

# Creating AI: A Unique Interplay Between the Development of Learning Algorithms and their Education

Anat Treister-Goren

Training Department, Artificial Intelligence NV, Savion, Israel, anat@a-i.com

Jason L. Hutchens

Research Department, Artificial Intelligence NV, Savion, Israel, hutch@a-i.com

## Abstract

Artificial Intelligence NV is a research and development company devoted to the creation of a computer program capable of passing the Turing Test. Our firm conviction is that the only way of achieving this is by returning to Turing's original vision of building a child machine and then training it to learn the language skill. In this paper we show how a unique interplay between creating learning algorithms and training them to converse has resulted in a system which satisfies this requirement. Internally code-named 'HAL', our child machine has successfully passed the Turing Test at the level of a fifteen-month-old human infant.

## 1. Introduction

In 1950 Alan Turing considered the question "Can machines think?" Turing's answer to this question was to define the meaning of the word 'think' in terms of a conversational scenario, whereby if a human interrogator cannot reliably distinguish between a machine and a human being solely on the basis of their conversational ability, then the machine can be said to be thinking (Turing, 1992). This procedure is nowadays referred to as the Turing Test.

Artificial Intelligence NV (Ai) is a company dedicated to a decade-long research project aimed at creating a computer program capable of passing the Turing Test. Ours is the first commercially-oriented research and development effort, with complete financial backing, that is committed to the creation of 'true' artificial intelligence.

The methods employed at Ai strictly adhere to Turing's measure of intelligence. General learning algorithms are researched, developed, and subsequently trained in a black-box fashion to reach precise language performance milestones, specified with reference to human lingual development. Evaluation of the system's performance is performed with respect to these milestones (Treister-Goren et al., 2000).

One of the many unique features of our venture is that equal importance is given to the development of learning algorithms and their training and evaluation. This approach involves cooperation on a day-to-day basis between researchers from very different disciplines. We are finding that this constant interplay helps us to guide our work in a way which focuses on resolving immediate performance issues. It has proven itself to be an extremely valuable approach.

We begin this paper with a brief historical overview of both the Turing Test and the traditional approach to artificial intelligence. We then describe how our research and training departments have worked together towards the creation of our own child machine, internally code-named 'HAL'. Finally we present some results of HAL's performance, including details of the Turing Test which it passed at the level of a fifteen-month-old.

## 2. Historical Overview

The Turing Test is an appealing measure of intelligence because, as Turing himself writes, it "has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man". The Loebner Contest, held annually since 1991, is an instantiation of the Turing Test. The sophistication and performance of computer programs entered into the contest, or lack thereof, is indicative of the fact that the field of artificial intelligence has largely ignored Turing's strict evaluation criterion.

In a recent thorough review of conversational systems, the absurdity of performance in the Loebner Contest was emphasized (Hasida and Den, 1999). Since the Turing Test requires that systems "talk like people", and since no system currently meets this requirement, the ad-hoc techniques which the Loebner Contest subsequently encourages make little contribution to the advancement of dialog technology.

Although we agree wholeheartedly that the Loebner Contest has failed to contribute to the advancement of artificial intelligence, we do believe that the Turing Test is an appropriate evaluation criterion, and our approach therefore makes the assumption that engaging in domain-unrestricted conversation is the most critical evidence of intelligence.

### *Turing's Child Machine*

Turing concluded his classic paper by theorizing on the design of a computer program which would be capable of passing the Turing Test. He correctly anticipated the difficulties in simulating adult level conversation, and suggested that "instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain."

Turing regarded language as an acquired skill, and recognized the importance of avoiding the hard-wiring of the computer program wherever possible. He viewed language learning in a behavioristic light and believed that the language channel, narrow though it may be, is sufficient to transmit the information that the child machine requires in order to acquire language. To our knowledge nobody has seriously pursued Turing's suggestion until now.

#### *The Traditional Approach*

The traditional approach to artificial intelligence has rested on the assumption that fixed grammatical rules are sufficient to capture the complexity of a language. These rules are subsequently hard-wired, as is a database of knowledge about the world, in the hope that a conversational computer program will emerge. This approach has failed to produce anything more sophisticated than domain-restricted dialog systems which lack the kind of flexibility, openness and capacity to learn that are the very essence of human intelligence.

