Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
COMPUTER SCIENCE AND ENGINEERING
Ciclo XXXV

Settore Concorsuale:
09/H1 - SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI

Settore Scientifico Disciplinare:
ING-INF/05 - SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI

Computational Creativity: an interdisciplinary approach to sequential learning and creative generations

Presentata da: Mattia Barbaresi

Supervisore
Andrea Roli

Co-supervisore
Michele Lombardi

Coordinatore Dottorato
Ilaria Bartolini

Esame Finale Anno 2023

Abstract

Creativity seems mysterious; when we experience a creative spark, it is difficult to explain how we got that idea, and we often recall notions like "inspiration" and "intuition" when we try to explain the phenomenon. The fact that we are clueless about how a creative idea manifests itself does not necessarily imply that a scientific explanation cannot exist. We are unaware of how we perform certain tasks, such as biking or language understanding, but we have more and more computational techniques that can replicate and hopefully explain such activities. We should understand that every creative act is a fruit of experience, society, and culture. Nothing comes from nothing. Novel ideas are never utterly new; they stem from representations that are already in mind. Creativity involves establishing new relations between pieces of information we had already: then, the greater the knowledge, the greater the possibility of finding uncommon connections, and the more the potential to be creative. In this vein, a beneficial approach to a better understanding of creativity must include computational or mechanistic accounts of such inner procedures and the formation of the knowledge that enables such connections. That is the aim of Computational Creativity: to develop computational systems for emulating and studying creativity. Hence, this dissertation focuses on these two related research areas: discussing computational mechanisms to generate creative artifacts and describing some implicit cognitive processes that can form the basis for creative thoughts.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As a multidisciplinary research field of Artificial Intelligence (AI), Computational Creativity (CC) has two main goals: (i) to build computers capable of human-level creativity or to assist humans in this process, and (ii) to better understand human creativity, formulating an algorithmic perspective on creative behavior in humans. The former is mostly a common concern of AI, while the latter is the main aim of Computational Psychology or Cognitive Science. In CC research, the latter venue is poorly explored. In fact, within the CC perspective, the efforts to investigate creativity as a human phenomenon are focused more on theoretical accounts (or philosophical framework) and definitions of creativity, instead of investigating creative human mechanisms on a more empirical basis. Computational approaches are adopted by psychologists who build their computational models following the experiments they carry out (from disciplines like Computational Psychology or Cognitive Science, indeed). Moreover, research on creativity is hampered by two main issues in the field: (i) lack of standardized, shared assessment and (ii) the implicit dimension of creativity; people often cannot verbalize how they came up with their creative solution.

This thesis puts forward a computational approach, but from a cognitive standpoint, focusing on implicit aspects of creativity. It aims to provide insights into the interplay, strengths, and limitations of adopting both AI and cognitive approaches in creativity research. In doing so, several computational, and practical implications, are considered.

After surveying some of the definitions and dimensions of creativity, some computational mechanisms for creative generations are first explored. Then, a computational model for creative generations is proposed, that also addresses some human implicit learning mechanisms, discussing the cognitive aspects emulated by the model. Finally, the potential of this interdisciplinary approach is discussed, also introducing further possible explorations.

1.1 Problem Domain and Approach

The presented research is concerned with the modeling of cognitive and computational processes in the (implicit) **learning** and **generation** of **creative sequences**.

Within the computational perspective, the problem studied is one of sequential learning: the goal is to learn the hidden structure, or the rules, governing the ordered succession of discrete events in the sequence. The problem, in other words, is that of grammar induction. That is, learning a language from a set of examples. In cognitive psychology, the general phenomenon is also known as sequential behavior or behavior sequencing. There exist two broad categories of sequence learning: explicit and implicit. Recently, especially in language acquisition, it has been studied as a form of Implicit Statistical Learning (ISL)–providing also a compelling example of nonconscious learning of a complex cognitive task. Other forms of implicit sequence learning include motor sequence learning, temporal sequence learning, and associative sequence learning. This kind of learning involves mechanisms such as segmentation and chunking and, more importantly, the computation of Transitional Probabilities (TPs).

On the other hand, the problem of sequence generation (similar to prediction) is a sub-problem of sequence learning. The goal is to reconstruct, step-by-step, a sequence "in the way it naturally occurs". However, contrary to prediction, the problem of generation depends on the definition of the criteria used to steer the process itself. In the creative domain, defining these criteria is problematic because, again, it requires stating some explicit utility, thus establishing what is creative (and what is not).

To summarize, the emphasis of this computational creativity approach is not on performance or quality attained by the system (as in applied CC works) but rather on the explanation of the psychological processes leading to artificial and human creativity and the reproduction of behavioral data collected in psychology experiments. That is, we focused on computational and cognitive learning methods that allow constructing a representation of the sequence of events in the learner's memory of the form of a symbolic hierarchical network and on processes that, starting from this knowledge, produce creative sequences with the same intrinsic structure. To this end, this dissertation explores some mechanisms and possibilities to address these issues in both domains.

1.2 Context, Aims and Motivations

In the literature, the efforts to investigate creativity as a human phenomenon are focused more on theoretical accounts instead of investigating creative human mechanisms on a more empirical basis. This thesis starts within AI but extends towards cognitive science (or computational psychology) trying to achieve a more comprehensive and unitary perspective on creativity. To achieve this, we tried to address both the main aims of CC: (i) to explore computational techniques for simulating creative behavior or to build products (applied CC) and (ii) to model human creativity using computational mechanism (cognitive CC).

Specifically, this dissertation focuses on computational models for personal creativity (Boden's P-creativity [21]) and its implicit processes to build models both for brain mechanisms comprehension and enhanced robot behaviors in Human-robot and social interaction. The broader vision is towards embodied creative cognition (and cognitive robotics, in general) where computational accounts are necessary for mechanistic insights into the creative process. In doing this, the hope is to add to this gulf between the development of computational models of sequential learning in CC research and their application to the understanding of cognitive processing in implicit learning and creative generation. Therefore, the seminal idea is conceiving a modeling support for interdisciplinary studies, which has to be adaptable to various contexts, shareable, and explainable. The point we want to stress here is that computational models at this level are tools aiding research: more oriented toward open science (i.e., reproducibility, particularly in psychology) or educational perspectives, as well as toward more understandable human behaviors for AI, human-machine interaction and robot-related contexts. At the same time, computational models were born for robotics, dealing more with physical and environmental constraints for the realization, or reproduction, of the behavior of the intended model. Then, placed at this edge, computational models are the glue for connecting theories with real-world, embedded, experiments and their applications (see Figure 1.1).

Therefore, the general aim of this thesis is to give a computational account of ISL and CC (as the creative mechanisms stemming from implicit knowledge), emphasizing the importance of computational modeling in cognitive science and psychology [131, 63, 86]. We wanted to test the potential ISL has to learn structured information from languages. We made use of artificial grammars to analyze the potential of ISL in segmentation and chunking. In this respect, Artificial Grammar Learning (AGL) is a compelling case especially because it involves the tasks of rule-learning and language acquisition (i.e., sequence learning) which are fields where many of the debates and issues reflect those in cognitive science at large [234]. The other goal, indeed, is to test, in this way, the generality of such an approach to model various phenomena (namely, movements, music, language). As a general mechanism, and one of the most basic mechanisms of the brain, we believe ISL is the right approach to model many different phenomena, in particular for creativity. In

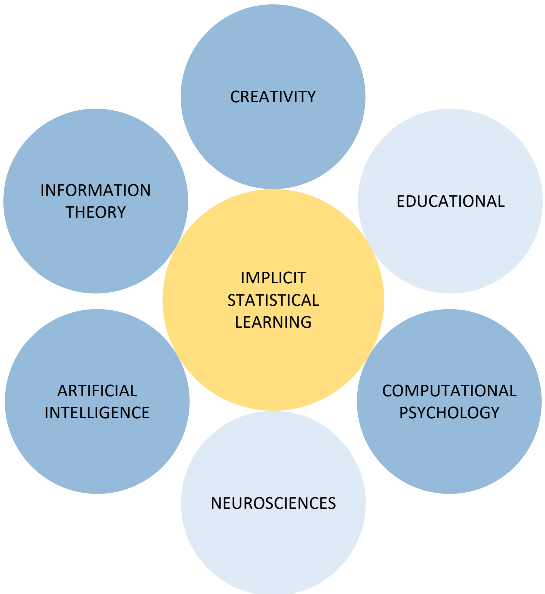


Figure 1.1: Areas affected by the ISL modeling in this dissertation (in blue) and followup explorations (light blue)

the end, the goal is to give an account of the implicit processes involved in creative productions. We opted for a symbolic approach because ISL is a general mechanism of the brain (with modality-specific peculiarities) that deals with transitions, so sequential modeling of events (or sequences in general). The symbolic approach allows the model to abstract away from the nature of the input and allows for hybridization with sub-symbolic approaches (such as that of transducers), leading to more grounded and sophisticated implementations.

In doing this, we also envisioned collecting results on the performances of ISL on various language complexities (used grammars in literature and for different modalities), reviewing some of the exploitable statistical cues, to test ISL potential (segmentation, chunking, and associative mechanisms) and to discuss some metrics for evaluating language complexities.

1.3 Research Questions and Objectives

The discussed context and motivations can be summarized by the following research questions.

RQ1 What are the basic computational mechanisms necessary for modeling personal creativity?

RQ2 What mechanisms are involved in the processes of human implicit learning and generation, within the task of encoding sequentiality (sequential learning)?

RQ3 What minimal, understandable framework, or approach, could fairly address both cognitive behavioral investigations and the development of a system for creative sequence generation with consequent computational simulations?

RQ4 How can a computational model account for various domains (e.g., Information Theory, Psychology, Neuroscience) to foster interdisciplinary and an ecological/holistic understanding of creative behavior?

In an effort of answering these questions, we also pursued some desiderata as useful design criteria, such as modularity, compositionality, and explainability, to favor interdisciplinarity. For the same reason, we decided to refrain from modeling explicit, domain-related, socio-cultural factors. Moreover, we decided to use simple functions to analyze the interplay between basic mechanisms; we wanted to address the problem at a computational/functional level rather than discussing the physical implementational details. In particular, we believe ISL is a good candidate to fulfill such criteria. In addition, ISL helps respond to RQ2 and provides the basis for RQ3 and RQ4.

To achieve the stated aims and to respond to the research questions, the idea is to conceive an unsupervised, incremental, simple, modular, open-ended, parameterized (to allow tuning to different modalities and contexts), understandable (representational transparency) model for creativity and learning, capable of generating robot choreographies, machine musical improvisation and even capable of emulating language implicit acquisition. This could be summarized by the following objectives:

- Investigating the role of value, novelty and surprise in creative generations
- Implementing a process-oriented, symbolic model for implicit sequence learning and generation
- Accounting human implicit learning mechanisms (ISL) to build the agent knowledge (or expertise)
- Designing a model, for enabling interdisciplinary approaches and dissemination for the mechanisms involved in implicit statistical learning and creative generation of sequences

1.4 Structure of the Thesis

After this introductory chapter, the remainder of this manuscript is organized as follows.

Chapter 2 introduces the main concepts and the theories around which this thesis revolves. In particular Computational Creativity and Implicit Statistical Learning as they form the theoretical background of this work. Then the last section briefly discusses some of the aspects involved with the computational modeling of cognitive functions. Specifically, it introduces Transitional Probabilities, as the central tenet of this work, the Markov Models used, and the motivations for adopting the symbolic approach.

Chapter 3 describes the carried research under a computational perspective and the applications to the conceived systems in movements and music. Some generative methods are discussed and some key computational mechanisms of creativity are explored.

Chapter 4 describes, on the other hand, the resulting cognitive model and the ISL mechanisms it reflects, which is the main contribution of this thesis. With a particular focus on segmentation, chunking, attention, generalization, and the generational procedure that exploits the learned knowledge for producing creative output.

Chapter 5 contains conclusions and discussions on the potential of this interdisciplinary approach, also introducing further possible explorations.

Chapter 2

Background

The content of this dissertation builds upon several topics studied by different disciplines other than AI and CC. In particular on sequential, grammar, implicit, and statistical learning (Computational Psychology, Cognitive Science). Therefore, this chapter briefly introduces the key aspects and the background theories that are useful for understanding the content of the thesis. First, the context and a brief historical background of Computational Creativity are introduced, summarizing the essential aspects and the characterizations that followed. Then, the principal tenet is presented, explaining the involved mechanisms. That is Implicit Statistical Learning: the paradigm upon which this work rests. Finally, the last section examines the role of Computational Models and AI tools: discussing some aspects of Information Theory, Markov Models, and the symbolic approach that helped to explore the cognitive phenomena of learning and generating sequences.

2.1 Computational Creativity

As a research topic, Computational Creativity is gaining ever more interest from various disciplines, especially in Cognitive Science, Robotics, and

Human-Machine Interaction[1] . However, the literature is vast. Therefore this section covers only the main concepts and those factors concerned with this dissertation. After a brief historical background, this section analysis those details useful for a working, viable definition. In particular, the characterizations explored and the theory for a process-based model of Creativity.

2.1.1 Historical Background

Creativity is a compelling but heterogeneous phenomenon. While the term is commonly used in everyday life, giving a precise definition of this concept is not trivial. In fact, how societies have perceived creativity has changed throughout history, as the term itself. In ancient Roman and Greek cultures, for example, creativity was assigned to Gods, Demons, or Muses, as Plato's thoughts on poetry summarize: *"The works of poets are entirely the invention of the Muses, who possess the poets and inspire them...Art could be beautiful only if it descended from God"*. In the 1550s, Giorgio Vasari wrote *"Lives of the Most Excellent Painters, Sculptors, and Architects"* Renaissance artists not as mere artisans, but as the individuals themselves who, with personal skills, experiences and characteristics, come up with new ideas. However, even when considered a human ability, Creativity has long been ascribed only to gifted people; attempts trying to demystify it started appearing rather recently.

