Discoveries of the Algebraic Mind: A PRIMs Model

Mark Y. Ji (y.ji@rug.nl)
Jacolien van Rij (j.c.van.rij@rug.nl)
Niels A. Taatgen (n.a.taatgen@rug.nl)
Department of Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, Netherlands

Abstract

The PRIMs (primitive element) cognitive architecture addresses the issue of deterministic programming of production-rules (Taatgen, 2017). Motivated by infants' flexible discoveries of simple rule-like algebraic patterns (e.g., *a-a-b*, *a-b-a*, and *a-b-b* types of patterns, with variable individual syllable tokens), this study illustrates how the gradual integration of primitive operations to task-related contexts can be made possible through a reward-guided contextual learning mechanism. The promise of this prototypical model is demonstrated in its ability to (a) learn and generalize simple algebraic patterns, and (b) to account for infants' differential focusing time on a learned pattern and other unexposed new patterns. The modeled results are summarized from a developmental plasticity perspective.

Keywords: infant, discovery; robust; plastic; PRIMs.

Introduction

The ongoing presentation of distinct new micro-theories in cognitive psychology makes it difficult to see the forest for the trees (see Newell, 1990). Thanks to the advent of ACT-R (adaptive control of thought, rational; Anderson et al., 2004), seemingly disparate task-related aspects of cognition are now frameable within an overarching cognitive architecture. However, *a priori* programming of just one solution of a task again neglects other possible task-related solutions, and fails to capture the trial and error discovery processes observed in real-life task performance (see Taatgen, 2017). The need for modeling flexible discovery is especially motivated by infant learning, since young infants cannot be taught as how to complete a task but must arrive at their own solutions. In this paper, we present a model with particular focus on the flexible discoveries of simple algebraic patterns during infancy (see Marcus et al., 1999). We start by introducing the paradigm, and then briefly review previous models from a multilevel view (see Taatgen, 2017).

Infants seem innately capable of detecting simple algebraic patterns, and generalize them without relying on statistical features of the learned patterns. In an experiment reported by Marcus et al. (1999), infants with brief exposure to audible sequential presentation of the *a-b-a* or the *a-b-b* type of pattern (note the symbols *a* and *b* here refer to distinctive syllable tokens that are variable), during the test phase showed longer preferential focusing time for the other unexposed types of patterns as compared to the just learned pattern. This held even when each of the syllable token within the pattern were drawn from a different set (e.g., focusing longer for "*ko-ko-ga*" as compared to "*ko-ga-ko*" at test, after just exposing to "*le-we-le*"). Recent years have seen replication of this phenomenon with visually presented patterns (for a meta-analytical review, see Rabagliati, Ferguson, & Lew-Williams, 2019), and its extension to more complex variants of algebraic patterns (see Wilson et al., 2018). For instance, the acquisition of the *a-b-a* type of pattern is now considered a specific case of non-adjacent dependency learning, when an infant predicts that the first token *a* always matches the third token *b*.

With sensory perception as a point of departure, some modelers speculated that the acquisition of algebraic patterns is merely a basic form of feature detection (see McClelland & Plaut, 1999). Nevertheless, most models were built from a slightly higher level of abstraction (see Altmann, 2017), in assuming infants to be capable of forming representations from features (see Saffran & Thiessen, 2007). These representational models implicitly assume that young infants can deriving relational rules from a complex representation. However, very young infants cannot flexibly retrieve for instance a syllable "*le*" as parsed from a complex pattern of "*le-we-le*" (see Richmond & Nelson, 2007). Even when this constraint is suspended, a recent biologically-inspired representational model is only capable of generalize simple algebraic patterns at chance level (Alhama & Zuidema, 2018). This observation calls into question whether feature and/or representation alone are indeed sufficient or plausible in explaining the acquisition of simple algebraic patterns (see Dawson & Gerken, 2012; Frank & Tenenbaum, 2011).