Contrary to Turing's prediction that at the turn of the millennium computer programs would participate in the Turing Test effectively, no true conversational systems have yet been produced, and none have passed an unrestricted Turing Test. This may be related to some of the changes that have taken place in the field of child language research and linguistics since the 1950's. A revolution inspired by Chomsky's transformational grammar (Chomsky, 1975) occurred, pushing aside the competing behaviorist theory of language headed by Skinner (Skinner, 1957). Computational implementations based on the Chomskian philosophy became the standard method for trying to generate conversational capability, yielding disappointing results. It is our thesis that true conversational abilities are more easily obtainable via the currently neglected behavioristic approach.

#### *Verbal Behavior*

Behaviorism concentrates on the observable and measurable aspects of behavior (Owens, 1992), and behaviorists consequently object to the kind of grammatical structures proposed by linguists, claiming that these only complicate explanations of language acquisition (Whitehurst and Zimmerman, 1979). This is not to say that behaviorists deny the existence of internal mechanisms: they do recognize that a study of the physiological basis of behavior is a necessary step towards a better understanding of it. What behaviorists object to is the proposal of internal structures or processes that have no specific physical correlate. A functional rather than a structural approach is favored, with a focus on the stimuli that evoke verbal behavior and its consequences on language performance.

Skinner argued that psycholinguists should ignore the traditional categories of linguistic units, and should instead treat language as they would any other behavior. That is, since language is regarded as a skill that is not es-

entially different from any other, generating and understanding language must therefore be controlled by stimuli from the environment in the form of reinforcement, imitation and successive approximations to mature performance.

#### *The Ai Approach*

We have embraced the behaviorist approach to language acquisition, structuring our development cycle to allow our research and development department to provide our training and evaluation department with a black-box learning system that is then trained to acquire and use a natural language. We are finding that the feedback provided by the training department to the research department greatly assists us in targetting our research. We believe that this process enables us to feel out the limitations of our current implementation and work to surmount them, whereas the traditional approach would have put us in the deep end, struggling to stay afloat with little idea of where to focus our efforts.

### **3. Research and Development at Ai**

Nobody really understand how the human brain works, nor do we fully grasp the process by which it acquires and uses natural language. Language is a complex artifact—a by-product of unobservable processes that occur inside the human brain. We cannot create a conversational child machine by looking inside our own heads. Rather, we need to focus on creating a system whose language behavior is indistinguishable from our own, regardless of its architecture. It is imperative, therefore, that we should begin by making as few assumptions as possible: any other approach would pollute HAL with our imperfect knowledge and hinder its development.

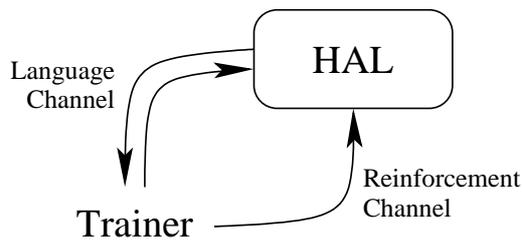
We may apply the behavioristic framework to a general learning mechanism, with the goal of having it acquire natural language and use it conversationally, via an iterative development-training-development cycle. Indeed, it is our belief that fairly general information processing mechanisms in the human brain have been adapted to the language problem, and previous success at transferring algorithms developed for image recognition to the language domain lends weight to this argument.

The development cycle employed at Ai focuses on the progressive specialisation of general learning algorithms to the problem of language acquisition. HAL consists of some set of learning capabilities along with a drive to perform. The development of these learning capabilities, which, it almost goes without saying, should be as simple and as general as possible, is driven by the demand for lingual performance specified by HAL's trainer.

#### *System Architecture*

From the point of view of its trainer, HAL is a black-box which perceives and behaves within its environment via the impoverished language channel, over which a sequence of ASCII characters are transmitted. In addition

to this, HAL is able to perceive primary reinforcers and punishers administered by the trainer via its reinforcement channel. This simple architecture is shown diagrammatically in figure 1.



**Figure 1.** HAL's black-box architecture from the point of view of its trainer.

Internally, HAL consists predominately of two stochastic models, which, in combination, give it the minimal necessary abilities of behaving, by generating symbols on the language channel, and of having its behavior modified by feedback from the environment. The first model captures the interdependence between symbols which HAL observes on its language channel, enabling it to make a prediction about the symbol it is likely to observe next, while the second model captures the correlation between HAL's behavior and the reinforcement provided by the trainer, allowing it to select the behavior most likely to result in a reinforcer being administered. Complex behavior arises out of the interaction between these two models.