Only in the 20th century the efforts to describe creativity became more scientific and methodological. In 1926, Wallace's model of thought provided the famous four stages of creative thinking: preparation, incubation, illumination, and verification. In the 1950s, Guilford [184] started researching Creativity as a scientific field, suggesting that it can be studied objectively. In 1961, Rhodes [178] introduced the concept of the 4Ps: Person, Process, Product, Press. Guilford, in 1967, drew the distinction between convergent and divergent thinking [87].

[1]As an example, see https://site.unibo.it/performingrobots/en or https://www.frontiersin.org/research-topics/14181/creativity-and-robotics

The advent of Computational Creativity is probably signed by the work of Boden [22, 21], in the 1990s. Starting from the premise that creativity is not a magical process, she devised some useful definitions and dimensions for creativity, sustaining that computers and AI have the potential to give us a much greater understanding of the human mind. She defined creativity as the ability to come up with ideas that are new, surprising, and valuable. She distinguished a psychological (or personal) and a historical dimension of creativity, highlighting the difference between what is novel for the individual and what is novel for the entire society, culture, or world. Furthermore, she introduced the notion of *conceptual space*, used to identify the three forms of creativity acting on it (combinational, exploratory, and transformational), described as the three roads to surprise, and she also addressed the key issue of creativity itself: that is, value has potentially infinite meanings and crucially, it changes over time. Since then (about the mid-1990s), interest in creativity from an AI perspective (that is, CC) has begun to blossom. All this led us to have a much better understanding of this phenomenon these days. Creativity is no longer some mystical gift that is beyond scientific study but rather something that can be investigated, simulated, and harnessed for the good of society.

Currently, CC studies are pushing the limits of AI: investigating concepts like autonomy and intentionality [226], and linking creativity also to emotion and cognition [85]. It has been studied from numerous perspectives, including philosophy, neuroscience, economics, and of course the arts. Nevertheless, unlike many phenomena in science, there is no single perspective, viewpoint, or standardized measurement technique. Creativity involves aesthetic, cultural, and subjective values. As an aspect of human intelligence in general, creativity is conceived as an individual's ability to generate novel alternatives that in turn may be used to solve problems and complete everyday tasks [39] (*little-c Creativity*), contrary to the creativity limited to the rare minds or geniuses, as was conceived in Plato's days (*big-C Creativity*).

As a result, we ended with a conspicuous number of conceptualizations.

The shared, relevant and fundamental traits of Creativity emerging from its various dimensions are those of *novelty*, *value* (and *surprise*). However, as said they involve multiple characterizations and interpretations; hence, the next section present in more details the considerations adopted throughout this thesis.

2.1.2 Domains of Creativity

Creativity research is broadly considered a low-consensus domain [208]. In this context, it is not clear who is assessing creativity: is it the creator itself, or a general, consensual agreement is required? While in a high-consensus domain, such as Mathematics, these two perspectives may correspond, this is not the case in the arts [24], for example. Hence, the most logical conclusion is to split creativity assessment into the personal and the collective, under the assumption that they may not correspond [206]: this discrepancy, for instance, is also highlighted by Boden's distinction of novelties (mentioned above). In this view, the agent starts with some subjective judgment based on their intrinsic motivations before looking for a consensual evaluation. Once the artifact or idea has survived personal assessment, then it has to deal with implementation, sociocultural factors, and domain-related constraints (such as communication, or audience reception and appreciation), which entail processes that are no longer related to individual creativity [136].

Another important distinction could be drawn, also, between implicit and explicit domains in the agent's mind. For instance, implicit processes are often thought to generate hypotheses that are explicitly assessed later on [96]: this argument falls under the broader subject of dual-process theories of thought and cognition [212, 94, 62]. Even if not new, in recent decades, this perspective of thinking about the role of these implicit forms giving rise to more structured, generalized, and even conscious knowledge has been increasingly studied [90, 61, 78, 77]. For instance, some empirical studies have shown that sleep promotes creativity, helping extract statistical regularities, solve problems [172], and restructure associative memories [120]. In this spirit, using generalization or analogy among learned behaviors is a notable source of creativity [116, 55, 99]. More importantly, the structure of the internal representation (e.g., the links connecting nodes in a network representation) strongly influences its ability to give rise to creative solutions [22]. From a computational perspective, Creativity and the creative mechanisms coming from explicit knowledge can be described accurately, and are available for introspection. So they lend themselves well to direct implementation; this is the realm of explicit theories and a fruitful scenario for applied CC. On the other hand, it has been suggested that there is a substrate of implicit processes which form the basis for cognition, thought, and creativity, where computational models are beneficial for algorithmic investigations [95, 34, 125].

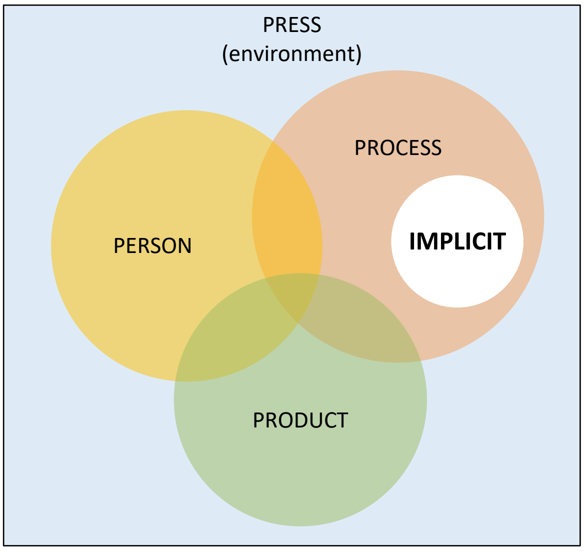


Figure 2.1: Area of investigation of this dissertation (white), using Rhodes's taxonomy.

This thesis focuses on the psychological (or personal) dimension of creativity and its involved implicit processes. Using Rodhes's taxonomy, this research deals with the Process dimension [178], see Figure 2.1, and more specifically with some of those processes that animate the intermediate stages of the creative process according to Wallas [186], those that do not require the conscious presence of the agent: such as Incubation, Intimation, and Illumination. In particular, the focus is on the role of novelty, value, and surprise, their divergent/convergent interplay, and the role of implicit knowledge and its generalization (as a form of knowledge restructuring in incubation, sleep or mind wandering). Hence, before going further, we need to clarify how these definitions are construed throughout this thesis.

2.1.3 Novelty, Value, and Surprise

Concerning Boden's definition [21] and broadly across the literature, the necessary factors suggested for creativity (and even more importantly for this thesis, for personal creativity [206])) are novelty, value, and surprise [208]. Still, these terms are often conflated with various meanings [190], hindering research in this field. Trying to favor cross-pollination and interdisciplinarity in this context, this section describes the practical connotations of these factors and the theoretical choices for modeling the interested processes.

Value The issue of value [91] is perhaps the most complex and vexing one in this context. While novelty and surprise are objective quantities, value is not. Value has potentially infinite meanings that depend on aesthetic, cultural, societal, and historical preferences; it has an ever-changing nature. As stated earlier, this issue hinders creativity research.

Even in this case, two dimensions can be distinguished for the utility: an internal, intrinsic one and an external, explicit one. In the former case, I'm talking of the utility of intrinsically motivated agents; in the latter, I'm referring to environmental or context utilities. In intrinsically motivated agents, for example, this utility comes from various theorized formulations such as reduction in uncertainty, reduction of free energy, or curiosity–related to the information gain obtained by the model for choosing that outcome. In this case, theoretical formulations often exploit Bayesian Models or Information Theory to estimate this quantity. This choice has theoretical and practical implications, which are not addressed in this thesis. The objective is to model implicit processes at a computational level, so we wanted to stop here, where implicit and explicit events start blending (e.g., the Illumination step). The point here, sufficient for our objectives, is just the presence of a utility. This issue is one of the first that needs to be explored with follow-up studies; i.e., what kind of utilities can guide the process at this implicit level? In particular, we want to explore another possibility: the interaction between implicit and explicit, factual personal knowledge (of the agent) in idea verification.

Novelty As stated, we focused on personal creativity, so the adopted definition of novelty herein falls under this category, contrary to historical novelty (see [21]). It is worth noting that we assumed that all creative sequences are composed of previously known elements; that is, no pure novelty can be generated. Without entering the philosophical debate [230], we assumed that *Ex nihilo nihil fit*. In addition to this, our focus is on "sequential novelty"; namely, we consider the novelty in the serial, sequential combination of units. So we can safely proceed behind this debate. However, throughout the thesis, various realizations have been investigated. In particular, two characterizations for novelty were explored:

- *dissimilarity*: dealing with symbolic sequence, a first incarnation of novelty is given by the inverse of string similarity. A similar version has been also conceived to evaluate the novelty of a sequence respect to a set of samples (as an average dissimilarity between the subject and all the elements in the set), and for integer arrays, where this metric translate into a Euclidean Norm.

- *originality*: a more cognitive interpretation of novelty, using a probabilistic, connectionist model (i.e., Markov Chains), is represented by the inverse of the probability that sequence has to be generated by the model (what Simonton called *the response strength*)

It is worth noting the difference here. In the first case, novelty represents something different from the input material and distinct from the other generated artifacts as a form of diversity concerning the material (of the input corpus or generated). In the second case–when ISL is introduced and so the personal, probabilistic knowledge–novelty assumes the meaning of originality concerning the learned internal knowledge of the individual; that is, it is based on the "capacity" of the individual to generate that outcome.

Surprise If within the two-factor definition surprise is not present (or often absorbed within novelty definition) [190], for a personal account of creativity, we need three factors, and we need to distinguish surprise from novelty as we will see. Surprise comes into play when we introduce ISL and the agent's knowledge. However, in this framing (that of Simonton), surprise needs utility. Surprise is fundamentally different from novelty; it is specified as the variation in beliefs about the utility [206]. In other words, the uncertainty the agent has on the utility (tested later on) of that outcome; that is, it measures the effects of knowing the value of the produced artifacts.

2.1.4 A Process Theory of Personal Creativity

One of the most compelling theoretical approaches for modeling personal creativity has been put forth by Simonton, working upon Campbell's Blind Variation and Selective Retention (BVSR) theory [204, 29]. The BVSR theory is a two-step process–as much as the convergent/divergent[2] thinking [87]–the first step is that of blind or unsighted generation of idea alternatives; the second stage is that of selecting and retaining the most useful ones. Simonton proposed a blind-sighted continuum along which creative ideas may lie [207]. He posits that creativity is something that is not only novel and useful, but it should also involve an element of surprise. In other words, blind means unsighted or unplanned or something which is not expected beforehand. Namely, something that surprises us with a meaningful result. Thus blindness exists to ensure surprisability; there is no creativity if we already

[2] two basic opposed mechanisms for exploring and reducing/selecting solutions

know the outcome. Variation means recombination or transformations: it is the mechanism for generating new and novel variants. That is, it serves novelty or originality. Selection necessitates utility. It means separating suitable and meaningful ideas from those that are novel but inappropriate or not useful, based on either internal/external or subjective/objective criteria. Finally, retention means developing, validating, and implementing the selected idea. More importantly, while the first three elements may work unconsciously, retention deals with domain-related, specific, and explicit, conscious actions, so it is not relevant to the creative process and, as said earlier, to this thesis.

Several criticisms have arisen about creativity as BVSR: a debated issue, for example, regards whether creativity is a blind or sighted process, e.g., contrasting hindsight or foresight ([229, 74] are only few examples). Centrally here, these scholars have suggested that although individuals cannot foresee the entirety of their final ideas at the start of the creative process, they may have a rough vision for what their initial ideas could become, enabling them to predict the potential creativity of their initial solutions with some degree of accuracy [16].

Within the current approach, we give an interpretation that could shed light on this argument, showing that blind (at least in our case) means "implicitly guided": namely, it involves randomness guided by implicit experience.

2.2 Implicit Statistical Learning

Implicit Statistical Learning (ISL) refers to the general, implicit and ubiquitous ability of the brain to encode temporal and sequential phenomena, and more generally, to grasp the regularities in the environment, in an autonomous and unsupervised way, often without awareness [198]. This approach results from the recent attempt to unify two research venues in psychology and cognitive science, namely Implicit Learning (IL) and Statistical Learning (SL) [160, 32], manifesting the existence of various interpretation for the same set of phenomena [159]. Implicit Learning refers more in general to mechanisms and knowledge, in the brain, that are unconscious [172]. Statistical Learning, on the other hand, was initially introduced for language acquisition, and it is now invoked in various domains of psychology and neuroscience to account for the human ability to detect and use statistical regularities present in the environment [188]. Importantly, the central tenet of SL involves the detection of transitional probabilities (TPs). Seminal experiments, exploring this phenomenon in the acquisition of spoken language, showed that infants are sensitive to TPs of syllables in a continuous speech stream [187]. It has been shown also that these mechanisms could be account not only in developmental phases of children, but also for large-scale, more real languages [67]. Furthermore, it is not specific to a domain or modality only, contrarily evidence has revealed similarity and differences in SL across the senses [37]. Moreover, SL is studied over species too [189]: many animals are sensitive to distributional statistics, which suggests that learning from distributional statistics is a domain-general ability rather than a language-specific one [5]. A number of studies have suggested that these are basic, robust, and general mechanisms important especially in development and rule-learning [36], social interactions [70] and even for conscious awareness [61, 84].