On the contrary, models applying rule-based processes not only successfully modeled the learning of simple algebraic patterns (Seidenberg & Elman, 1999), but also stimulated a wide range of studies in the field of algebraic pattern acquisition (Frank & Tenenbaum, 2011). The Bayesian model of Frank and Tenenbaum (2011) demonstrated multiple algebraic solutions based on a hypothesis space of primordial rules. Moreover, the results modeled with this approach reflected emergent distinctions between *type*- and *token*-based processes (Frank & Tenenbaum, 2011). In other words, there is an empirical age-related distinction between young infants' early capability of detecting the *types* of *a-b-a*/*a-b-b* patterns; and their slightly delayed ability to detect the invariant *a-b* token pair in the *a-x-b* pattern where it is separated by a variable *x* (Dawson & Gerken, 2012). The distinction between *type*- and *token-based* processes are often interpreted in terms of the *exogenous*-to-*endogenous* transition, when early infants' passive *exogenous* reactions to the environment are gradually augmented by their active *endogenous* flexible retrieval of information as parsed from a complex representation (Diego-Balaguer et al., 2016).

How early infants might flexibly learn to recognize simple algebraic patterns remains to be explained. Emergent evidence now suggests that the infant brain possesses a modular architecture (see Dehaene-Lambertz & Spelke, 2015), thus calling for its conceptual implementation in studying the cognition of infants. Specifically, the *exogenous* reactions can be mapped to the passive encoding and comparisons at various modules, and the *endogenous* processes can be mapped to the active retrieval from declarative memory (Colombo & Cheatham, 2006; cf., Stocco & Anderson, 2008). Moreover, recent evidence indicates that the language-related prefrontal area is already functional during infancy (see Dehaene-Lambertz & Spelke, 2015), which can facilitate simple task-relevant processes such as the detection of syllable repetition (Bristow et al., 2009). Nakano et al. (2009) further reported selective activation of the prefrontal cortex in infants upon repetition of a syllable, and upon alteration of the syllable, which demonstrates inherent sensitivity of frontal structures to the establishment and alteration of the task requirement. It is possible that frontal activation follows a reward-guided mechanism that integrates and strengthens the currently acquired adaptive skills for future use (cf., Duncan, 2010).

Based on this empirical background, a modular and adaptive architecture is a well-suited tool for studying infant learning. Here, the PRIMs (primitive element) cognitive architecture is a promising candidate (Taatgen, 2013). It follows a modular structure pioneered by ACT-R, with additional prospects for the flexible discovery of rule-like patterns. This discovery mechanism is comprehensible from the perspective of functional development (see Bateson & Gluckman, 2011). To illustrate, initially randomly fired lower-level processes may occasionally lead to the successful detection of a repetition. This then entails a higher-level reward-guided mechanism that integrates various just applied lower-level operations to their associated task contexts, thus making them context-sensitive.

In this paper, we first aim to show how simple algebraic patterns can be acquired and generalized. Based on that, we attempt to account for the empirical findings in infants' focusing time differences in reacting to learned and other unexposed types of simple algebraic patterns.

Model

There follows a brief description of PRIMs operations at both the lower- and higher-levels.

Primitive Operations

The PRIMs cognitive architecture breaks down artificially programmed production-rules into elemental processes that can copy and compare information between separate slots in the input channel and the various memory modules (see Figure 1). These processes are called primitive operations, and they can be flexibly fired during task exploration.
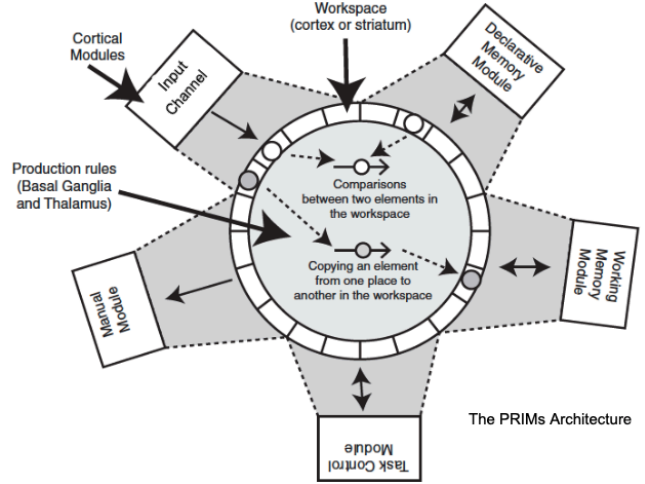