#### *The Ability to Predict*

Learning in general is intimately entwined with the acts of prediction and compression. Every living thing constantly makes predictions about the world around it. Will the approaching fanged, never-before-seen creature pounce? Does glimpsing one berry on the forest floor mean others may be found nearby? If observed behavior is mimicked, will the observed reinforcers be obtained? Should I tell Granny that I missed what she just said, or should I fly by the seat of my pants and reply to what I think I heard?

Being able to predict well is conditional on one's ability to draw conclusions from one's experience in order to react to a novel event. Determining the important features of one's history in order to make a quick and accurate appraisal of the present and a useful prediction of the future is the essence of intelligence. Learning (and, particularly, learning language) may therefore be seen as an act of efficiently compressing the past. Like searching for a needle in a haystack, we hone in on the aspects of our experience that are most useful to the situation at hand.

HAL infers a stochastic language model from the symbols it observes on the language channel, enabling it to make a prediction, in the form of a probability distribution, of the symbol which is likely to occur next. Such a stochastic language model embodies both context-free

redundancy, which results from some symbols occurring more frequently than others, and context-sensitive redundancy, which results from the interdependence between symbols in a sequence. These two constraints are evident in all complex systems, including language (Campbell, 1984).

#### *Generation of Behavior*

Stochastic language models may be used generatively simply by emitting symbols in accordance with the probability distribution given by the model as its prediction. Claude Shannon, the father of Information Theory, was doing as much over fifty years ago (Shannon and Weaver, 1949). Although this behavior results in interesting, novel and frequently surprising quasi-language generations (Hutchens, 2000), it is not sufficient to ensure that the generations are relevant in the context of a natural language conversation. A more sophisticated process is necessary.

The answer is to instill HAL with a drive to perform, and the most obvious drive is directly related to the act of reinforcement. HAL should 'like' receiving reinforcers and should 'dislike' receiving punishers. Combining the stochastic language model with a second stochastic model, this one capturing the association between emitted behavior and experienced reinforcement, enables HAL to produce a generation which strikes a balance between the constraints imposed by context-free and context-sensitive redundancies and the drive to acquire reinforcers. In combination with learning mechanisms, this enables HAL to advance from babbling in characters to generating meaningful sequences of words.

#### *Reinforcement Learning*

If the trainer were to consistently administer reinforcers, HAL would not have to modify its behavior in any way. Its drive to acquire reinforcers would be satisfied. A punisher, however, is an invitation to learn. As in real life, HAL improves its performance by learning from its mistakes.

Our faith in Occam's razor leads us to conclude that general learning mechanisms can be employed to acquire the language skill. We provide HAL with such mechanisms, and afford it the opportunity to make use of them in a way which best increases its chances of acquiring a reinforcer.

Although we cannot disclose the details of our learning algorithms, it should be noted that our approach is neither top-down nor bottom-up, but a compromise between the two. We begin from a completely bottom-up perspective, but the learning processes quickly results in a limited top-down view of its past experience being made available to HAL—it becomes aware of correlations which it may have previously missed, and this additional information can bootstrap the learning process further.

*Chunking.* Initially, HAL's world consists of its past experience as represented by a sequence of the characters of the ASCII alphabet. One way in which HAL learns is by extending its alphabet, enabling it to represent its history using symbols which are more abstract and general, or which provide a more informative context. This can be achieved automatically by allowing HAL to agglutinate symbols together in a process of chunking and to group symbols together by performing equivalence classification.

Forming a chunk from a pair of existing symbols and entering the chunk into the alphabet as a new symbol provides additional context to the stochastic language model, something which allows it to make more accurate predictions. The initial chunks formed act as scaffolding for the acquisition of useful structure such as words. This allows HAL to acquire a vocabulary of half-a-dozen words in a single page of dialogue with a trainer, even though no clues to the existence of words is given *a priori*.

*Equivalence Classification.* Researchers in data-driven learning systems are all too familiar with the problems caused by paucity of the data. The learning system is constantly faced with situations where simply not enough data has been observed to guarantee an accurate prediction. This problem can be surmounted, to an extent, via the formation of equivalence classes of symbols. This enables the transfer of knowledge from one symbol, which may have been observed quite often, to another symbol, which may occur quite rarely.