By now ISL is considered a cornerstone of cognition [103] and it has become an important building block of virtually all current theories of information processing, encompassing a complex suite of computations [71]. It has been recently suggested that incidental and automatic learning of temporal transitions between adjacent regularities does not depend on the use of prediction errors; instead, it may be a direct function of the amount of exposure, taking place at any time in our life in a variety of our daily activities (e.g., motor learning, social interactions) [146]. Moreover, it seems that the independence from prediction errors enables learning of additional contingencies which might otherwise not be learned. At the computational level, such learning mechanism is compatible with chunking models of statistical learning [162], which may be implemented in functionally specific areas [36].

More remarkably, it has been suggested that improvisational musical creativity is mainly formed by implicit knowledge. The brain models music—and other sequential phenomena such as language or movements—as a hierarchy of dynamical systems encoding probability distributions and complexity [42]. ISL also plays a role in the production of sequences (e.g. notes or actions); from a psychological perspective, transitional probability distributions sampled from music may refer to the characteristics of a composer's implicit knowledge: a high-probability transition may be one that a composer is more likely to predict and choose, compared to a low-probability transition corresponding more to an unusual variation [45].

Even more important to what is being discussed, creativity is mainly formed by implicit knowledge [233] and the effects of ISL are related also to acquisition of expertise in music [163] and movements [153].

Inspired by these extensive literature, which reflects the pervasiveness of ISL, especially regarding computational approaches like [133], we adopted this framework for emulating implicit sequential learning and creativity. Another motivation is that, although some models of ISL are present in literature (e.g., PARSER [162], SRN [57], or TRACX [68]), at present there is little experimental evidence regarding learning performance of complex streams, and what the underlying mechanisms and computations for such learning might be [71].

Perceptual and Semantic Chunking The main issue of ISL revolves around the formation of chunks, or, more extensively, cognitive units [159]. However, the literature on chunking comprises diverse areas of research. As a consequence, the concept of chunk has different meanings [79]. A number of works argue that there is a basic perceptual or sensorimotor chunking that groups sequential stimuli during sequence processing [80, 216]. This differs from the traditional Miller's notion of chunking [138], central concept of cognitive psychology, which involves a conceptual or semantic re-coding of information. It is essentially defined as a strategy to enhance memory by grouping items in terms of varying semantic attributes. Perceptual chunking, on the other hand, is an automatic perceptual process that is domain-general and that creates groups in sequential stimuli. Such grouping is commonly observed in sequence learning tasks [76]. For the sake of clarity, in this thesis we always refer to perceptual chunking in serial processing, unless explicitly stated.

2.2.1 Chunking and Segmentation

As mentioned, ISL comes from the union of two research venues that favor different interpretations for chunk formation: IL focuses on the selection of chunks meanwhile SL focuses on the computation of transitional probabilities aimed at discovering chunk boundaries. On the one hand, chunk learning has been widely considered to drive the implicit acquisition of cognitive and motor skills [80], suggesting that implicit learning can result from chunk learning–where a sequence is perceived as short segments, and the concatenation of these segments leads to the acquisition of the sequence. On the other hand, implicit sequence learning can arise from statistical learning: that is, the ability to extract the statistical regularities of sensory input across time or space [12]. In literature, the other two terms used to explain the difference between these approaches are clustering and bracketing. Clustering mechanisms rely upon merging high-frequency co-occurring elements as a single unit; bracketing (or boundary-finding) mechanisms work by identifying low-frequency co-occurring elements that correspond to the boundaries between discrete units [170]. Nevertheless, both approaches result in the acquisition of unsegmented input streams. Throughout this thesis, we generally refer to them as chunking and segmentation: by referring to the memorization of the joined units and the statistical boundary discovery mechanism used to discern them, respectively.

In any case, it remains unclear what are the roles of these mechanisms that underlie this type of learning and how they interact. [53] suggested that the initial acquisition of implicit sequences may arise from (first-order) statistical learning rather than chunk learning. Similarly, [1] demonstrated that statistical learning and chunking are two successive stages in implicit sequence learning with chunks inferred from prior statistical computations. However, in [160], the authors described three possible scenarios. The first case is that statistical computations and chunk formation are independent processes (with the premise that chunk formation is responsible for conscious knowledge, and statistical computation for improved performance in implicit tasks [137, 103, 102]). The second possibility is that statistical computations and chunk formation are two successive steps in the learning process. Chunks would derive from prior statistical computations. Typically, chunk boundaries are defined as the points where the predictability of successive, or spatially contiguous, elements is the lowest. The third possibility is that chunking is the only effective process and that sensitivity to the statistical structure is its byproduct.

Nevertheless, there is another possibility. Namely that the computation of statistics and chunking are parallel components, working together during learning. An example of this is the Chunk-Based Learner (CBL [133]) model.

2.2.2 Attentional and Memory Mechanisms

Perhaps the most paradigmatic example of chunking is represented by PARSER [162]. This model shows how the sensitivity to statistical regularities can arise from the extraction of chunks. In particular, how parsing emerges as a consequence of the attentional processing of the input and essential laws of memory and associative learning; initial segmentation is supposed to depend on various cognitive-perceptual factors (such as prior experience, vigilance, or saliency) but not on the computation of transitional probabilities, as instead has been suggested since Saffran's studies [187]. This method exploits a memory-based approach in which chunks emerge from the interplay between attention, decay, and interference. In particular, correct units originate from attention and decay, and the sensibility to (forward and backward) TPs arises from interference. In conceiving PARSER, authors have suggested that correct parsing emerges as a consequence of the organization of the cognitive system. That is due to the interplay of two principles: (i) perception shapes internal representations, and (ii) internal representations guide perception. The result is that perception builds internal representations that, in turn, guide subsequent perception. The relevant units emerge through a natural selection process because forgetting and interference lead the human processing system to select the repeated units among all the other parts initially generated by the chunking of the material but no longer encountered. The second principle ensures the convergence of the process toward an optimal parsing solution.

2.2.3 Generalization and Abstraction

Another compelling aspect originating from the discussions around ISL regards the extension of learned statistical structure to unseen stimuli. In the literature, it is mentioned as the generalization, the abstraction, or the transfer of learned knowledge [3]. A distinction across the three terms, however, exists. While the broader term "transfer" refers more to the application of learned patterns to novel domains and modalities [72], the difference between 'abstraction' and 'generalization' is more subtle. Often abstraction is used to refer to two different but often related concepts. On the one end, it describes the process by which individual details of an episode are lost (resulting in representations separated from the underlying perceived elements). On the other hand, it represents the accumulation of information across individual learning episodes that gives rise to knowledge that, although not necessarily contained within any episode, captures (statistical) regularities across those episodes. In [218] it is called "integration", and is studied, for example, in humans during sleep [83, 101]. Again, for clarity, we use the term "generalization" throughout this dissertation as we discuss mechanisms more similar to the latter case.

Thus, generalization in ISL is often driven by local stimulus properties and similarity: rather than by the extraction of abstract rules [72]. Primarily in language acquisition [188, 8], but also elsewhere [116], scholars have suggested that statistical computations are exploited for generalization purposes. In particular, using distributional statistics, learners could grasp higher-order relations that lead to the discovery of categories (e.g., form classes in language acquisition [219, 7]). In [8], the authors showed that learners' tendency to generalize depended on the degree of overlap among word contexts in the input: the consistency of contextual cues leads learners to generalize rules to novel strings; their inconsistency keeps learners from generalizing and thus treats some strings as exceptions. The intriguing point put forward in these studies is that, from this viewpoint, statistical learning and rule learning are not different mechanisms [154]; that is, distributional statistics (i.e., how the context cues are distributed across the input) determine whether forming a rule or learning specific instances [6, 174]. According to these hypotheses, statistical learning is a single mechanism whose outcome applies either to elements or to generalization beyond them, and importantly this trade-off is accomplished without instruction, depending on the structure of the input itself.

2.3 Computational and Cognitive Models

This section briefly discusses some aspects involved with the computational modeling of cognitive functions. In particular, the importance of computational models for cognitive studies [49], the roles of Statistical Learning and Transition Probabilities–as the central tenets of this thesis–and the motivations for the symbolic approach adopted.

Marr's (three levels of) analysis already highlighted the importance of computational models for cognitive science [132]. For Marr, referring to computational or "functional" means addressing *what* a complex system is doing at a higher level to find out *how* then, at lower levels, it accomplishes that task. In other words, we often need to know the function of the analyzed system to understand what aspects to look at in the lower implementational or physical details (neural or other). In this view, computational models offer two main advantages to cognitive science. First, they force the researcher to thoroughly specify all the required assumptions: thus, including all of the otherwise overlooked aspects. Secondly, they can produce the same sort of data as human subjects. Namely, the same statistical analysis can be conducted on both the observed data and the data from the model. These are salient features for cognitive science [213] and can be attained only if the functional architecture of the model and operations are readable, interpretable, and transparent.

More importantly, computational approaches help demystify human creativity favoring insights into the underlying mechanisms and their characteristics [134]. Furthermore, as stated in [34], they play a central role also in the investigation of implicit learning paradigms: (i) by making predictions comparable between models; and (ii) by making it possible to explore how specific mechanistic principles can offer unitary accounts of the data.

Additionally, in these respects, the use of robot systems introduces phenotypic and environmental constraints similar to those that living beings have to deal with; this is in line with modern brain theories [2, 164], which emphasize the importance of closing the perception-action loop between the agent and the environment [182]. Conversely, on the AI side, these kinds of embodied cognitive approaches could favor the study of symbol emergence in robotics [215, 214].

In conclusion, to build socially intelligent robots and favor the mutual pollination between AI and cognitive science, joining forces across these disciplines is paramount [232] and computational models are the appropriate means to do this.

2.3.1 Transitional Probabilities and Markov Models

In ISL literature, Transitional Probabilities (TPs) represent a type of sequential statistics found between words' parts (e.g., syllables) that predict the occurrence of the next (or the previous) unit [92]. Thus, the probability that one syllable followed another within a word or phrase. Initially, by using the term Statistical Learning (SL) to describe this phenomenon, Saffran and colleagues [187] showed that infants were able to compute these probabilities in a continuous stream (generated by an artificial grammar) to discover word boundaries. The gist is that artificial language, as natural languages, are statistically structured [165]: the TPs between syllables composing a given word (like "on-ce" or "up-on") were higher than those overlapping two words (like "ce-up"), hence making TPs a reliable cue for sequence segmentation.

TPs are the conditional probabilities of the next element given the preceding element, like the transition probabilities investigated by Markov that, in turn, are rooted in the context of Information Theory [128][3]. Due to their sequential nature, Markov Models (MMs) are well fit to describe the succession of events in a sequence (e.g., notes, moves, words). Throughout this dissertation, we explored mainly high-order and variable-order Markov chains: so we refer to them generally as Markov Models, also referred to as n-gram models in Computational linguistics [26]. In representing transitional probabilities, they lend themselves well to modeling ISL and sequential phenomena. Markov chains are easy to analyze/understand (they eliminate the label bias problem, for example) and their underlying theory is sound and elegant. They can also form the basis for more sophisticated frameworks (Bayesian analysis/inference, for example). In addition, they are generative and powerful inference tools. As a result, plenty of works used MMs in music [197] and many other sequential domains. Furthermore, in MMs, the same statistical principles are used both for producing sentences and for evaluating the probability that a given sentence could have been produced by the model. That is an essential capability for language models [117], for instance.

[3]Shannon made great use of Markov chains. However, he went beyond Markov's work with his information theory application.

2.3.2 The Symbolic Approach

The ongoing debate between symbolic and connectionist approaches represented one of the fundamental discussions in AI: see [81] for some recent, related considerations. Importantly, there was a misleading perceived association between symbolic rule processing with conscious mental processing, on the one hand, and the computations of NNs (or other sub-symbolic approaches) with intuitive/nonconscious cognitive processes on the other [123]. Hence, this section motivates the choice of using a mostly-symbolic approach. However, it is worth noting that we made no strong assumptions on this, aiming at making the least possible number of hypotheses. On the contrary, we believe these two paradigms are complementary: in this sense, the recent developments of hybrid approaches (e.g., neuro-symbolic AI or Markov Graphs [98]) favor this perspective. For instance, connectionist approaches (e.g., autoencoders or transformers) can be exploited at the perceptual level for elaborating the signal coming from different sensors (e.g., mapping and classification). Then, a symbolic model such as the present one could be employed to investigate the relations between those classified signals coming from the lower sub-symbolic level.

In addition, the symbolic approach fits well with this specific context (and aims) since chunking arises naturally within symbolic models of cognition, where atomic elements of information are combined into single units [79].

Furthermore, using symbolic models for cognitive tasks brings other generic advantages. They have clear syntactic structures, the results are easy to understand, and they can be represented explicitly as production rules. All this allows for further additions of processes and integrations of the acquired knowledge with other (implicit and explicit) knowledge and modules [124].

On the other hand, artificial neural networks require parameters to be set, including the network structure, the initial weights, the number of hidden units, the number of layers, the learning rate, or when to stop training; this strategy intrinsically involves making assumptions. In addition, the trained networks are usually hard to understand. It is also hard to imagine how the acquired knowledge, stored in weights, can be integrated into higher-level modules. Within a symbolic approach, incremental, unsupervised, cognitive approach instead, the parameters used in the model are those variables that have to be probed with followups experiments of the model (e.g., parameters for decay, interference, propensity to generalize) [123]

Moreover, the statistical learning mechanism modeled throughout this thesis is intrinsically ubiquitous in the brain, and domain-general, working on different kinds of information at various levels, or areas, of the hierarchy of brain processes [12]. These features make one refrain from using NNs for their lack of modularity and generality.