Figure 1: The PRIMs architecture for skill acquisition.

This PRIMs model of infant learning includes all possible lower-level primitive operations that encode (e.g., encode information within the input channel to the working memory module; see L_{encode}) or compare (e.g., compare whether information within the input channel matches to the working memory module; see $L_{compare}$) information between various modules (see Figure 1, Table 1, *ik* in table refers to $slot_k$ in $chunk_i$). However, a constraint is placed upon infants' processing capacity. This constraint acknowledges that the infants cannot yet simultaneously process multiple representations (e.g., encode/retrieve distinct representations "*le*" and "*we*" at the same time), and neither can they retrieve detailed information (e.g., syllable token "*le*") as parsed from a more complex representation (e.g., representation of the pattern "*le-we-le*"). When a condition is met for repetition detection, a scaffolding operation ($L_{scaffold}$) enables state transition to evaluation. Note that state transition to evaluation may also be flexibly entailed without scaffolding.

Table 1: Primitive operations.

L_{encode}	$input_{ik} \not\Rightarrow working/decl.\ memory_{ik}$	ignore
	$input_{ik} \Rightarrow working/decl.\ memory_{ik}$	encode
	$working/decl.\ memory_{ik} \Rightarrow control_{ik}$	encode
$L_{compare}$	$input_{ik} = working/decl.\ memory_{ik}$	compare
	$input_{ik} \neq working/decl.\ memory_{ik}$	compare
$L_{scaffold}$	"evaluation" $\Rightarrow control_{ik}$	transition

Reward-Guided Contextual Learning

In addition to the firing of lower-level operations, adaptive skills need to be arranged to satisfy and accomplish a defined task more efficiently. This is achieved by another higher-level evaluation operation. In this model, the evaluation operation is activated only when the presented stimulus at the input channel matches to the stored representation at the task control module ($H_{evaluation}$, see Table 2). This operation quickly entails a reward-guided contextual learning mechanism that reinforces the associations of just fired operations with their relevant task contexts – namely, *which* operation to fire at *what* context. For instance, to successfully detect a repetition in "*le-we-le*", the model always needs to encode the first token "*le*" with reference to its task contexts such as its general position "*first*" or its specific value "*le*". Gradually, the flexible firing of operations starts forming robust context-sensitive skills (i.e., encode-"*first*", or encode-"*le*"), which may be employed during relevant future contexts. Primitive operations can also be compiled to process more efficiently (e.g., "input ⇒ memory" and "memory ⇒ control" may be compiled into "input ⇒ control").

Table 2: Task-related operations.

$H_{evaluation}$	$input_{ik} = control_{ik}$	match
	$input_{ik} \neq control_{ik}$	mismatch

In this model, the weight of contextual association between a certain $operation_j$ and its relevant task context - in this case the specific syllable token stored in $slot_i$ of $module_k$ - is reflected in the following equation:

$$\Delta S_{jik} = \beta \left(S_{jik \ (current \ trial)} - S_{jik \ (previous \ trial)} \right)$$

in which, $S_{jik \ (current \ trial)} = default \ association \times (expected \ time - actual \ time) / expected \ time$

In this equation, *actual time* is the *actual* trial completion time, while *expected time* is hypothetically set initial trial completion time. The changing rate of the association weight S_{jik} is moderated both by (a) how efficiently the task was completed (when *actual time* < *expected time*) and (b) a learning rate parameter β. The *default association* sets the maximum weight for any contextual associations.