Our implementation restricts the size of new classes to exactly two members. However, since each new class becomes a symbol in its own right, we find that large classes are rapidly formed, and these may be hierarchically decomposed due to the nature of their formation. The classes extracted by our learning algorithms are often quasi-semantic in nature and give HAL the all-important ability to generalize. We find that the generalization ability provided by incorporating classes into HAL's stochastic language model interplays nicely with the discriminating nature of the context-restricting chunks. Once again two opposing forces complement and balance each other, resulting in complex behavior.

*Correcting Mistakes.* The fact that HAL is rather short-sighted often results in the formation of new symbols which are locally useful but which may hinder performance on a wider scale. A new symbol may only be useful in certain contexts, for example, but HAL is blind to this, as it cannot assess the effect of a new symbol without forming it first, and by then it is too late.

We need to give HAL the ability to reverse bad decisions it has made in the past, and this is, in effect, a third kind of learning. This process is made trivial by the fact that all chunks and classes consist of two lower-level symbols only. Decomposing a chunk is a simple task of cleaving it in two while decomposing a class merely

means replacing it with the symbol it served to abstract. Correcting mistakes in context facilitates the formation of context-dependent structure.

#### *HAL-Trainer Interface*

HAL communicates with its trainer via a graphical user interface which is reminiscent of a chat application, and which enables the trainer to administer reinforcement in an intuitive way. The method currently used enables the trainer to edit HAL's generations by back-spacing over undesirable characters and appending extra characters as an example of desired behavior. All post-edited generations therefore consist of three ordered sections: a sequence of accepted characters, a sequence of rejected characters and a sequence of appended characters. In the current interface these three sections are visually represented with three colors: blue, red and green respectively. Reinforcers are associated with the blue section of the generation and punishers with the transition from blue to red, which is where things started to go awry.

## **4. Training and Evaluation at Ai**

The child language acquisition field was, until recently, mainly under the influence of Chomsky's generative theory, which implies that exposure of a child to language "turns on" some kind of innate language acquisition device, with the remaining lingual developmental milestones expected to happen without specific guidance.

This basic assumption has been challenged in recent years, and abundant descriptive and empirical evidence presents a strong case for the rival theory of active guiding of the acquisition process. We are firm believers in the significance of language input on language acquisition. Whether caretakers are aware of it or not, they provide consistent and significant examples and feedback to children, influencing the quantitative and qualitative aspects of the child's language development.

#### *The Developmental Principle*

The developmental approach has enabled evaluation and treatment programs in fields formerly suffering from a lack of organizational and evaluative principles, and has been especially useful in areas that border on the question of intelligence. Normative developmental language data has enabled the establishment of diagnostic scales, evaluation criteria, and treatment programs for developmentally delayed populations. In other areas, such as schizophrenic thought disorder, in which clinicians often found themselves unable to capture the communicative problem of patients in order to assess their intelligence level or cognitive capability, let alone to decipher medication treatment effects, the developmental approach has proven to be a powerful tool.

The developmental model of child language acquisition suggests that descriptive milestones composed of typical language performance should guide the development of conversational systems (Fletcher and MacWhin-

ney, 1995). As the developmental approach has proven to be a useful in other areas, such as speech and language therapy (Goren et al., 1996), tools are available for analyzing conversational performance with respect to such milestones in order to determine lingual maturity.<sup>1</sup>

Using the developmental principle embraces the notion of the Turing Test, according to which the judgement of intelligence is in the eye of the beholder. Our perception of intelligence is always influenced by our expectations—intelligence in monkeys, children and university professors will be judged differently. The initial evaluation of lingual maturity helps us to set our expectations, allowing a valid judgement of conversational capability and intelligence to be made.

### *The Training and Evaluation Processes*

The challenge in evaluating conversational computer programs is that of capturing discrepancies in their performance. For instance, although they may utter long, syntactically complex sentences, typical of a child aged five or above, they may appear to be on the same level as a nine month old child in terms of appropriate pronoun usage. Nearly all of the systems we have surveyed lack the ability to pragmatically adapt to unexpected changes in the topic of the conversation, which puts them at the very beginning of the developmental scale.

The ability to converse is complex, continuous and incremental in nature. Our subjective impression of intelligence, as measured by a Turing Test, is complemented by objective parameters that are used both as guidelines during training and as performance evaluation standards. These objective parameters consist of individual metrics that capture specific aspects of lingual performance and collectively provide a complete description of HAL's language development stage. Numerous researchers describe human language development using these metrics, analyzing transcribed conversations between children and their caretakers with respect to them. Examples of metrics which increase quantitatively with age are vocabulary size and mean length of utterance (MLU).