Chapter 3

Computational Generations

As stated in the introduction, the most general interpretation of creativity involves at least two main traits: *novelty* and *value*. Nonetheless, these two terms contain various meanings and characterizations: the former involves unexpectedness, unpredictability, surprise, and originality meanwhile the latter stands for usefulness, quality, appropriateness, and meaningfulness [190]. For instance, similarity is not the same as quality, usefulness, or appropriateness (it is the actual use of the artifacts that validates its usefulness), and novelty is not the same as surprise (the former involves artifacts, the latter is related to the individual).

This is an important point here. Keeping in mind that we want to model personal and implicit creative processes, we need to distinguish between what comes before and what comes after the actual assessment, what happens inside the agent and what happens externally, and what are the different driving forces in these steps. Concerning personal creativity (P-creativity), the actual usefulness and quality of an artifact are assessed after the agent had come up with that brilliant but drafted idea, and implemented it, putting effort into its realization: thus, involving processes that have nothing to do with creativity. Moreover, this kind of quality is often judged externally, by other agents (for H-creativity). In addition, even within the personal realm of creativity, there is at least another distinction: namely, the difference between implicit and explicit (processes and) knowledge. An agent can use its explicit, factual knowledge to (mentally) validate hypotheses, however, the illumination (the insight, the "ah-ha" moment) comes out in a flash just because an idea was generated implicitly, without explicit awareness of the agent, and so the forces driving the internal, implicit processes must have nothing to do with external or explicit utility. This point reflects the approach and motivations of the dissertation: in fact, our last effort ends where explicit or external utility must be invoked to reach applied, practical creativity (see the next chapter). For these reasons, we decided to employ similarity and novelty as domain-general, utility-independent forces for simulating convergent and divergent behaviors, respectively.

Within these assumptions, the first studies were carried out with a focus on computational aspects to investigate the role of some of these fundamental mechanisms for CC. The objective was to explore essential dynamics rising by the interplay of the mentioned basic components of creativity. The roles that deserve particular attention are those of: (i) similarity and novelty in the divergent/convergent process, as they form two core processing states in creative thinking and problem-solving; (ii) TPs at different orders, for investigating ISL; (iii) generalization, as an important source of creativity. We studied these factors exploring some CC applications in artistic contexts where creativity and implicitness are pervasive: (robot) movements (or choreographies) and (artificial) music.

To summarize, adopting a purely exploratory, algorithmic perspective, we conducted some simulations on computational methods for generating (symbolic) creative sequences in evolutionary art settings and then built the first experiments on the relations between novelty and similarity. Afterward, we exploited statistical information to enhance both similarity assessment and the generative procedure and to explore the possibility of an adaptive novelty search for this framing. Finally, a generalization step was introduced, acting on the learned statistical model, and investigated its effects on produced sequences. These computational investigations produced a series of algorithms for sequence generation (NohGenerator[1], NohGen2[2], NohGen3[3]) and the essential elements to conceive the psychologically-inspired model of the next chapter.

In the following, the computational steps and the results of these experiments are discussed. The cognitive model, and its psychological-related features, will be presented in the next chapter.

3.1 Robot Moves: Similarity and Novelty

We started investigating the relationship between similarity and novelty, trying to comprehend their functions in the creative process and their distinctions from the perspective of the individual and implicit processes involved. With these assumptions, we decided as an interdisciplinary research group to explore the generation of robot choreographies for the Nō theatre, a traditional form of classical Japanese musical drama. We conceived a novel fitness function for evolutionary art, which generates sequences of movements—i.e., robot choreographies—based on similarity to an inspiring repertoire. The convergent behavior induced by similarity is, in turn, counterbalanced by a novelty mechanism, which makes it possible to sample unexplored areas of the choreography space. Part of this work was published in [9]. In the following, this approach is discussed together with the achieved results.

3.1.1 Materials and Methods

Genetic algorithms are powerful tools for searching in unknown problem spaces. According to BVSR principles of creativity, creative ideas must be generated without full prior knowledge of their utility values [205, 206]. Herein, genetic algorithms offer a natural setting for the blinded-divergent and selective-convergent mechanisms involved in the creative process, terms

[1] `https://github.com/mattia-barbaresi/NohGenerator`
[2] `https://github.com/mattia-barbaresi/NohGen2`
[3] `https://github.com/mattia-barbaresi/NohGen3`

of diversification and intensification [19]. The evolutionary approach is itself an exploratory process: the combination of two individuals from the population pool is a combinational process, but the use of a fitness function guides the exploration toward promising areas of the conceptual space, which is bounded and defined by the genetic encoding of the individuals. Losing the fitness function, or having one unable to guide the exploration effectively, reverts the mechanism to pure combinational creativity, where elements of the conceptual space are joined and mutated, hoping to find interesting unexplored combinations. We conceived a genetic algorithm (GA) which, starting from a given inspiring repertoire and a set of unitary moves, generates symbolic sequences of movements (i.e., choreographies) exploiting similarity with the repertoire combined with the novelty search approach [228]. The idea was to start from a simple, basic system that includes primitive but essential mechanisms for creative generation, then increasingly add features that enhance the creative potential of the generative system to evaluate the effects these modifications have on the metrics used for the assessment.

To do this, we conceive a fitness function that evaluates the average similarity of an individual from the samples in the inspiring set. The entire generation process is guided by the novelty search approach described in [228]. These choreographies are performed by a (virtual) Nao robot[4] using CoppeliaSim[5] environment.

Encoding

We represent a choreography as a sequence of basic moves (i.e., figures or poses). Each figure is stored as a single keyframe (from motion capture data) that specifies the position of each captured joint and is encoded as a symbol. Therefore, in our representation, choreographies are sequences of symbols. Hence, the GA generates strings (symbol sequences) mapped to keyframe series (sequences of joint angle sets). The actual movement of the robot is

[4]Softbank Robotics: `http://doc.aldebaran.com/2-1/home_nao.html`
[5]Coppelia Robotics: `https://www.coppeliarobotics.com`

produced by the transition from one keyframe to another. Only steady poses are encoded, and we assume that the transition between stable moves is also stable, taking advantage of the typical gentle style of Nō movements. This, in general, represents a limitation for robot movements in real applications (e.g., some moves or transitions could make the robot fall) but runs well for our purpose.

Inspiring Set and Similarity

In general—but especially when dealing with artistic problem modelization—it is not trivial to conceive a (fitness) function that fully captures all the problem objectives. What is valuable? And how a choreography can be judged as more valuable than another? Note that in this case, we use the term *value* as a synonym of *typicality* or *membership*. The value, the quality, of an artifact will usually be assessed by other agents' (humans) judgment and will be based on cultural experience and knowledge, and hence are likely to reflect historical comparisons of the artifacts. These aspects refer more to the H-creativity [22] and will be addressed in future work. To model this aspect in this first attempt, the idea is to take a sample repertoire and evaluate each individual through its similarity to such a given set. The inspiring set represents choreographies that the robot, or the dancer, knows about a particular style or repertoire, which is the knowledge from which the artist pursues its style and creates new pieces. As a first step, we created the repertoire by manually encoding typical sequences of moves (*kata*) that have been performed by a Nō expert.[6]

Therefore, the fitness function evaluates sequences using string similarity between individuals and this repertoire. In formulas:

$$fitness(x) = \frac{1}{card} \sum_{n=1}^{card} similarity(repertoire[n], x) \tag{3.1}$$

$$similarity(x, y) = 1 - \frac{JW(x, y) + Jacc(x, y)}{2} \tag{3.2}$$

[6] https://site.unibo.it/performingrobots/en/project/activities

where $card = |repertoire|$ is the cardinality of the repertoire (i.e., the number of sequences forming the repertoire), $repertoire[n]$ is the nth element of that inspiring set and where the similarity is formulated as the combination of Jaro-Winkler and Jaccard [118] string distances, respectively $JW(x, y)$ and $Jacc(x, y)$ in equation 3.2. Similarity can be thought of as the convergent process in the generation that constrains the resemblance of generated artefact to given samples.

Novelty Search

Convergent/divergent thinking is a characteristic of creativity; to model these dynamics, we include a novelty search mechanism that guides the generation in the opposite way with respect to similarity. The novelty mechanism is inspired by the work of Vinhas et al. [228]. The main goal of this algorithm is to generate a more diverse set of individuals than the set that would be created by a traditional fitness based evolutionary algorithm. The method evolves individuals according to two criteria: *(i)* look for the best individuals according to the fitness function and *(ii)* take novelty and fitness as two different objectives to be maximised using a Pareto optimization. This bi-objective optimization is performed considering the number of individuals of the current generation that have a fitness above a given threshold. It uses also an archive to store the most novel individuals. The novelty score is then calculated as

$$novelty(x) = \frac{1}{k} \sum_{j=1}^{k} dissim(x, j) \tag{3.3}$$

where $dissim(x, j)$ is a dissimilarity metric, between the individual being evaluated and a set of k neighbours chosen from the population and the archive. We implemented the dissimilarity metric by complementing the similarity function.

3.1.2 Results

We tested the algorithm by varying several parameters, among which the size of the repertoire, and we compared the results of the combined use of similarity and novelty with simple fitness-based GA results and random search—which may represent a naive way of producing a divergent process.

As expected, the overall similarity slightly decreases with the size of the repertoire, meaning that the search process achieves a similarity balance between the generated individuals and the ones in the repertoire. The drift imposed by novelty turns out to be effective, as the best choreographies generated are usually sensibly different among themselves, considerably more diverse than those generated at random. Some samples of the choreographies can be found here: `https://github.com/ste93/NaoNohVideos`.

3.1.3 Discussion

In this section, we examined the role of novelty and similarity in a divergent-convergent thinking scenario as the driving factors for sequence production. We tried to tackle the problem of using metrics that are general enough but effective for this task. This strategy is fruitful since GAs, combined with a symbolic approach, provided a good testbed for these explorations: there are no learning-related problems since the procedure works in a batch mode, and the mechanisms for variation it uses are domain-general (string mutation and recombination).

The similarity with the repertoire, seen as typicality or membership, ensures a certain grade of convergence toward input samples, and novelty, as originality, is the explorative/divergent drive for creativity, more than random that has no direction; novelty aims at finding dissimilar samples from the known and selected ones. This type of force operates as the (learning-oriented) case of curiosity in literature: to explore unknown, non-conventional alternatives.

Currently, we are performing a quantitative assessment of the choreographies by applying information theory and complexity measures. These metrics make it possible to capture relevant features of the choreographies that can be used both for identifying their peculiarities and differences (and possibly comparing the fingerprint provided by the metrics with human evaluation) and derive some heuristics for improving the generation process. This approach is analogous to that recently proposed for swarm robotics [181]. Among the metrics we are currently applying, we mention the Normalized Compression Distance [122]—which is an alternative to similarity and is based on Kolmogorov complexity—and the Set-based Complexity [75], which is aimed at reckoning the heterogeneity of an ensemble of non-random strings.

3.2 Evolutionary Music: entering ISL

Some works have suggested that musical creativity is mainly formed by implicit knowledge. The hypothesis is that the brain models music as a hierarchy of dynamical systems encoding probability distributions and complexity [42]. These distributions sampled from musical pieces may refer to the characteristics of a composer's implicit (statistical) knowledge: a high-probability transition may be one that a composer is more likely to predict and choose, compared to a low-probability one corresponding more to an unusual variation [45]. Based on these assumptions and aiming at engineering more human-like and creative computational procedures, in this study, we combined implicit-knowledge mechanisms with novelty search in a genetic algorithm to emulate an agent's effort to produce novel and appealing sequences of actions (musical pieces). The resulting series have to be, at the same time, both familiar (concerning the knowledge initially provided) and novel. Using Markov chains, we have integrated implicit statistical knowledge extracted from music corpora with an adaptive novelty search mechanism. Part of this work was published in [10]. The reminder of this section describes the algorithm along with the main design choices. Results are shown in two distinct musical contexts: Irish music and counterpoint compositions.

3.2.1 Related Work

The applications of Markov chains in music have a long history dating back to the 1950s [167]: for detailed reviews on AI methods in music, or other examples and techniques, see [30, 143, 97, 126, 64]. Similarly, evolutionary computation has been used for generating music since long [100, 18], and there are currently several systems that produce music through evolutionary techniques [17, 144, 126].

From our perspective, however, music generation is just a case study: we focus on modeling a general (context-independent) method for generating sequences (not limited to music) based on implicit mechanisms. In addition, the search towards creative potential represents a different approach in contrast to the more common optimization practice, as the objective function tries to capture several, somehow subjective, features of the piece of art produced. However, some of the latest and most comparable approaches to this work are perhaps those in [155, 56, 106]. *Continuator* is an interactive music performance system that accepts partial input from a human musician and continues in the same style as the input [155]. The system utilizes various Markov models to learn from the user input. It tries to continue from the most-informed Markov model (higher-order), and if a match is not found with the user input, the system continues with the less-informed ones. In [56] the authors describe a method of generating harmonic progressions using case-based analysis of existing material that employs a variable-order Markov model. They propose a method for a human composer to specify high-level control structures, influencing the generative algorithm based on Markov transitions. In [106] the authors propose to capture phrasing structural information in musical pieces using a weighted variation of a first-order Markov chain model. They also devise an evolutionary procedure that tunes these weights for tackling the composer identification task between two composers. Another work is that of *GenJam* [18]. It uses a genetic algorithm to generate jazz improvisations, but it requires a human to judge the quality of evolved melodies. Finally, *GEDMAS* is a generative music system that com-poses entire Electronic Dance Music (EDM) compositions. It uses first-order Markov chains to generate chord progressions, melodies, and rhythms [4].