Default-Mode Operations

At the stage when the task is well learned, the firing of task-related operations become more efficient, increasing task-negative transitional spaces between them. The transitional spaces can become frequently occupied by the default-mode operation ($H_{default-mode}$, see Table 3), which are reinforced also by the contextual learning mechanism (cf., Smith et al., 2018). In other words, default-mode operation starts also to bind with task contexts whenever task-relevant operations are not active. The activation of default-mode operation is initially set at a low magnitude, but is gradually increased when it is more often fired and integrated also to the task contexts.

Table 3: Default-mode operations.

$H_{default-mode}$	$input_{ik}$ = "empty"	task-irrelevant processes

Methods

The modeled algebraic paradigm is adapted from Marcus et al. (1999, Exp. 2 and 3). In the first simulation, the focus is placed on the learnability of the *a-a-b*, *a-b-a*, and *a-b-b* patterns, each based on 100 model runs. The individual tokens of the trisyllabic pattern are presented each for 330 ms, with an ISI of 250 ms following each syllable token, and an ITI of 1000 ms following each pattern. The patterns are randomly drawn from a pool of 16 examples for each pattern type as adapted from Marcus et al. (1999, *a-b-a*/*a-b-b* type from Exp. 2 and *a-a-b* type from Exp. 3). The modeled trial is considered successful when repetition is detected during the evaluation operation (e.g., if $input_{ik} = control_{ik}$, estimated success = 1). This will in turn issue a reward to the model, strengthening associations of the manifestly adaptive lower-level operations with their relevant task contexts. Otherwise, the model is considered unsuccessful (e.g., if $input_{ik} \neq control_{ik}$ or no comparisons were made, estimated success = 0), and contextual associations for the operations during this trial will remain unchanged. To illustrate the gradual learning progression, 400 trials for each of the *a-b-a*, *a-b-b*, and *a-a-b* patterns are included.

Alternatively, simulations 2 and 3 focus on the generalization of the learned patterns, each based on 100 model runs. In the learning phase, the models are identical with the first simulation, except for the number of learning trials included. Specifically, simulations 2 was run for 150 learning trials, while simulation 3 was run for 500 trials to illustrate the effects of overlearning. The capacity of the model to generalize was then tested with novel examples of the learned pattern (e.g., *ko-ga-ko*) or other unexposed types of patterns (e.g., *ko-ga-ga, ko-ko-ga*). In the test phase that exams pattern generalization, specifications of primitive operations were not included. Instead, the models directly apply those operations and skills acquired from the learning phase to generalize them in the novel task contexts. To illustrate the trajectory of generalization, 150 transfer trials were included for each of the *a-a-b, a-b-a*, and *a-b-b* patterns.

Finally, the same learning and transfer models are applied in simulation 4 to illustrate critical differences in the empirical finding - in other words, infants' preferential longer focusing time on other unexposed types of patterns versus the learned pattern during test phases. The simulation consists of 100 learning and 10 transfer trials for each pattern, whereupon the frequency of default-mode operations during transfer trials is then calculated.

Results

Learning

Results of *simulation 1* demonstrate the model's ability to learn simple algebraic patterns (see Figure 2, averages with 95% CI error bars). Acquisition of all patterns converged to high percentages of correct predictions, albeit at different learning rates.
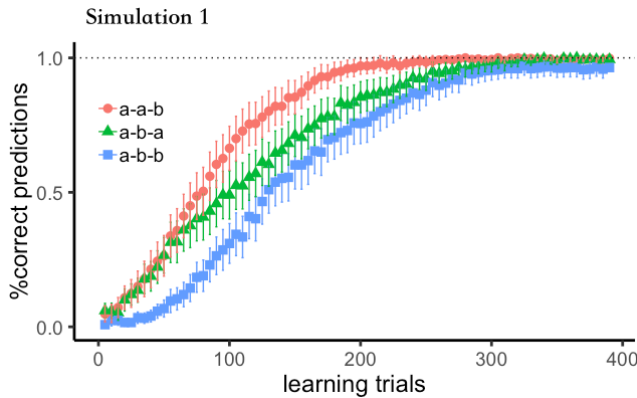

Figure 2: The discoveries of *a-a-b*, *a-b-a*, and *a-b-b* algebraic patterns. Horizontal axis shows learning trials from 1 to 400. Vertical axis shows the averaged percentages of correct predictions across 100 model runs (with 95% CI error bars).

