential results of instruction at the preoperational/concrete operational transition. *Psychology in the Schools*, **33**(1), 70–83. https://doi.org/10.1002/(sici)1520-6807(199601)33:1<70::aid-pits9>3.3.co;2-t

Pasnak, R., Kidd, J. K., Gadzichowski, K. M., Gallington, D. A., Schmerold, K. L., & West, H. (2015). Abstracting sequences: Reasoning that is a key to academic achievement. *The Journal of Genetic Psychology*, **176**(3), 171–193. https://doi.org/10.1080/00221325.2015.1024198

Pears, R., & Bryant, P. (1990). Transitive inferences by young children about spatial position. *British Journal of Psychology*, **81**(4), 497–510. https://doi.org/10.1111/j.2044-8295.1990.tb02375.x

Peckham, R. H. (1933). Visual discrimination in preschool children. *Child Development*, **4**, 292–297. https://doi.org/10.2307/1125767

Pennington, B. F., Bennetto, L., McAleer, O., & Roberts, R. J., Jr. (1996). Executive functions and working memory: Theoretical and measurement issues. In G. R. Lyon, & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 327–348). Baltimore, MD: Paul H Brookes Publishing Co.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Special Issue: Annual Research Review*, **37**(1), 51–87. https://doi.org/10.1111/j.1469-7610.1996.tb01380.x

Perner, J., & Mansbridge, D. G. (1983). Developmental differences in encoding length series. *Child Development*, **54**(3), 710–719. https://doi.org/10.2307/1130059

Perner, J., Steiner, G., & Staehelin, C. (1981). Mental representation of length and weight series and transitive inferences in young children. *Journal of Experimental Child Psychology*, **31**(2), 177–192. https://doi.org/10.1016/0022-0965(81)90011-4

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, **20**(3), 249–262. https://doi.org/10.1016/0028-3932%2882%2990100-2

Piaget, J. (1928). *Judgment and reasoning in the child*. London: Kegan Paul, Trench, Trubner & Co. Ltd.

Piaget, J. (1952a). *The child's conception of number*. London: Routledge and Kegan Paul.

Piaget, J. (1952b). *The origins of intelligence in children* (**8**). New York: International Universities Press.

Piaget, J. (1968). *Genetic epistemology*. Paper presented at the series of lectures by Jean Piaget, New York: Columbia University.

Piaget, J. (1972). *The child and reality: Problems of genetic psychology*. London: Muller.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). The child's conception of geometry. London: Routledge and Kegan Paul.

Piaget, J., & Inhelder, B. (1971). *The child's conception of space*. New York: Norton.

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities*. London: Routledge and Kegan Paul.

Piaget, J., & Szeminska, A. (1941). *La genese du nombre chez l'enfant*. Neuchatel: Delachaux & Niestle.

Pick, H. L. (1992). Eleanor J. Gibson: Learning to perceive and perceiving to learn. *Developmental Psychology*, **28**(5), 787–794.

Plunkett, K., & Elman, J. L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations*. Cambridge, MA: MIT Press.

Pollack, J. B. (2014). Mindless intelligence: Reflections on the future of AI. In P. A. Vargas, E. A. Di Paolo, I. Harvey, & P. Husbands (Eds.), *The horizons of evolutionary robotics*. Cambridge, MA: MIT Press.

Quinlan, P. T. (2003). *Connectionist models of development: Developmental processes in real and artificial neural networks*. New York, NY: Psychology Press.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Reese, H. W. (1968). *The perception of stimulus relations: Discrimination learning and transposition*. New York: Academic Press.

Riley, C. A., & Trabasso, T. (1974). Comparatives, logical structures, and encoding in a transitive inference task. *Journal of Experimental Child Psychology*, **17**, 187–203.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, **5**(5), 266–281. https://doi.org/10.1017/S1355617798455073

Ross, T. P., Hanouskova, E., Giarla, K., Calhoun, E., & Tucker, M. (2007). The reliability and validity of the self-ordered pointing task. *Archives of Clinical Neuropsychology*, **22**(4), 449–458. https://doi.org/10.1016/j.acn.2007.01.023

J. Rubenstein, & P. Rakic (Eds.). (2013). *Neural circuit development and function in the brain, Part II*. New York: Elsevier.