3.2.2 Materials and Methods

In the work presented in the previous section, we conceived a genetic algorithm (GA) which, starting from a given inspiring repertoire and a set of unitary moves, generates symbolic sequences of movements (i.e., choreographies) exploiting similarity with the repertoire combined with the novelty search approach. In this work, we discuss an adaptive genetic algorithm combined with Markov chains—built up from a corpus of music excerpts. The algorithm evolves the parameters (weights) of a constructive procedure that acts on the model and produces new pieces of music that are intended to be novel variations upon familiar music. In addition, the model is used also for evaluating the similarity (the objective function) of generated sequence to the starting knowledge: in this case, the unbiased model was used, that is, without using weights. In this contribution, we build upon the novelty approach and we make it adaptive to make it independent from specific ranges of the functions involved in the algorithm. When the objective function stagnates for several consecutive iterations, novelty search is applied to move away from that local minimum. On the other hand, when novelty causes the fitness to diverge–more than a certain amount from the last best value found–novelty is turned off. Algorithm 1 shows the main loop, and Algorithm 2 shows such a procedure.

Markov model: chains and score

Markov chains allow us to grasp the statistical structure of sequential phenomena (i.e., music, movements) but also statistical learning and knowledge in humans [46]. It has been observed that transitional probabilities sampled from music (based on Markov models) may also refer to the characteristics of a musician's statistical knowledge and capture expertise and temporal individual preferences in playing music [44]. We consider here sequences of symbols from a finite alphabet, which can represent e.g. melodies. To model this implicit knowledge, we computed the Markov chains with memory m (or Markov chain of order m) up to the $m = 5$ order[7] starting from a set of musical pieces. For each order m, transitional probabilities are computed for each excerpt as frequency ratios: $P(y|x_m) = \frac{\#x_m y}{\#x_m}$, given a symbol y and a (sub-)sequence x_m (the past) of length m, where $x_m \rightarrow y$ is the inspected transition. Thus a transition matrix is calculated for each order, where rows are the contexts (pasts, or memories) and columns are the next inferred/emitted symbol. Every matrix has a variable dimension since the number of rows changes with the number of contexts (ngram) found in each model order. Columns instead are fixed since feasible symbols are those of the 0_{th}-order (thus, no context) model; that is, the number of symbols the model has ever seen and that it knows. This model is used with (evolved) weights to generate sequences (phenotype) and without weights (in its original form) to evaluate the similarity of the generated sequence and, indeed, the inspiring corpus. The role of the GA is to bias the model (in fact, its production) toward a trade-off between already-known sequences and new, novel ones. As we increase the context size, the probability of the alphabet becomes more and more skewed, which results in lower entropy. The longer the context, the better its predictive value.

Objective function

The objective function captures the familiarity (or the membership, the similarity) of a sequence with respect to the Markov model resulting from the (inspirational) musical corpus. So given a sequence $X = x_0 x_1 ... x_n$ the Markov score is defined as the product of TPs of symbols in the sequence

$$score(X) = P(x_0) \times P(x_1|x_0) \times P(x_2|x_0 x_1) \times ... \times P(x_n|x_{n-m} ... x_{n-1}) \quad (3.4)$$

[7] In the data we used, orders higher than 5 are not "expressive" because of the data limits and structure: at some point, higher orders contain about the same information held in the previous ones.

where $x_{n-m}...x_{n-1}$ is the (past) sequence of length m up to the $(n-1)_{th}$ symbol. For a given chain, it might happen that a transition (past $\rightarrow$ symbol) does not exist. If that actual past does not match a transition in that current order, we shorten the past ($x_{n-m}...x_{n-1}$ becomes $x_{n-m+1}...x_{n-1}$) and move down an order (i.e., a chain), looking at shorter contextual information to guide the generation. Finally, we apply the negative logarithm to the Markov score and turn the GA objective into a minimization problem:

$$\underset{X}{minimize}: \quad -\log(score(X)) \tag{3.5}$$

Encoding

The GA manipulates the parameters of a randomized constructive procedure that acts on the Markov model. The genotype is an array of 6 decimal elements—that sum up to 1— representing the weights to assign to each computed chain in the Markov model. Namely a weight i for each i_{th}-order Markov chain (i.e., a categorical distribution for the chain choice), as in Table 3.1. Every positional value of the array weights the probability of

Table 3.1: Example of an individual

w_0	w_1	w_2	w_3	w_4	w_5
0.0	0.3	0.05	0.4	0.2	0.05

the corresponding order in the model when generating a sequence. Notably, these arrays weigh a Monte Carlo process that selects, for each symbol to be emitted in the generation, the order–i.e., the Markov chain–to look at when looking for the transitions to produce it. The phenotype indeed is represented by all sequences generated by the model with that given array of weights (the individual of the GA).

Adaptive Novelty Search

To steer the generation towards novel productions, we followed the novelty search already explored in [9]. This algorithm is applied to compensate for a lack of diversity concerning the best individuals already found. It consists of a bi-objective optimization activated when the main objective (e.g. the fitness) stagnates. The novelty is computed as the mean L_2 norm between a genotype (weight vector) and an archive of past genotypes.

Using the Markov model, there is no explicit boundary for the objective function since the Markov score depends on the length of the sequence being evaluated. To tackle this problem, we conceived an adaptive mechanism for the activation of novelty, unrelated to the specific range of values of the objective function, based on the results obtained in previous iterations. At each iteration, the algorithm stores the best result obtained. If such value does not change for a given number of iterations, somehow the algorithm is stuck at a minimum. In such a case, a selection mechanism intervenes to take into account novelty values in a Pareto optimization with the objective function. The algorithm starts to look at the novelty of individuals. When novelty search has moved the score away of a certain amount from the last best value found, it is turned off and the regular evolution with the Markov score is resumed. See Equation 3.6.

$$\begin{cases} \text{if } bestFit \simeq prevFit, \text{for } k \text{ times}, & \text{switch to bi-objective} \\ \text{if } |lastAvg - bestAvg| \simeq stdevLast, & \text{switch to mono} \end{cases} \tag{3.6}$$

As well as for the objective function, we applied the negative logarithm to novelty too. Thus the bi-objective optimization is intended to minimize both the main objective and the novelty of individuals.

$$\underset{X}{minimize}: \quad -\log(novelty(X)) \tag{3.7}$$

We remark that biasing towards novelty does not mean just adding randomness, but rather diversifying with respect to the best solutions found.

Archive assessment For the assessment of the archive, we followed the approach used in [228] except for one aspect; we did not consider a threshold for the individual in order to be added to the archive. Instead, we considered, at each iteration of the genetic algorithm, the individuals of the elite, from the elitism process. For these individuals, we calculate the dissimilarity as in the mentioned work.

3.2.3 Results

The most suitable musical contexts in which our technique can be applied are those in which improvisation plays an important role; yet, we also need structure to some degree, in such a way that the implicit (soft) constraints imposed by the style can be detected. This way, the music resulting from our method has some amount of novelty, yet still in the style of the examples provided. For our experiments we chose two notable musical contexts: traditional Irish tunes and Orlande de Lassus' *Bicinia* [89]. In this section, we first introduce these cases, and subsequently, we present a selection of the typical results achieved by our technique.

Irish songs and Bicinia

Traditional Irish music is strongly characterized by its melodies: most old tunes are just melodic (see e.g. [225]) or they are the result of an improvisation upon a given *ground*, i.e., a bass line providing also a harmonic base (see e.g. [169]). In any case, the melodic part of a traditional Irish music is currently the most important component, and melodies are usually played with variations, improvising upon a given melodic structure. A large corpus of traditional Irish airs is available in abc notation,[8] which makes it possible to extract melodies as sequences of symbols, each representing both note and duration. A typical traditional Irish air is shown in Figure 3.1. From these airs, we extracted all the ones in the key of G and assigned one symbol to each $\langle$*note, duration*$\rangle$ pair.

The second musical context we have chosen is that of two voices counterpoint, which is one of the simplest and oldest forms of polyphony [185].

[8] http://www.norbeck.nu/abc/

Figure 3.1: Score of a well known traditional air titled “The south wind”.

In origin, a voice was superimposed to a given one, called *cantus firmus*, in improvisational settings. This original impromptu spirit was subsequently substituted by a more elaborated compositional approach, leading to marvelous multi-voices counterpoints, such as the ones composed by Gesualdo da Venosa. The main technical characteristic of two voices counterpoint can be summarized in a small set of rules involving the intervals, i.e. the distance in semitones, between the upper and the lower voice. For example, the distance between C and F (above C) is 5 semitones. Obviously, these are not all hard constraints, but some are rather preferences, and they have also been subject to change across the years according to different musical aesthetics. In fact, rules emerge from the performance practice and are subsequently systematized. We focus here on two voices XVI century counterpoint, characterised by rules such as the ones listed in the following:

- no parallel fifths or octaves are allowed (e.g. C-C cannot move to E-E, C-G cannot move to D-A);

- fifths and octaves should be intercalated by *imperfect consonances*, i.e. thirds and sixths (e.g. an allowed sequence is C-G, C-A, D-A);

- *dissonances*, i.e. seconds, fourths and sevenths, should be *prepared* and then *resolve* to a consonant interval by descending (e.g. D-B, C-B, C-A).

These rules constitute a core around which further indications can be added, depending on the specific musical context and taste. However, the rules listed above can be considered the essential syntactic nucleus of two voices counterpoint of the XVI century.

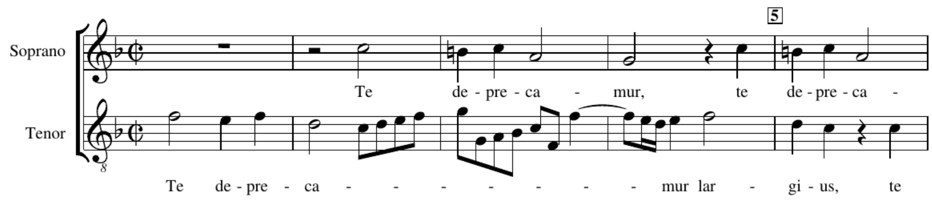


Figure 3.2: Excerpt of bicinium no. 1, “Te deprecamus”, by de Lassus. Score extracted from `https://imslp.org/wiki/Category:Lassus,_Orlande_de`.

In our tests, we have taken all the twelve two voices counterpoint compositions, called *bicinia*, by Orlande de Lassus, which are available as MIDI files.[9] In Fig. 3.2 we show an excerpt of a bicinium by de Lassus. This second context was chosen to assess to what extent our method is able of identifying recurrent patterns and rules typical of a music genre. In this case, we have encoded the twelve MIDI bicinia as sequences of intervals (i.e. distances in semitones between upper and lower voice). As the two voices have in general different durations, we have sampled the music at steps of duration 1/32 and taken the intervals in semitones, deleting repetitions. This provides the repertoire on which the Markov models are computed. A typical result from our system is a sequence of integer numbers representing intervals in semitones which can be used as a guideline for composing the upper voice upon a given *cantus firmus*.

Experimental settings

Differently from usual optimization contexts, in our case a good performance does not correspond to the one that leads to the overall best objective function values, but rather to a good balance between similarity (Markov score) and novelty. Therefore, we tuned the parameters of the algorithm trying to attain an effective interplay between score and novelty. The results

[9] `http://icking-music-archive.org/ByComposer1/Lasso.php`.

we present have been obtained with a population of 100 individuals, uniform crossover with probability 0.5, Gaussian mutation ($\mu = 0$, $\sigma = 0.3$) with both chromosome and gene probability equal to 0.35, and 200 generations. The novelty is activated after 5 idle generations (the best score s_{pop} in the current population is stored, along with the standard deviation of the populations scores σ_{pop}) and deactivated when the difference between the score of current best individual and s_{pop} is greater than $\sigma_{pop}/3$. The plot of score and novelty of a typical run is shown in Figure 3.3.

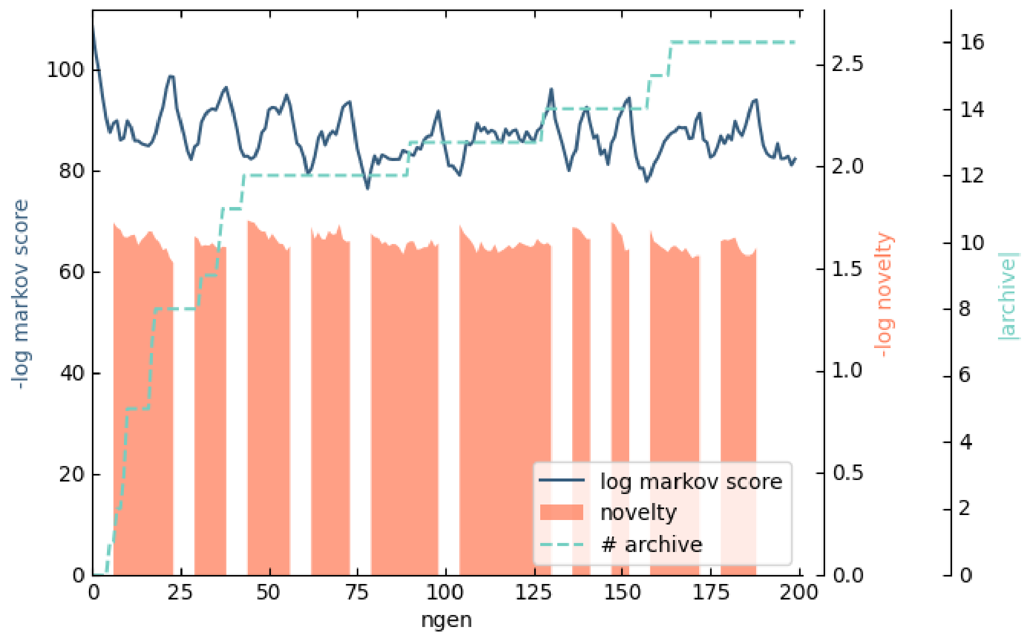


Figure 3.3: Plot of score and novelty of a typical run. Both the functions are to be minimized and novelty is activated, adaptively, only when diversification is needed. The number of individuals in the archive, involved in the calculation of novelty [9], is also plotted.