It is easy to grasp that the learning of *a-b-a* is slightly more difficult than *a-a-b*, since irrelevance of second item *b* in *a-b-a* needs to be additionally acquired for repetition detection. However, it is less straightforward to explain the slower learning rate of *a-b-b*. This is nevertheless consistent with a recent finding showing 11-month-olds difficulty in detecting repetition in the *a-b-b-c* pattern (Schonberg, Marcus, & Johnson, 2017), and the slight advantage of initial versus late repetition (*a-a-b* versus *a-b-b*) in neonates at the neural level (Exp. 3, Gervain, Berent, & Werker, 2012). Note both findings were interpreted in terms of the primacy effect. Similarly, the simulation results similarly show a primacy effect at the skill level (see Figure 3). The first token in *a-b-b* must be "ignored" (orange) against the readily firing of various "encode" operations at the first position that are otherwise essential in learning *a-b-a* and *a-a-b*.

Another feature of the model is found in its ability to select an initial range of operations, while remaining capable of converging on to relatively invariant solutions when robust skills are formed (Figure 3). For instance, the model can flexible encode the item to the task control module, either from the working memory module (brown) or from the declarative memory module (purple). Nevertheless, when selection of the declarative memory route gradually organized into a robust state, it then becomes difficult to return to the initial flexible state in selecting an alternative working memory route. The modeled results in Figure 3 also revealed a gradual increase of default-mode operations (blue)

when the selection of task-relevant operations gradually stabilize.

Generalization

Simulations 2 (150 training trials) and *simulation 3* (500 training trials) demonstrates the generalization of algebraic patterns from learning, based on 100 model runs (see Figure 4, averages with 95% CI error bars). Results in *simulation 2* show that an optimal level of learning facilitates transfer of a learned pattern for other novel patterns (Figures 4A, 4B, and 4C, cf., Taatgen, 2013). Note that the transfer rates are moderated also by the degree of difficulty to learn that pattern. To the contrary, modeled results of *simulation 3* predict hindrance of transfer due to overlearning (Figures 4A, 4B, and 4C). Although infants may not realistically be expected to participate in a prolonged learning session, overfitting to a particular context may still render the system less adaptive to a slightly altered context (e.g., a deterioration of prediction rates even for the same pattern with altered tokens).

Lastly, results of *simulation 4* shows a higher frequency of default-mode operation when a pattern has been learned (Figure 5, averages with 95% CI error bars). Default-mode operations may cause infants to divert from the task, and are therefore likely to have an inverse relation to the time they would be focusing on the task. These simulated results are consistent with the findings of Marcus et al. (1999).
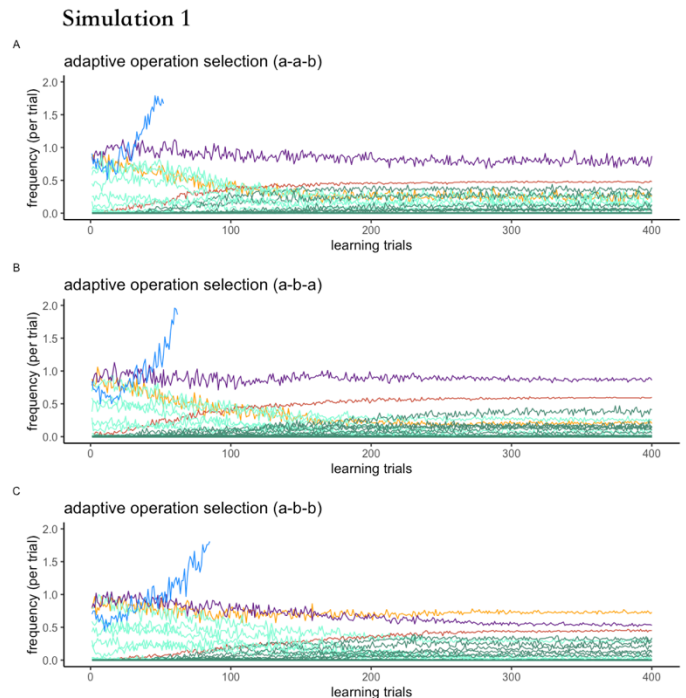