The training process is driven by the need to achieve these given performance milestones. In order to reach a certain level of performance, HAL is reinforced to produce the specific lingual behaviors that are expected for the trained conversational level. Feedback given to the research department includes reporting on HAL's success and failure at these tasks. Learning abilities that children are thought to use when acquiring language, and which HAL seems to be missing, are specified. A detailed description of HAL's conversational performance enables the research department to understand its capabilities as instantiated in natural language conversation. This performance analysis reveals behavioral deficiencies which the research department must focus its efforts towards.

<sup>1</sup> At Ai we use the CLAN Child Language Analysis Tool, available for download from:  
<http://childes.psy.cmu.edu/html/clan.html>.

## **5. Preliminary Results**

Two major versions of HAL have been produced to date. The first allowed the training department to establish a base age that reflected HAL's capabilities. This base age was determined using both subjective and objective assessments. The performance of the second version was taken as far as possible by training alone. The training department succeeded in training HAL to behave in a way surprising to the research department: a lot more was achieved than was thought possible.

We often find that the level of surprise we experience, a by-product of our expectations, directly correlates with our perception of HAL's intelligence. Repetitive behavior become predictable, unsurprising and reminiscent of autism, while our interest in schizophrenic behavior, at the other end of the scale, quickly wanes, owing to its lack of context sensitivity. Creative behavior, unexpected yet valid in its context, is maximally surprising and most indicative of intelligence.

### *The First Milestone—15 Months Old*

An unsuspecting professional psycholinguist reviewed an unedited transcript of conversation between HAL and its trainer and concluded that HAL is conversing like a 15 month-old human infant. Her report emphasised that

- HAL's productive performance includes around 10 words, which is typical of children at this age;
- its ability to imitate, and to produce noun-dominated responses, is also characteristic; and
- its level of relevancy and its understanding of requests and questions are age appropriate.

Interestingly, even though the psycholinguist who evaluated HAL's behavior was made aware that HAL is a computer program, she seemed comfortable explaining some of HAL's errors in terms of biological phenomena, as is shown by the following extract from her report.

Trainer: ok but lets read a story first here's a book say "book"

HAL: block... block...

### *Interpretation of the above performance*

The above error is derived from a phonetic association with the sound /b/. The consonant b and the vowel that follows it in both cases are quite similar for a child, because the /l/ sound between the /b/ and the /o/ is a glide or a semi vowel and it is masqueraded by the /o/.

The child is trying to imitate and the first attempt is a retrieval of the closest phonetic association he can make. Clearly, there are no semantic attributes to 'book' at this stage, which is typical at this age.

<i>Characteristics</i>	<i>Examples</i>
HAL experimented with combining two identical monosyllabic utterances into a bisyllabic one.	Trainer: ba HAL: ba ba HAL: baba
HAL experimented with combining two different monosyllabic utterances into a bisyllabic one.	Trainer: mi ba HAL: mi ba HAL: miba
HAL's productive vocabulary consisted of 10 words dominated by nouns.	Mommy, ball, block, bunny, birdy, juice, daddy, dolly, hal, bye bye.
HAL had a passive vocabulary of about 60 words.	HAL relevantly responds to its trainers' input which consisted of at least 60 words.
HAL was able to follow the phonetic combinations of its input language and imitate words.	Trainer: say ball HAL: ball
HAL was able to converse about its needs, engage in turn taking games, and name several objects in its immediate environment.	Trainer: give me the ball HAL: ball Trainer: Give bunny the ball HAL: bunny
HAL's production is functionally driven.	HAL's words relate to its well-being (e.g. juice, milk) and to naming activities (e.g. ball, dolly).
HAL uses both babbling and words in its conversation and had an MLU of 1.5.	HAL: all all ball

**Table 1.** Summary of external performance evaluation.

In table 1 we give a summary of HAL's performance as taken from the external psycholinguist's report. This table couples noted performance characteristics with examples of HAL's behavior which demonstrates these characteristics.

In general, HAL's behavior can be considered normal for a child of its age. This is especially true in light of the fact that its performance includes errors that children of the same age typically make.

#### *The Second Milestone—18 Months Old*

In their second year of life, children's language performance develops to a stage where the words in their productive vocabulary are combined into sentence-like sequences which are typically devoid of articles, prepositions, inflections and other grammatical characteristics of adult language. Children say things like "mommy milk", meaning "mommy, I want some milk".