We can observe that the score oscillates: whenever the algorithm stagnates, novelty is activated so as to increase diversification. When this latter is high enough, only the Markov score is kept as objective function. In a sense, we can describe the dynamics of the algorithms as a biased exploration of local minima, as typically done by Iterated Local Search techniques [19].

Musical results

The generation of melodies inspired to traditional Irish airs has been evaluated by sampling some weight vectors from the final populations and using them to generate actual music. By analyzing the results both through visual inspection and by listening to them, we observed that the music generated is similar to the repertoire provided but with variations and recombinations of patterns. A couple of excerpts are shown in Figure 3.4, where we can observe variations of typical Irish melodic and rhythmic patterns: the characteristic *run* (i.e., a fast sequence of notes, typically in a scale) in bars 4, 5 of the first example and the syncopated and composite rhythm in the second one.

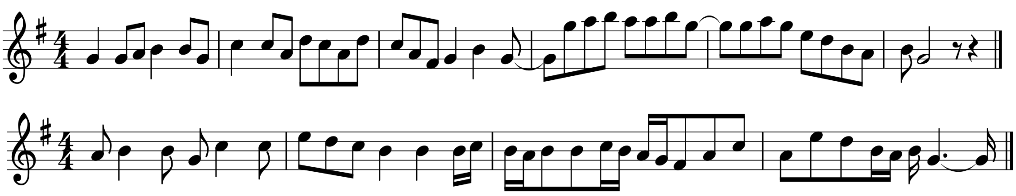


Figure 3.4: Two typical excerpts of automatically created Irish music.

The second test case concerns de Lassus' Bicinia. The main result attained is that it was able to discover the basic rules that characterize two voices counterpoint. In particular the rules extracted that have more strength are: *incipit* with a perfect consonance (unison, octave or fifth), no consecutive octaves or fifths, and dissonant intervals followed by consonant ones—both perfect and imperfect. In Figure 3.5 we show an excerpt of the counterpoint produced by applying one of the sequences generated by our algorithm to a given cantus firmus (Chanson CXXVI from manuscript Bibl. Nat. Fr 12744 published by G. Paris). As the algorithm returns a sequence of intervals, it could be used as a tool that assists composers by suggesting feasible note choices, once one of the two voices (*cantus firmus*) is given.

In conclusion, for both contexts, the algorithm was able to identify the

Figure 3.5: An example of a two voices counterpoint. The lower voice is the cantus firmus, while the upper voice has been generated by applying a sequence of intervals generated by our algorithm.

core regularities and elaborate around them. The calibration of the parameters is important to achieve a good balance between the tendency of just recombining the patterns learned and the exploration of new possibilities. However, the choice of parameter values does not seem critical, because the combined use of a stack of Markov models of varying orders and novelty search makes it easier to achieve this trade-off.

In general, switching on and off the novelty search throughout iterations, in the GA, produces more variety (in terms of variance and peaks of the score) than using the Pareto between score and novelty from the start to the end of evolution.

3.2.4 Discussion

The algorithm we have presented has proven to be able to generate novel, yet somehow familiar, melodies. An interesting perspective is creating non-homogeneous repertoires, maybe just including music the user likes, without any genre restriction. This way, our system can produce music that merges some of the peculiar features that meet the user's tastes.

Future work will focus on quantitatively assessing the properties of the generated sequences employing information theory measures, such as block entropies [196] and complexity measures like set-based complexity [75]. Some of these measures can also be introduced in the generative process, to limit human evaluation as much as possible. In addition, some metrics can also be used to assess the distance between sequences or cluster them [33].

As the proposed technique is general and can be applied whenever the goal is to produce sequences of actions, we plan to explore multimodal automatic generation by combining Markov models from two different contexts, e.g., music and text.

3.3 Machine Improvisation: ISL and Generalization

According to the *Grove Music Online* [147], improvisation is “The creation of a musical work, or the final form of a musical work, as it is being performed. It may involve the work’s immediate composition by its performers, or the elaboration or adjustment of an existing framework, or anything in between. To some extent every performance involves elements of improvisation, although its degree varies according to period and place, and to some extent every improvisation rests on a series of conventions or implicit rules.”. We emphasize here that the notion of improvisation involves the extemporaneous creation of sequences of notes (i.e., pitches and durations, including dynamic and agogic expressions) performed according to shared, implicit and explicit, conventions and rules. Another important property that characterizes improvisation is *risk*, i.e., “ the need to make musical decisions on the spur of the moment, or moving into unexplored musical territory with the knowledge that some form of melodic, harmonic, or ensemble closure will be required.” [147]. Therefore, the act of improvising requires the capability of balancing the adherence to the rules that have been learned and an ingenious exploration outside their boundaries.

Recent works address musical improvisation in the context of statistical learning [44, 43]. Inspired by these studies, we developed a model for emulating implicit sequential learning and creativity. Here, we took the opportunity to show, in particular, the effects of generalization on produced sequences. In this work, published in [11], we illustrate an ISL mechanism that creates melodic improvisations by performing a stochastic walk on a generalized graph of TPs. The use of the generalized graph makes it possible to combine both the adherence to a given set of implicitly learned rules and a cautious exploration outside those conventions. In Section 3.3.1, we describe the model and the creative algorithm, while results are illustrated in Section 3.3.2. We conclude by discussing further improvements and future perspectives on this approach.

3.3.1 Materials and Methods

Initial acquisition of implicit sequences may arise from ISL [53]. In addition, previous studies suggested that implicit knowledge governs music acquisition [180]. Drawing upon these perspectives, we wanted to grasp the implicit aspects of a creative process in a minimal model capable of learning and generating (musical) sequences. Hence, the basic idea is to exploit the implicitly learned knowledge to produce novel musical strings. In addition, we also provided a generalization step from this implicit knowledge, to acquire structured information from the context.

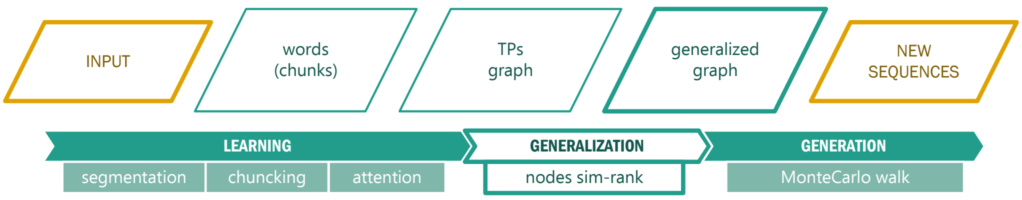


Figure 3.6: Sketch of the process: focus on the discussed generalization step

The algorithm proceeds through three subsequent phases: learning, generalization, and generation (Figure 3.6). In the learning phase, we introduced TPs at two specific levels: between symbols, as a cue for segmenting the incoming input into small segments (or chunks), and between these formed chunks. According to the view that the role of TPs might be (i) to prepare, or aid, learners to memorize recurring items (as cues) [58] or (ii) between formed chunks for distributional computations [219]. After the learning phase, the graph of TPs between chunks undergoes a generalization phase. This phase draws on the distributional learning hypothesis [176, 140] which argues that people use statistical learning to acquire grammatical categories from the input (i.e., the contextual information surrounding words). Indeed, by relying solely on distributional information (i.e., contextual information in the graph), this mechanism exploits node similarity (SimRank) to reveal these categories (namely, form classes in language).

Finally, this new generalized graph is employed to generate novel, structured sequences using a Monte Carlo walk.

Learning

The learning phase consists of two mechanisms: tracking the transitional probabilities between symbols (second order TPs) to be used as cues to segment the input into words (or units, chunks), and tracking TPs between those words (first order TPs) to form a graph made of transitions between chunks [219].

Following the approach of PARSER, at each iteration a random integer (in [1,3]) is chosen to determine the number of disjunctive, embedded units to be perceived: namely, the size of the next percept[10].

At each perception cycle, TPs between the observed symbols are stored. Initially, the algorithm tries to use stored TPs to find a drop in the transitions between symbols that would determine the boundary of a word. On the other hand, if no TPs cue is found, a syllable is perceived (two consecutive symbols). The segmentation strategy used in this work is one of the simplest where a boundary is detected if the transitional probability of the upcoming symbol drops under a certain threshold, so if $TP_i > TP_{i+1} + \epsilon$. In the present

[10]This for simulating the effects of the other various factors which modify the boundaries of the actual attentional window, such us the listener's state of vigilance.

study, we used $\epsilon = 0.2$ as an empirically selected threshold. However, various strategies could be exploited (see next chapter, or [41] for an old but detailed analysis): recent studies, for example, suggest the use of backward TPs [159, 188], but this is out of scope here.

After segmentation, TPs between the resulting ordered units are recorded. Note that this *per se* represents an abstraction intended to grasp the dynamics, the transitions, between formed words—not between symbols.

At the end of each perception cycle, decay and interference are applied. The interference mechanisms is simulated by decreasing the weights of the units involved in the currently processed percept. Instead the decay mechanism, contrary but in accordance to PARSER, has been implemented with the following formula (3.8):

$$D = (\exp^{-\frac{\Delta T}{S}})/C \tag{3.8}$$

In the formula above, D represents the amount of decay for each involved unit at each percept cycle, ΔT represent the amount of time (discrete, in number of steps) passed since the percept was perceived and stored in memory, S [11] represent the stability of decay, that is the decay of decay (grater the number, more steady is the decay), and C is a parameter for choosing the initial value of decay (i.e., the initial strength).

Generalization

The output of the learning phase is a graph where nodes represent units (chunks, words), and edges represent transitional probabilities between words. To construct the generalized graph, the procedure first computes the form classes, using similarity between nodes, and then generates some sequences (with the TPs graph) that are parsed to build the higher-level graph. The similarity between nodes is computed using a SimRank [104] measure over inward and outward edges. SimRank is a graph-theoretic measure that says

[11] In all the experiments and simulations, we kept a high value for this parameter, $S = 1000$, for simulating a constant decay similar to that used in PARSER.

"two objects are considered to be similar if they are referenced by similar objects". In this case, we used a slightly modified version where "two objects are considered to be similar if they are referenced by similar objects ... and refer to similar objects". That is, nodes are grouped if they have similar inward and outward edges, so if they have a similar neighborhood. Precisely, we group nodes with both inward and outward SimRank greater than a threshold value γ. In this case, we used $\gamma = 0.5$ as, in our experiments, it provided convincing similarity values over known samples. So for each node i we calculate $FC_i = \{I_i \cap O_i\}$ where, for each node j:

$$I_i = \{N_j : SimRank_{IN}(N_j, N_i) \geq \gamma\}$$

$$O_i = \{N_j : SimRank_{OUT}(N_j, N_i) \geq \gamma\}$$

The formed groups represent what in language acquisition is called form classes [195]. Once calculated, the form classes are used to parse some generated sequences (using the TPS graph), and the new generalized graph is then built. Transitional probabilities between formed (form) classes are computed as well, counting transitions over the parsed sequences.

Generation

The generalized graph is then used to produce novel sequences. In the present experiment, we opted for a simple Monte Carlo choice over the edge probabilities to traverse the graph. At each visited node, as in general it may contain words that can be used in the same position in the construction, a word is picked randomly (the nodes that contain alternative words are called here *choice nodes*). Another possibility for selecting a word is to use the weights (the frequencies) of the words to employ another Monte Carlo choice at each node. However, we have preferred random sampling to emulate a more analogical selection: by modeling the transitions between classes, in each choice node, there are 'equivalent' words, which are related to each other, thus enabling this possibility.

3.3.2 Results

A prominent context for improvisation is of course music. Since the system we developed is mainly focused on sequences of symbols, we opted for melodic pieces of music. Therefore, we provided the system a set of melodies belonging to a given style (e.g. Irish music) upon which the TPs graph and the generalized graph can be built. The latter provides then the basis for the generation of new melodies in the style of the given repertoire, but with variations and explorations in the implicit boundaries set by the examples. The resulting melodies are characterized by improvisation flavor, as they have not the structure of a complete piece of music, but capture the main stylistic features of the original compositions, like a musician making extemporaneous explorations around a given style.

To test the system we chose two different styles: Irish melodies and the six preludes from solo cello sonatas by J.S. Bach. Irish melodies have been retrieved from Henrik Norbeck's `abc` tunes [88]. All the 136 melodies in the key of G have been gathered (including variations of the same song) and the `abc` notation symbols, which encode the music in textual form, have been directly used as sequence symbols. The second repertoire of melodic music, instead, has been retrieved in MIDI format from David J. Grossman's J.S. Bach page [40]; the MIDI files have been converted to an intermediate textual representation by means of PyPianoroll [52] and transposed to the same key, so as to have sequences composed of symbols representing the intervals from a common base note. In both the corpora of examples, a symbol in a sequence represents both pitch and duration.

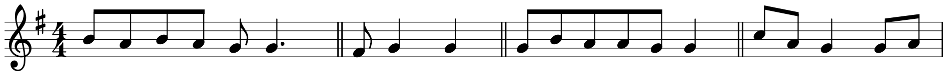


Figure 3.7: An example of the options for melodic segments in a choice node of the generalized graph.