Figure 3: Operation selection over learning trials. Horizontal axis shows the number of learning trials. Vertical axis shows the frequency of various operations applied in a trial as averaged over 100 model runs (with 95% CI error bars). Color coding: purple, declarative mem. encode; brown, working mem. encode; orange, ignore; light-green, other primitive operations; dark-green, other compiled operations; blue, default-mode operations.
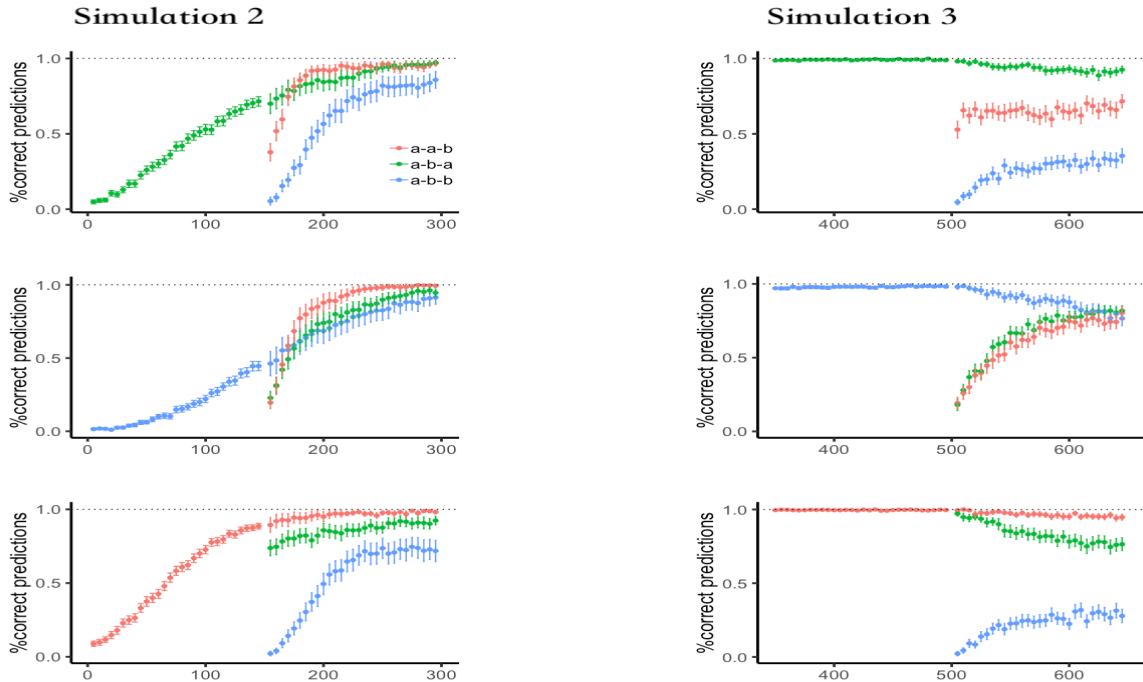
Simulation 2