Russac, R. J. (1978). The relation between two strategies of cardinal number: Correspondence and counting. *Child Development*, **49**(3), 728–735. https://doi.org/10.2307/1128241

Russell, J. (1981). Children's memory for the premises in a transitive measurement task assessed by elicited and spontaneous justifications. *Journal of Experimental Child Psychology*, **31**(2), 300–309. https://doi.org/10.1016/0022-0965%2881%2990019-9

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, **9**, 331–349. https://doi.org/10.1111/j.2044-835X.1991.tb00881.x

Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, **23**(10), 1164–1179. https://doi.org/10.1016/j.neunet.2010.07.012

Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: Children's understanding of cardinality and equinumerosity. *Cognitive Science*, **37**(8), 1493–1506. https://doi.org/10.1111/cogs.12043

Schlesinger, M., & McMurray, B. (2012). The past, present, and future of computational models of cognitive development. *Cognitive Development*, **27**(4), 326–348. https://doi.org/10.1016/j.cogdev.2012.07.002

Schliemann, A. D. (1983). Relational terms and logical operations. *Termos relacionais e operacoes logicas*, **9**(2), 17–25.

Schnall, S., & Gattis, M. (1998). Transitive inference by visual reasoning. In M. A. Gernsbacher, & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 929–934). Hillsdale, NJ: Lawrence Erlbaum.

Scott, P. J., & Nicolson, R. I. (1991). *Cognitive science projects in Prolog*. Hillsdale, NJ: Lawrence Erlbaum.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, **3**(3), 417–424.

Semenov, L. A., Chernova, N. D., & Bondarko, V. M. (2000). Measurement of visual acuity and crowding effect in 3-9-year-old children. *Human Physiology*, **26**(1), 16–20. https://doi.org/10.1007/BF02760711

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, **35**(6), 679–697. https://doi.org/10.1080/87565641.2010.508546

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Siegel, L. S. (1972). Development of the concept of seriation. *Developmental Psychology*, **6**(1), 135–137. https://doi.org/10.1037/h0032205

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology*, **28**(3), 225–273. https://doi.org/10.1006/cogp.1995.1006

Siegler, R. S. (1996). A grand theory of development: Commentary on Okamoto and Case. *Monographs of the Society for Research in Child Development*, **246**(61), 266–276. https://doi.org/10.1111/j.1540-5834.1996.tb00550.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation. *Psychological Science*, **14**(3), 237–250. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., Strauss, S., & Levin, I. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, **46**(2), 1–74.

Sigel, I. E. (1968). Reflections. In I. E. Sigel, & F. H. Hooper (Eds.), *Logical thinking in children* (pp. 503–528). New York, NY: Holt, Rinehart and Wilson.

I. E. Sigel, & F. H. Hooper (Eds.). (1968). *Logical thinking in children*. New York: Holt, Rinehart and Winston.

Sigel, I. E., Roeper, A., & Hooper, F. H. (1966). A training procedure for acquisition of Piaget's conservation of quantity: A pilot study and its replication. *British Journal of Educational Psychology*, **36**(3), 301–311. https://doi.org/10.1111/j.2044-8279.1966.tb01883.x

Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a “button-press” go/no-go task. *Journal of Experimental Child Psychology*, **94**(1), 18–26. https://doi.org/10.1016/j.jecp.2005.10.003

Smedslund, J. (1963). Development of concrete transitivity of length in children. *Child Development*, **34**, 389–405. https://doi.org/10.2307/1126735

Smith, L. (1993). *Necessary knowledge: Piagetian perspectives on constructivism*. Hove, East Sussex: Lawrence Erlbaum Associates, Ltd.

Smith, L. B., & Breazeal, C. (2007). The dynamic lift of developmental process. *Developmental Science*, **10**(1), 61–68. https://doi.org/10.1111/j.1467-7687.2007.00565.x

Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, **8**(4), 553–560. https://doi.org/10.1016/0010-0285%2876%2990018-9

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, **15**(6), 626–636. https://doi.org/10.1037/0012-1649.15.6.626

Spence, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review*, **43**, 427–449. https://doi.org/10.1037/h0056975

Stenning, K., & Lambalgen, V. M. (2012). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Stevenson, H. W., & McBee, G. (1958). The learning of object and pattern discriminations by children. *Journal of Comparative and Physiological Psychology*, **51**(6), 752–754. https://doi.org/10.1037/h0039756

Stock, P., Desoete, A., & Roeyers, H. (2007). Early markers for arithmetic difficulties. *Special Issue: Arithmetical Difficulties: Developmental and Instructional Perspectives*, **24**(2), 28–39.

Stock, P., Desoete, A., & Roeyers, H. (2009). Predicting arithmetic abilities: The role of preparatory arithmetic markers and intelligence. *Journal of Psychoeducational Assessment*, **27**(3), 237–251. https://doi.org/10.1177/0734282908330587

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Swanson, R. A., Henderson, R. W., & Williams, E. (1979). The relative influence of observation, imitative motor activity, and feedback on the induction of seriation. *The Journal of Genetic Psychology*, **135**, 81–91. https://doi.org/10.1080/00221325.1979.10533419

Terrace, H. S., & McGonigle, B. (1994). Memory and representation of serial order by children, monkeys, and pigeons. *Current Directions in Psychological Science*, **3**(6), 180–185. https://doi.org/10.1111/1467-8721.ep10770703