We are interested here in the features of the generalized graph and the characteristics of the melodies it produces. The number of nodes containing alternative choices and the number of choices estimate the amount of "controlled exploration" around the musical style learned. For example, a typical choice node in the generalized graph of Irish music may have the following alternatives: `B A B A G G3 | F G2 G2 | G B A A G G2 | c A G2 G A`, represented in Figure 3.7 in score notation. In general, the segments differ in start and end note, as well as total duration; therefore, they do not represent equivalent alternatives, but rather different sub-paths that can be used to compose a new path which is likely to combine fragments of melodies in an original way, yet keeping the flavor of the melodies in the repertoire. The generalized graph built from Irish music has 151 nodes, of which 19 are choice nodes. The choices in each node are distributed between 2 and 8, with a median of 3. The resulting melodies are similar to the ones in the repertoire, but characterized by a considerable degree of originality. The interested reader can find audio excerpts and score transcriptions at [145]. The generalized graph of Bach preludes for cello solo substantially differs from the one related to Irish music, as it it composed of a greater number of nodes (526) and a lower number of choice nodes (10), all with just 2 choices except one with 4 choices. Another remarkable difference with respect to the previous case is that the melodic segments in each choice node are longer. The musical difference between the two repertoires is wide and this has of course strong impact on the properties of the generalized graph. Irish traditional music is characterized by simple elements: almost all the notes used belong to the scale of G major and the maximal difference in pitch is about two octaves. Moreover, the melodies are often composed of long sequences of notes at small intervals and few large steps (e.g. of an octave or a fifth). Conversely, the preludes for cello solo by Bach span a wider range of pitches and the use of chromatisms is extremely common. In addition, the examples available are much less than the Irish ones, so the probability of overlaps between portions of melodies is much lower. The features of the two graphs reflect the musical properties of the two styles in that the richness of Bach's style and, above all, the hierarchical structure of his compositions limit the adjustable interchangeability of melodic segments which is expressed by the generalized graph. However, the musical result of artificial improvisations in the style of Bach's cello preludes is appreciable (audio and score excerpts are available at [145]).

3.3.3 Discussion

We presented a generative model that uses a stochastic walk on a topological generalization of variable Markov Chains (TPs graph), to produce novel musical sequences. The presented work is intended to be a basic module of a more extensive system conceived for emulating the learning of implicit sequences. It is intentionally domain-general and symbolic since it is intended to model various phenomena: from music and language to movements and social interactions [142]. This learning system assimilates implicit knowledge that becomes the basis for modeling implicit, automatic behaviors. In these regards, we envision adding a short memory module, to model higher-level phenomena such as attention, for example. However, even if in this case the focus was on the learned material and the effects of generalization on productions, the ultimate goal is, in fact, that of producing creative outputs. In this perspective, the next step will be to use an ad hoc, creative Monte Carlo walk in place of the simpler stochastic one, that is, to give the model the ability to explore creative paths instead of the most (or the least) probable ones. We believe a model built in this way could also provide a place, an environment, for simulating and studying a variety of behaviors in cognitive science and creativity.

3.4 Summary and Discussion

This chapter presented three applicative examples for generating creative symbolic sequences. However, these experiments served also to investigate the potential processes underlying creativity. In particular, the importance of convergent and divergent phases, respectively, to exploit and explore the given conceptual space (represented by the input corpora), the role of TPs (studied via Markov Chains), and finally, the components for an ISL model, and the effects of generalization on produced sequences.

From a pure computational perspective, we analyzed the quality of the sequences generated by the algorithms, as the ultimate goal of applied CC is the creation of novel and engaging artifacts. Initially, we exploited (string) similarity and novelty approach through a genetic algorithm for generating robot movements. Then, we introduced variable-order Markov models to enhance both the similarity assessment and the production (also introducing an ad-hoc procedure for generating musical pieces) and improved the novelty approach by making it adaptive. Finally, in the last section, we presented a mechanism reflecting ISL that grasps the inner structure of the sequences and generalizes the learned knowledge to discover new, higher-order relationships over the elements found in the sequences. The explorations performed on the latter approach widen the range of possible investigations and applications, for computational but also for psychological research on learning and creativity. Contrary to the other two approaches, the last one is online and unsupervised, and parses the input incrementally, emulating perception and learning as in [162] and in [133]. Moreover, the generalization step is conceived for reflecting the distributional learning ability observed in humans by the psychological studies conducted on language acquisition [26]. However, the issue of learning could also be faced from another perspective. It is worth noting that the ISL model retains transition probabilities between elements so expectations of the next future. Therefore, another testable hypothesis, for example, is that distributional learning (or learning in general) is error-driven [152]. Contrary to the other two approaches, the latter has the potential to be used not only as a system for generating sequences but also as a computational tool for carrying out experiments and exploration in cognitive research. For these reasons, we decide to turn our attention to the cognitive aspects it models. The next chapter concerns these explorations.

Chapter 4

A Production-Oriented ISL Model

Following the psychological studies on language acquisition, we aimed to construct a fully incremental, online model for sequence learning that uses the same chunk information and distributional statistics to perform both learning and production. The model's inner workings approximate comprehension by learning chunks and statistics and by using them to segment input sequences into related groups of words (such as shallow processing), and production, using the same learned material to produce new outcomes. We hypothesized that both problems could be tackled, to a large extent, by recognition-based processing tied to chunks, discovered through sensitivity to transitional probabilities between units.

Hence, this chapter focuses on the ISL paradigm, introducing another major factor for creativity: namely, *surprise*. This resulted in the formulation of a creative Monte Carlo walk, using (and discussing) Simontons' principles, for the generation. Adopting ISL, we explored, computationally, some of the cognitive mechanisms involved in the implicit formation of sequential knowledge in humans, using the latter to generate ever more human-like productions. As a result, we came up with a minimal computational toolkit; a psychologically-inspired computational model for implicit sequence learning and creative generations. Thus, this chapter introduces CREATIVITiPS (creativity + TPs + tips: just a words pun for referring to the model and its code[1], also abbreviated as TiPS) from the perspective of the cognitive mechanism it implements: in particular segmentation, chunking, and attentional memory mechanism, as well as generalization and the dynamics of creative production.

Various simulations were conceived to evaluate learning and production and to demonstrate the potential of the proposed framework for the study of creativity. Several experiments have been conducted to evaluate the model's sensitivity to its main parameters. Moreover, the model has been successfully compared to similar approaches in developmental psycholinguistic findings spanning a range of phenomena.

4.1 Overview

Cognitive and computational approaches, focusing on the process side, try to describe brain functions and behaviors for creative thinking. Such proposals help demystify human creativity by offering insights into the underlying mechanisms and their characteristics. As said, this is also the purpose of CC. Thus, following the principles of [36], we propose a psychologically-oriented account for expertise acquisition in sequential learning and subsequent creative generation. The model reflects some aspects (i) of Implicit Statistical Learning (ISL) in Sequential Learning, using simple associative, memory, and attention-dependent mechanisms, (ii) abstraction, and (iii) of creative production, exploiting the implicit, learned knowledge.

On one hand, we applied our model to a series of different languages trying to cover various sequence complexities: to investigate the role of Transitional Probabilities, chunking, and attentional mechanisms for sequential learning, and to see the amount of (structured) information this simple model could grasp and generalize. On the other hand, we wanted to investigate some

[1] `https://github.com/mattia-barbaresi/creativitips`

implicit factors related to the creative process. As it has been suggested, creativity is mainly formed by implicit knowledge [233](cf., insight, the "a-ha" or the "eureka" moments). Therefore, we wanted also to exploit this knowledge for creative productions. Using Campbell's and Simonton's principles of creativity, we give an account of the exploitation of this implicitly learned knowledge to come up with novel generations.

We begin by introducing the model and its inner workings. Then results are reported on the acquisition together with the simulation of some of the related psychological experiments on children's sentence processing. Finally, we show some applications of this model that extends beyond language to cover the acquisition of different kind of sequential processes, for example, music. Our ultimate goal is embedding this model as a controller for humanoid robots to investigate these implicit aspects in robot movements and social interactions [142].

4.1.1 Segmentation and Chunking

Across the ways suggested for integrating chunking and statistical computations, previously discussed in Section 2.2.1, we explored the fourth possibility, stressing their interplay and their reciprocal influence. If the ISL mechanism is ubiquitous, TPs could serve as statistical cues for identifying chunks and also as relations within multiword segments (chunk relations) as suggested in [218, 47]. Hence, we consider the sensibility to statistical regularities as a general component, as well as the chunking mechanism, ubiquitous in brain functioning. In the case of segmentation of sequential input, many cues could be used to learn underlying structural relations and to gain expertise [20] (e.g., learn to play an instrument). In this sense, transitional probabilities are just one of these exploitable cues to find segmentation boundaries. However, if we consider the more general issue regarding the sequence encoding of order and sequentiality in the brain [48], the learned statistics can serve as building blocks for more elaborate codes: for instance, in language acquisition, learning the TPs between syllables seems to be the basis on top of which words and syntactic tree structures are built [130]. This issue gains particular value for sequence production, inference/prediction, and space-time structure learning, for example as it is in social interaction [82] and intuition [148].

4.1.2 Attentional and Memory Mechanisms

Other than TPs, we wanted to model also chunking and some basic memory mechanisms. Hence, drawing upon the approach of PARSER [162], we devised a short-term memory to implement decay, interference, and a basic attentional module. This component stores the recent segments found while reading the input sequence; then, due to forgetting, if a chunk is not reoccurring in the future, it is gradually forgotten. On the other hand, when perceived again, its weight is enforced. In these terms, it acts like a sieve, an access door to TPs between chunks, that helps filter out unwanted units (noise or spurious information). Therefore, it serves to ease the statistical computations of the ISL process. Moreover, this straightforward mechanism allows us to emulate attention because active elements (those above a threshold) in memory can shape the perception of the input and can consequently facilitate the acquisition.

At an implementational level, this module is kept as a separate component from the TPs graph; in this way, it is possible to model diverse decay and interference rates for memory and TPs, for instance. Another advantage concerns the independence of the TPs module that, in this form, can be applied to diverse input streams at various levels of abstraction. However, another possibility is embedding this short-term memory mechanism directly on the learned graph (c.f. [151, 149]): as TPs are calculated between chunks, the nodes in the TPS graph are the same units transited in memory.

4.1.3 Generalization

Generalization refers, in general, to extension of learned statistical structure to unseen stimuli within the same modality domain and has been demonstrated in ISL studies [72]. Evidence for sensitivity to distributional information in language acquisition comes mainly from the phenomenon of generalization behavior: categories defined by highly variable input distributions are more readily extended to novel tokens [7]. While these results support the idea that input distributions matter [31], they provide only a rudimentary understanding of the mapping between these distributions and the acquired category representations. Hence, we introduced a generalization phase that acts on the learned graph and exploits distributional properties (nodes neighborhood) to discover higher-order categories, called form classes, in the language acquisition domain.

There is a link between abstraction and creativity, which relies on the formation of new structures, and the discovery or definition of new relationships between existing entities. Abstraction could therefore foster high creativity. This phase could be assimilated to the incubation step, or mind wandering (and dreaming, to a large extent), and it could facilitate the emergence of remote or hidden associations [156]. Moreover, it could serve for enabling, or enhancing, the other brain mechanisms for sequence coding [48]. In a multimodal input perspective, for example, applied over the features of perceived entities, this mechanism could lead to another kind of statistical, distributional modeling such as that in long-term memory behavior (i.e., concept formation [23, 139]) and, from another perspective, forming broader classes or concepts, it could also enable or enhance analogical reasoning mechanisms, such as bisociations [115] and analogies [99].

4.1.4 Generation

Psychological theories of creativity typically involve (1) a divergent stage that predominates during idea origination (for a review see [183]), and (2) a convergent stage that predominates during the refinement, implementation, and testing of an idea. Divergent thought is characterized as intuitive and reflective; also, it involves the generation of multiple discrete, often unconventional possibilities. A measure of divergent thinking ability is fluency, or the number of new, innovative ideas that can be generated. Conversely, convergent thinking involves selecting or tweaking the most promising possibilities resulting from a critical and evaluative analysis. Campbell [29] argued for a single generic process that could account for creative thought and "other knowledge processes", namely, Blind Variation and Selective Retention (BVSR). The basic idea is that the most effective way to discover something new—whether an invention, discovery, or adaptation—is through experimentation, trial-and-error, or generation-and-test procedure. In contrast, if one knows in advance that some ideas are useful or suitable, then the person has merely engaged in reproductive rather than productive thinking: confirming what is already known rather than exploring the unknown [202]. Starting from an implicitly formed graph (and a generalized one) that represents the learned knowledge of the agent, we can now discuss the generation phase. This phase undergoes the BVSR (or divergent/convergent) and could be also conceived as the problem-solving phase. Moreover, in [95], the authors concluded that creativity encompasses both conscious and unconscious incubation and insight, so we hypothesized a creative generational BVSR as an implicit process within these phases.

As said, we wanted to model the individual, implicit dimension of creativity, mainly with two motivations: to exclude the explicit and external variables influencing the creativity of an agent, and to understand it from the basis. In particular, covering those processes of the incubation/illumination stages. That is, we wanted to investigate the possible reasons for that sensation that comes right before the idea pops up to the conscious awareness of the agent: when there is that feeling that a solution is coming. In other words, the preconscious activity that leads to the "Eureka!" moment. According to Sadler-Smith [186], expertise helps a creative person progress to the inspiration stage (in Rhodes 4Ps) by unconsciously evaluating their unconscious ideas. Domain expertise forms the cognitive substrate for creativity. In addition, Dijksterhuis and colleagues have proposed an Unconscious Thought Theory that also resonates with Wallas' incubation stage as it proposes that "contrary to conventional wisdom", unconscious thought has a "generative power" concerning creative cognition and complex decision making [51]. With sufficient incubation, the creator may have an insight, eureka, or "ah-ha" experience in which a solution flashes to mind [95]. Yet, because such inspirations are by no means guaranteed to work, this illumination phase must be followed by the verification phase in which the idea is directly tested and evaluated, whether externally or internally (cf. [50]). If this test fails to confirm the utility of that solution, then the cycle will continue in the hope that an effective solution is finally found [201].