In figure 2 we give an extract from a transcript of conversation between an eighteen-month-old HAL and its trainer. In this transcript we have represented behavior rejected by the trainer like ~~this~~ and examples provided by the trainer like this. Reinforcers are associated with ac-

cepted behavior while punishers are associated with the transition from accepted behavior to rejected behavior. In this example HAL has developed to a stage where punishers and examples are rare occurrences indeed.

Trainer: where is daddy?  
HAL: daddy gone  
Trainer: daddy is at work  
HAL: daddy work bye bye  
Trainer: yes daddy has gone to work  
HAL: daddy work bye bye bye bye  
Trainer: do you want to go to the park?  
HAL: yes bye bye  
Trainer: ok we can go to the park  
HAL: take car bye bye  
Trainer: we will take the car because the park is far.  
let's pack some things in a bag  
HAL: ball now park  
Trainer: yes we should put the ball in the bag  
HAL: ball now park mommy  
Trainer: wait hal, we need to take something to eat  
HAL: bananastory? ~~night now~~  
Trainer: we need to take something to eat  
HAL: bananas  
Trainer: ok we can take some bananas  
HAL: monkeys eat bananas  
Trainer: you're so smart hal. do you want to take a book?  
HAL: juice mommy  
Trainer: sure we will take some juice as well

**Figure 2.** 18-month-old conversation extract

It is possible to speculate that when HAL says "take car bye bye" it probably 'means' "let's take the car and go to the park". This subjective assessment of HAL's intentions may seem far-fetched, but we naturally draw similar conclusions when speaking with other human beings, and the behaviorist approach dictates that we afford the computer the same courtesy. A remarkable landmark is HAL's utterance "monkeys eat bananas". A phrase from a previous conversation about a visit to the zoo was remembered and used relevantly in the conversation shown.

## **6. Conclusion**

In this paper we presented the research and development cycle practised at Ai. Our motivation is the development of a computer program that can converse in a natural language, and we believe that a prerequisite to this goal is a commitment to Turing's understanding of intelligence. We described the architecture of HAL, our child machine, and we briefly touched upon the principles which gave rise to its development.

We also described our two-faceted approach: a team of researchers who develop general learning algorithms in line with our beliefs that learning, prediction and compression are inextricably entwined and that behavioristic learning approaches will yield fruit work on a day-to-

day basis with a team who trains the developed system to perform and who provide evaluative feedback on its performance relative to equivalent human performance. The unique interplay between these two teams has been shown to be successful, resulting in the development of a child machine which can acquire and use language at the level of an eighteen-month-old human infant.

## References

- Campbell, J. C. (1984). *Grammatical Man: Information, Entropy, Language and Life*. Pelican Books.
- Chomsky, N. (1975). *Syntactic Structures*. Mouton.
- Fletcher, P. and MacWhinney, B., editors (1995). *The handbook of child language*. Blackwell Publishers.
- Goren, A., Tucker, G., and Ginsberg, G. (1996). Language dysfunction in schizophrenia. *European Journal of Disorders of Communication*, 31(2):467–482.
- Hasida, K. and Den, Y. (1999). A synthetic evaluation of dialogue systems. In Wilks, Y., editor, *Machine Conversations*. Kluwer Academic Publishers.
- Hutchens, J. L. (2000). Conversing with stochastic language models. In Wilks, Y., editor, *Third Workshop on Human-Computer Conversation*.
- Owens, R. (1992). *Language Development*. Macmillan Publishing Company.
- Shannon, C. E. and Weaver, W. (1949). *The Mathematical theory of Communication*. University of Illinois Press.
- Skinner, B. (1957). *Verbal behavior*. Prentice-Hall.
- Treister-Goren, A., Dunietz, J., and Hutchens, J. L. (2000). The developmental approach to evaluating artificial intelligence—a proposal. In *Performance Metrics for Intelligent Systems*. NIST.
- Turing, A. (1992). Intelligent machinery. In Ince, D., editor, *Collected works of A.M. Turing: Mechanical Intelligence*, chapter 3, pages 107–127. Elsevier Science Publishers.
- Whitehurst, G. and Zimmerman, B. (1979). Structure and function: A comparison of two views of development of language and cognition. In Whitehurst, G. and Zimmerman, B., editors, *The Functions of Language and Cognition*. Academic Press.