Thayer, E. S., & Collyer, C. E. (1978). The development of transitive inference: A review of recent approaches. *Psychological Bulletin*, **85**(6), 1327–1343. https://doi.org/10.1037/0033-2909.85.6.1327

Thomas, R. K., & Crosby, T. N. (1977). Absolute versus relative class conceptual behavior in squirrel monkeys (*Saimiri sciureus*). *Animal Learning & Behavior*, **5**(3), 265–271. https://doi.org/10.3758/BF03209238

Timmons, S. A., & Smothergill, D. W. (1975). Perceptual training of height and brightness seriation in kindergarten children. *Child Development*, **46**(4), 1030–1034. https://doi.org/10.2307/1128419

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2011). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities*, **44**(6), 521–532. https://doi.org/10.1177/0022219410387302

Tomlinson-Keasey, C., Eisert, D. C., Kahle, L. R., Hardy-Brown, K., & Keasey, B. (1979). The structure of concrete operational thought. *Child Development*, **50**(4), 1153–1163. https://doi.org/10.2307/1129343

Trabasso, T., Riley, C., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Psychological studies of logic and its development* (pp. 201–230). Hillsdale, NJ: Lawrence Erlbaum.

Trabasso, T., & Riley, C. A. (1975). On the construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 381–410). Mahwah, NJ: Lawrence Erlbaum.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. London: Harvester Wheatsheaf.

Van Gelder, T. (1995). What might cognition be if not computation? *Journal of Philosophy*, **92**(7), 345–381. https://doi.org/10.2307/2941061

Voneche, J., & Vidal, F. (1985). Jean Piaget and the child psychologist. *Synthese*, **65**, 121–138.

Wasserman, T., & Wasserman, L. D. (2013). Toward an integrated model of executive functioning in children. *Applied Neuropsychology: Child*, **2**(2), 88–96. https://doi.org/10.1080/21622965.2013.748394

H. M. Wellman, & K. Inagaki (Eds.). (1997). *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena*. San Francisco, CA: Jossey-Bass.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive development*, **6**(1), 59–76. https://doi.org/10.1016/0885-2014%2891%2990006-Y

Werner, H. (1948). *Comparative psychology of mental development*. New York: Harper.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wright, B. C., Robertson, S., & Hadfield, L. (2011). Transitivity for height versus speed: To what extent do the under-7s really have a transitive capacity? *Thinking & Reasoning*, **17**(1), 57–81. https://doi.org/10.1080/13546783.2010.544548

Wynn, K. (2008). Some innate foundations of social and cognition. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Foundations and the future* (**3**, pp. 330–347). New York, NY: Oxford University Press.

Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, **83**(3), B55–B62. https://doi.org/10.1016/S0010-0277%2802%2900008-2

Xu, C., & LeFevre, J. A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology*, **52**(6), 854–866. https://doi.org/10.1037/dev0000124

Yeh, J. C. Y. W. (1970). Transposition and transfer of absolute responding as functions of learning-set training and stimulus similarity. *Journal of Experimental Child Psychology*, **10**(1), 57–66. https://doi.org/10.1016/0022-0965%2870%2990044-5

Young, A. R., Gurm, M. K., & O'Donnell, K. A. (2017). Assessing executive functions in young children. In M. J. Hoskyn, G. Iarocci, & A. R. Young (Eds.), *Executive functions in children's everyday lives: A handbook for professionals in Applied Psychology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199980864.001.0001

Young, R. M. (1976). *Seriation by children: An artificial intelligence analysis of a Piagetian task*. Basel: Birkhäuser.

Young, R. M., & Lewis, R. L. (1999). The SOAR cognitive architecture and human working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 224–256). New York, NY: Cambridge University Press.

Youniss, J., & Murray, J. P. (1970). Transitive inference with non-transitive solutions controlled. *Developmental Psychology*, **2**(2), 169–175. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0028763

Yule, P. G., Fox, J., Glasspool, D. W., & Cooper, R. P. (2013). *Modelling high-level cognitive processes*. Oxford: Psychology Press.

Zeiler, M. D., & Friedrichs, A. G. (1969). Absolute and relational diThough clearly different fromscriminations involving three stimuli. *Journal of Experimental Psychology*, **82**(3), 448–452. https://doi.org/10.1037/h0028423

Zeiler, M. D., & Gardner, A. M. (1966). Intermediate size discrimination in seven- and eight-year-old children. *Journal of Experimental Psychology*, **71**(2), 203–207. https://doi.org/10.1037/h0022887

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). VI. The Development Of Executive Function: Cognitive complexity and control-revised. *Monographs of the Society for Research in Child Development*, **68**(3), 93–119.