Simulation 3

Figure 4. Generalization after learning or overlearning. Horizontal axis shows learning trials (150 trials in A, B, and C; 500 trials in D, E, and F) and the transfer trials (150 trials followed from the learning trials). Vertical axis shows the averaged percentages of correct predictions across 100 model runs (with 95% CI error bars).
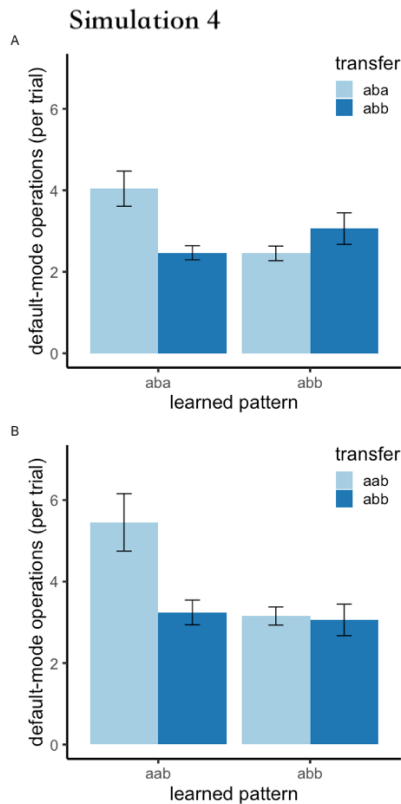
Simulation 4

Figure 5: Frequencies of default-mode operations in transfer. Horizontal axis shows the learned type. The bar colors denote types applied in the 10 test trials on generalization. The vertical axis shows frequencies of default-mode operation per trial, as averaged across 100 model runs (with 95% CI error bars).

Discussion

An Aristotelian axiom *nihil est in intellectu quod non sit prius in sensu* holds that there is nothing in the intellect that was not originally derived from the senses. However, more recent literature on cognition in infants has disputed whether the detection of simple algebraic patterns is purely a lower-level statistical process or follows higher-level rules. Towards reconciling these two disparate views, results of our present PRIMs model suggest that seemingly rule-like patterns can be gradually acquired from the bottom-up. The promise of the model is reflected in its ability (a) to learn and generalize simple algebraic patterns (cf., Marcus et al., 1999; Schonberg, Marcus, & Johnson, 2017); and (b) to account for differences in infants' preferential focusing time on learned patterns versus other unexposed types of patterns (cf., Marcus et al., 1999). The modeled results may be framed in terms of a contemporary view on developmental plasticity (cf., Bateson & Gluckman, 2011).

Contemporary biology and psychology may be said to be correcting an earlier overemphasis on whether cognitive development is innate or learned. It is now clear that altering an innate property (e.g., presence or absence of certain trait-related genetic factors) is not always equatable with changes in learned characteristics. Instead, environmental conditions are crucial in shaping the precise characteristics of a learned skill (Bateson & Gluckman, 2011). The present PRIMs model demonstrates equal possibility of various routes in detecting syllable repetition. For instance, when flexible retrieval is not yet developed, infants can still distinguish between algebraic patterns (Dawson & Gerken, 2012).

Nonetheless, an innate structural architecture undoubtedly provides the basis for primitive operations to function.

Furthermore, distinct characteristics such as robustness and plasticity are not as cleanly separated as once thought. For instance, people maintains certain typical ways of dealing with a problem, but can also become flexible when the problem is changed. An emergent view now holds that robust outcomes can be derived from individual's plasticity (Bateson & Gluckman, 2011). This present PRIMs model shows that robust context-sensitive skills can be gradually integrated through a reward-guided contextual learning mechanism, and that the achieved robustness also raise barriers against the application of other possible skills that were currently not integrated. In addition, robust skills may be co-opted in other task contexts achieving generalization (see Taatgen, 2013). On the other hand, the present model also points to the detriments of overlearning and extreme robustness during one learning instance, which hinder the system for instance to accommodate the same type of pattern with just the syllable tokens altered. This extreme case may be taken as similar to a deterministically programmed model that only monotonously performs one single task.