Authors

Maggie McGonigle-Chalmers, M.A., Ph.D., is a retired Senior Lecturer (now Honorary Fellow) from the University of Edinburgh, Psychology Department with research interest in Developmental Cognition and Autism Spectrum Disorders. In addition to the publications arising from that work, she has authored a textbook entitled *Understanding Cognitive Development* (SAGE Publications, 2015).

Iain Kusel, M.A., M.Sc., Ph.D., is a software engineer with a research interest in developmental cognition and computational modeling. He received his Ph.D. from the Department of Engineering and Innovation, Faculty of Maths, Computing and Technology, The Open University.

SUBJECT INDEX

Page numbers in *italics* represent figures and tables.

E

M

N

O

P

R

S

Information for subscribers
Monographs of the Society for Research in Child Development is published quarterly in March, June, September, and December. Institutional subscription prices for Monographs of the Society for Research in Child/Child Development/Child Development Perspectives for 2019 are: Print & Online: US$1285, US$1796 (Rest of World), €1168 (Europe), £921 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST/HST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2015, where available. For other pricing options, including access information and terms and conditions, please visit onlinelibrary.wiley.com/library-info/products/price-lists. Terms of use can be found here: onlinelibrary.wiley.com/terms-and-conditions.

Delivery Terms and Legal Title
Where the institutional subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers Free of Board (FOB) our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

Back issues: Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue - Ste 300, Hudson, NY 12534, USA. Tel: +1 518 822 9300, Fax: +1 518 822 9305, Email: psc@periodicals.com

MICROFORM The journal is available on microfilm. For microfilm service, address inquiries to ProQuest Information and Learning, 300 North Zeeb Rd., Ann Arbor, MI 48106-1346, USA. Bell and Howell Serials Customer Service Department (800) 521-0600 ×2873.

Postmaster: Send all address changes to MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, John Wiley & Sons Inc., C/O The Sheridan Press, PO Box 465, Hanover, PA 17331.

THE MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, (ISSN 0037-976X (print), ISSN 1540-5834 (online)) is published quarterly, on behalf of The Society for Research in Child Development, Inc., by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774.

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.
Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada).
Europe, Middle East and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315.
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000.
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Visit our Online Customer Help available in 7 languages at www.wileycustomerhelp.com/ask

Production Editor: Stephanie Hill (MONO@wiley.com)

View this journal online at wileyonlinelibrary.com/journal/mono

Wiley is a founding member of the UN-backed HINARI, AGORA, and OARE initiatives. They are now collectively known as Research4Life, making online scientific content available free or at nominal cost to researchers in developing countries. Please visit Wiley's Content Access – Corporate Citizenship site: www.wiley.com/WileyCDA/Section/id-390082.html

Monographs of the Society for Research in Child Development accepts articles for Open Access publication. Please visit http://authorservices.wiley.com/author-resources/Journal-Authors/open-access/onlineopen.html for further information about OnlineOpen.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts. Follow our progress at www.wiley.com/go/citizenship

For submission instructions, subscription and all other information visit: wileyonlinelibrary.com/journal/mono

ORCID® provides a persistent digital identifier that distinguishes a researcher from every other researcher. Through integration in key research workflows such as manuscript and grant submission, ORCID supports automated linkages between the researcher and his/her professional activities ensuring that the work is recognized. For more information, please visit http://orcid.org.

Printed in the USA by The Sheridan Group

ISSN 0037-976X (Print)
ISSN 1540-5834 (Online)

monographmatters.srcd.org

The Development of Size Sequencing Skills: An Empirical and Computational Analysis

In this monograph, Maggie McGonigle-Chalmers and Iain Kusel report an investigation into a phenomenon called size seriation. At around the age of seven years children suddenly become capable of systematically organizing objects in order of size. Using touchscreen tasks, they explore the differences between children of five and seven years when learning seriation tasks and when trying to identify size relations such as middle-sized. A computer model simulates the findings and shows how the act of size sequencing itself, together with an increase in memory capacity, creates a new solution for the older child that is not available to the younger child. Taken together, the findings and model reveal changes in mental functioning that explain spontaneous seriation and how the concept of a "unit" emerges during development.

Maggie McGonigle-Chalmers, MA, Ph.D.

is a retired Senior Lecturer (now Honorary Fellow) from the University of Edinburgh, Psychology Department with a research interest in Developmental Cognition and Autism Spectrum Disorders. In addition to publications arising from that work, she has authored a textbook entitled *Understanding Cognitive Development* (SAGE Publications, 2015).

Iain Kusel, MA, MSc, Ph.D.

is a software engineer with a research interest in developmental cognition and computational modelling. He received his Ph.D. from the Department of Engineering and Innovation, Faculty of Maths, Computing and Technology, The Open University. He was affiliated with the University of Edinburgh at the time this work was completed, and is now a software engineer at the Sword Group.

Front cover photo credit: Dan Weiss/Penn State/danw.smugmug.com