Furthermore, the present PRIMs model suggest that the empirical finding concerning infants' shorter focusing time on the learned versus the other unexposed types of patterns may be product of the degree of robustness. Specifically, efficient processing of robust skills encourages the firing of default-mode operation, and gradually diverts the system from the focused task. In real terms, this may be associated with displacement of the infant's attention to need for food, comfort, play and so forth, curtailing the focusing time for the simple algebraic pattern. As illustrated from the present model, the accumulation of default-mode operations could occur whenever the system is still exploring the task. This in turn suggests that focusing time difference may not be directly relevant to how well an infant habituates a representation or masters a rule. Currently we are applying the same model to account for the counterintuitive reversed focusing time findings (longer focusing time on the learned versus the other patterns) to the generalization of non-adjacent dependency pattern *a-x-b* (Gómez & Maye, 2005).

Conclusion

Our PRIMs model firstly shows that simple algebraic patterns can be discovered bottom-up through the interplay between flexible primitive operations and a reward-guided contextual learning mechanism. This adaptive process produces robust context-sensitive skills that not only satisfies a given task, but may be also generalized in other relevant tasks. Secondly, the present study shows that infants' differential focusing time on the learned versus other unexposed types of pattern may be indirectly related to the robustness/plasticity of skill integration. In other words, efficient skill processing may encourage default-mode operation that reduces task focus. The modeled results suggest a more cautious position on drawing a direct link between infants' focusing times and the habituation/rule-bound operation of simple algebraic patterns.

References

Alhama, R. G., & Zuidema, W. (2018). Pre-Wiring and Pre-Training: What does a neural network need to learn truly general identity rules? *Journal of Artificial Intelligence Research, 61*, 927–946.

Altmann, G. T. (2017). Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 20160060.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*, 1036.

Bateson, P., & Gluckman, P. (2012). Plasticity and robustness in development and evolution. *International Journal of Epidemiology, 41*, 219–223.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., & Baillet, S. (2009). Hearing faces: Crossmodal representations of speech in two-month-old infants. *Journal of Cognitive Neuroscience, 21,* 905–921.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior, 34*, 283–322.

Dawson, C. R., & Gerken, L. (2012). Can rational models be good accounts of developmental change? The case of language development at two time scales. *Advances in Child Development and Behavior, 43*, 95–124.

Dehaene-Lambertz, G., & Spelke, E. S. (2015). The infancy of the human brain. *Neuron, 88*, 93–109.

Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7*, 44.

Duncan, J. (2010). *How intelligence happens.* Yale University Press.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*, 183–206.

Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition, 120*, 360–371.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. Journal of *Cognitive Neuroscience, 24*, 564-574.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences, 3*, 166–167.

Nakano, T., Watanabe, H., Homae, F., & Taga, G. (2008). Prefrontal cortical involvement in young infants' analysis of novelty. *Cerebral Cortex, 19*, 455–463.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Rabagliati, H., Ferguson, B., & Lew-Williams, C. (2019). The profile of abstract rule learning in infancy: Meta-analytic and experimental evidence. *Developmental Science, 22*, e12704.

Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 68–86). Malden: Blackwell Publishing.

Schonberg, C., Marcus, G., & Johnson, S. (2017). The roles of item repetition and position in infant sequence learning. In: *CogSci 2017.* Cognitive Science Society.

Seidenberg, M. S., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*, 433–433.

Smith, V., Mitchell, D. J., & Duncan, J. (2018). Role of the default mode network in cognitive transitions. *Cerebral Cortex, 28*, 3685–3696.

Stocco, A., & Anderson, J. R. (2008). Endogenous control and task representation: an fMRI study in algebraic problem-solving. *Journal of Cognitive Neuroscience, 20*, 1300–1314.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review, 120*, 439–471.

Taatgen, N.A. (2017). Cognitive architectures: Innate or learned? *AAAI Fall Symposium Series,* 476–480.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, A., & Rey, A. (2018). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*.