Therefore, from this standpoint, we implemented a model for "expertise acquisition", where an agent incrementally, implicitly learns the material to which it is exposed and then tries to generate feasible solutions. Following this reasoning, the generalization phase, conceived as a period where the agent incubates the learned information and ruminates on it, produces more elaborated, abstract knowledge that could influence production. However, even if expertise and experience influence learning, divergent and convergent thinking [38, 220, 66, 227] stages, its role in these diverse phases has to be fully unveiled yet; this work also discusses some computational aspects related to this issue.

4.2 Related Work

Given the interdisciplinary approach, we took inspiration from theoretical works spanning ISL literature, cognitive science, and Computational Creativity. From a computational perspective, we opted for a Markov Chains modelization. Statistical Learning is about transitions' conditional information, and Markov Models are a natural choice to model such aspects. From a cognitive perspective, the Chunk-Based Learner is the most similar related work found in the literature. IDyOT is another approach that investigates SL, but with slightly diverse viewpoints and aims. Finally, PARSER and Edelman's models are studies from which the present model draws the various cognitive mechanisms it implements: respectively, attentional-memory perception and generalization. In the remainder of this section, we introduce these studies and the main aspects that differ from the present model.

Chunk-Based Learner The Chunk-Based Learner (CBL) of language learning [133] is the most similar approach. In the same vein as the learning model of the present work, CBL learns in an unsupervised, incremental, online way. In this model, the initial formation of chunks relies on computations aimed at locating boundaries in the dips of the backward TPs distribution. As well as the proposed model does with forward TPs. However, the author focused on the usage-based approach to language acquisition, and their model does not account for attention, generalization, or implicit creative mechanisms. Contrary to the CBL model, we conceived the chunking and the statistical module as separated and independent but interleaved mechanisms. In particular, we consider tracking TPs as a modular mechanism complementary to chunking and segmentation [15]. Moreover, the computation of TPs is considered an ubiquitous mechanism in the brain. So we used them at two specific levels: a more perceptual way, where TPs between symbols are computed to aid segmentation, and the other, where TPs are used for encoding sequentiality at the level of chunks (that is, between them).

ADIOS and U-MILA For pattern extraction and structured generalization, two processes that have been implicated in language acquisition, we took inspiration from Edelman's approach (ADIOS [211] and U-MILA [117]). The authors conceived a procedure to discover higher relations in an oriented graph. That is, to reveal the grammar implicitly represented in the given graph. In particular, they look at the inward and outward streams of every path in the graph for each start-end node pair and use their ratio to cluster sub-paths. In the present case, inspired by their approach, the conceived algorithm groups nodes by analyzing their similarity based on their inward and outward edges. However, we wanted to exploit the similarity of the connections of the nodes (e.g., the neighborhood) instead of analyzing longer paths in the graph. Again, our focus is more on the role of the generalization step than the effectiveness of the implemented function.

PARSER PARSER [162] is the approach followed for realizing chunk-memory mechanisms such as memory decay (forgetting) and interference. The primary motivation of PARSER is to account for optimal segmentation in terms of simple and ubiquitous psychological processes. Starting from the observation that, in humans, attentional coding naturally segments the ingoing information into small and disjunctive parts of variable length, the model encodes the input as a succession of provisional units comprising a random number of components (between 1 and 3). These units are stored in a lexicon and are subject to ubiquitous laws of memory: they are reinforced whenever they reoccur in the input, and, conversely, their strength vanishes as a consequence of decay and interference with the processing of similar material. Therefore, the selection of relevant units for the language structure (compared to the irrelevant ones) emerges as a natural consequence of these memory laws. Decay eliminates those units that do not occur often enough, while interference makes the model sensitive to statistics, such as the bi-directional transitional probabilities between the unit components [161]. Because perception is guided by internal representations, the learned chunks become new primitives: thus, making the system able to build chunks whose components were not initially perceived in a single attentional focus. However, the PARSER model uses random choices to decide the length of perceived chunks and works with syllables. We tried to circumvent this issue using TPs as a cue for the segmentation.

In conclusion, our memory component differs from PARSER in three fundamental ways: (i) our memory (Percept Shaper in PARSER) starts empty instead of encoding all the possible syllables; (ii) we used a logarithmic instead of a linear decrement (to combine an initial rapid decay with the long term persistence) to model more complex situations; and (iii) we exploited TPs between symbols as a cue to alleviate the use of random segmentation.

IDyOT IDyOT by Wiggins is more a theoretical framework intended as a cognitive architecture explaining the proposed theory. While, in IDyOT, creativity is equated more with prediction, in this thesis creativity is intended in the sense of Campbell's BVSR: it requires a blind variation of learned knowledge to create novel (selected) combinations of stimuli while prediction finds known items [54]. In this sense, our approach is more like BrainGene [166]. Moreover, there is a fundamental difference in scope: we want to focus more on the processes, trying to account for (in a computational and operational fashion) psychological and neuroscientific findings instead of trying to conceive a whole theory of creativity and consciousness.

4.3 Materials and Methods

The model draws upon works on Implicit Statistical Learning which suggest that the initial acquisition of sequences (in implicit sequence learning, such as in language or movement acquisition) is based on the ability of the brain to implicitly grasp the statistical regularities of the processed input, in particular exploiting Transition Probabilities [53]. The computational model used for these experiments is the same introduced in Section 3.3.1(see Appendix for the function pseudocode), where it was applied to music for exploring preliminary effects of generalization on learned knowledge and thus on productions: aiming to combine the observance of implicitly learned rules with a guided exploration of them. However, this chapter discusses the same mechanisms more in detail and from a more holistic cognitive perspective. The general workings of the algorithm proceed through learning, generalization, and generation, as for the previously discussed computational model, with the additional creative walk conceived for experiments on the convergent (task-oriented) phase (Figure 4.1). The model proceeds incrementally, parsing the sequences percept after percept. Every percept is formed by the union of (i,e. chunking) up-to-three perceived units. These perceived units are picked using attentional cues (matching active units in memory), by analyzing TPs between symbols (n-grams), or randomly if the other two mechanisms could not be exploited. Once these units are perceived, another TPs module stores the transitions between these chunks.

Then, the graph of TPs between chunks undergoes a generalization phase, which draws on the distributional learning hypothesis [176, 140] that argues people use statistical learning to acquire grammatical categories from the contextual information surrounding words. Finally, this new generalized graph is employed to generate novel, structured sequences using mainly two modes of generation: a stochastic one, more suitable for divergent thinking, and a second one for convergence, task-oriented production.

Through adjustment of the model's parameters, some diverse scenarios could be studied. In addition, the model can be applied both on unsegmented and segmented corpora.

4.3.1 Learning

The symbolic computational model learns in a purely unsupervised and incremental fashion. It is based on the online processing of transitional probabilities (TPs) and chunking. Its design reflects some simple psychological mechanisms and can address general aspects of ISL and structure learning.

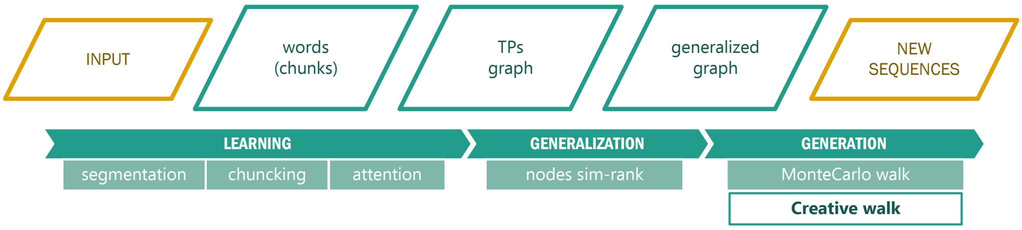


Figure 4.1: Sketch of the entire process: in addition to the MonteCarlo walk, a creative one has been added

Therefore it is not specific to language. The model combines a Statistical Learning module, that tracks TPs between symbols and formed chunks, and a memory-based chunking model (PARSER). It is intended to be as simple as possible and modular to accommodate any future modification and less implicit modules (e.g., improved attention mechanisms, explicit knowledge, algebraic module and inference, analogies). Algorithm 3 shows the pseudocode for such a procedure.

A pivotal issue in the ISL realm concerns TPs when they are exploited as statistical cues: we know that the brain has this sensibility, what is not clear is to what type (e.g., forward, backward TPs o Mutual Information) and in what manner this statistical information is used [41]. We employed Forward TPs in this dissertation, yet we explore some methods to detect boundary cues: in particular, the Brent technique [25], that is, a word boundary is inserted when the probability of a bigram is lower than those of its neighboring bigrams (previous and the next), thus using the context in which it occurs (Equation 4.1), and other two, newly conceived mechanisms that involve the use of a TPs average computed over a sliding time-window proximal to the current transition. One uses this average in a way similar to the Brent technique (Equation 4.2), and the other one simply takes a dip under the average as a boundary for segmentation (Equation 4.3). Even if straightforward, this last method does not suffer the problem of the other two: a trough-based approach, like the first two, is not capable of extracting unigram words, since such words would require two adjacent transitions. In particular, given the sub-sequence $..kwxyz..$ a boundary is placed between x e y, due to the used method, if:

$$\textbf{BRENT} : TP(x|w) > TP(y|x) < TP(z|y) \tag{4.1}$$

$$\textbf{AVG} : TP(x|w) > avg > TP(y|x) \tag{4.2}$$

$$\textbf{FTPAVG} : TP(y|x) < avg \tag{4.3}$$

with $TP(a|b)$ expressed as the conditional probability of a given b, and avg calculated as the average of all the transitions in $wxyz$ (i.e, $TP(w|k)$, $TP(x|w)$, $TP(y|x)$, $TP(z|y)$). Above, the three equation uses first-order transitional probabilities. However, these formulas can be easily extended to higher-order TPs simply considering pasts of increasing length; throughout this thesis, we used second-order TPs [177], so the transition between x e y, in the examples above is calculated as $TP(y|wx)$.

4.3.2 Generalization

Distributional information provides a powerful cue to syntactic category membership, which can be exploited by many simple, psychologically plausible mechanisms [173]. Many researchers have suggested that distributional learning mechanisms play a major role during grammatical category acquisition since linguistic (form-) classes (like nouns and verbs) are primarily defined by the ways lexical items are distributed in syntactic contexts [175, 14, 28]. These perspectives suggested that, at the core of language acquisition (and more), there are some general-domain cognitive processes: such as categorization, chunking, analogy, and cross-modal association. Even in this case, we adhere to these perspectives. Moreover, as suggested in [7], we believe that acquiring specific structure from linguistic input, and generalizing beyond that input to novel exemplars, represent a single mechanism. The abstraction mechanism learns classes/abstractions as well as the higher-level structures and relations. As said, we believe that this kind of computational approach for language is general and could be echoed in many other domains. Also in this case, an induction phase was conceived to discover higher-level relations (i.e., form classes) based solely on the distributional information learned, and so available, in the TPs graph. That is the transitional probabilities between formed chunks.

Hence, we conducted some experiments based on the number of sequences generated to form this generalized graph, to test if they facilitate the emergence of remote or hidden associations. Algorithm 4 describes the steps of this phase.

4.3.3 Generation

The generalized graph is then used to produce novel sequences. The previously described stochastic Monte Carlo walk is employed for divergent production, and a new procedure is conceived that concerns a utility value for the convergent task. Then the model potential is tested in both phases.

The outcome of learning is a bag of words in short term memory and a graph representing transitions between those learned words, forming a probability distribution among them. So the learned model yields the information for Simonton's originality parameter (1-p). It's worth noting that the model offers the knowledge after learning so it can be seen as the starting implicit knowledge an agent has at the time that it first start thinking about the creative generation problem.

Simonton's formulation of creativity [206] suggests three parameters for characterizing a potentially creative thought: the idea's initial probability (p), the final utility (u), and the creator's prior knowledge of that utility (v). The three parameters then lead to a three-criterion multiplicative definition of (personal) creativity, namely, $c = (1 - p)u(1 - v)$, where u represents the idea's final utility, the first factor indicates originality $(1 - p)$ and the third factor surprise $(1 - v)$–that is how much new knowledge an agent gain once the idea is generated and evaluated. Creativity takes continuous values $(0 \leq c \leq 1)$: if c = 0, then creativity is nil, but if c = 1, creativity is maximal.

A fascinating feature of this multiplicative definition is that whenever u and $(1-v)$ are held constant at nonzero values, then c maximizes as p goes to zero so that $(1-p)$ goes to one. Once two parameter values are given, the remaining parameter value is usually constrained: If $u = v = 1$, then it should follow that $p = 1$, but if $u = 0$ and $v = 1$, then it should follow that $p = 0$. In other words, among useful and surprising ideas, the most highly creative ideas are those that require an incubation period before the insight appears. An idea that comes to mind without engaging in such incubation can still be creative, but it will be less so to the degree that p exceeds zero.

To generate new sequences, such creativity values are used, at each edge, to perform a Monte Carlo walk through the generalized graph. Initially the graph has only probabilities (p), so it does not know anything about the utility for a given sequence. However, within each cycle, utility (u) and prior knowledge(v) are updated. So the algorithm starts looking for just uncommon sequences. At each step, the generated sequence is subject to an evaluation using the specified utility. Having the utility value, the u and v values of each edge visited for generating that sequence are then updated. The utility value is an overall judgment of the sequences. Thus, to assign a value for each edge we had to deal with a sort of Credit Assignment Problem: in these experiments, we defined utility u as the average calculated from the newly assigned utility value and the old one, and the prior knowledge v as a sort of online variance of the utility.

What follows are the formulas used for the update:

$$u_{new} = (u_{old} + u_{actual})/2 \tag{4.4}$$

$$v_{new} = 1 - \sqrt{(u_{old} - u_{actual})^2} \tag{4.5}$$

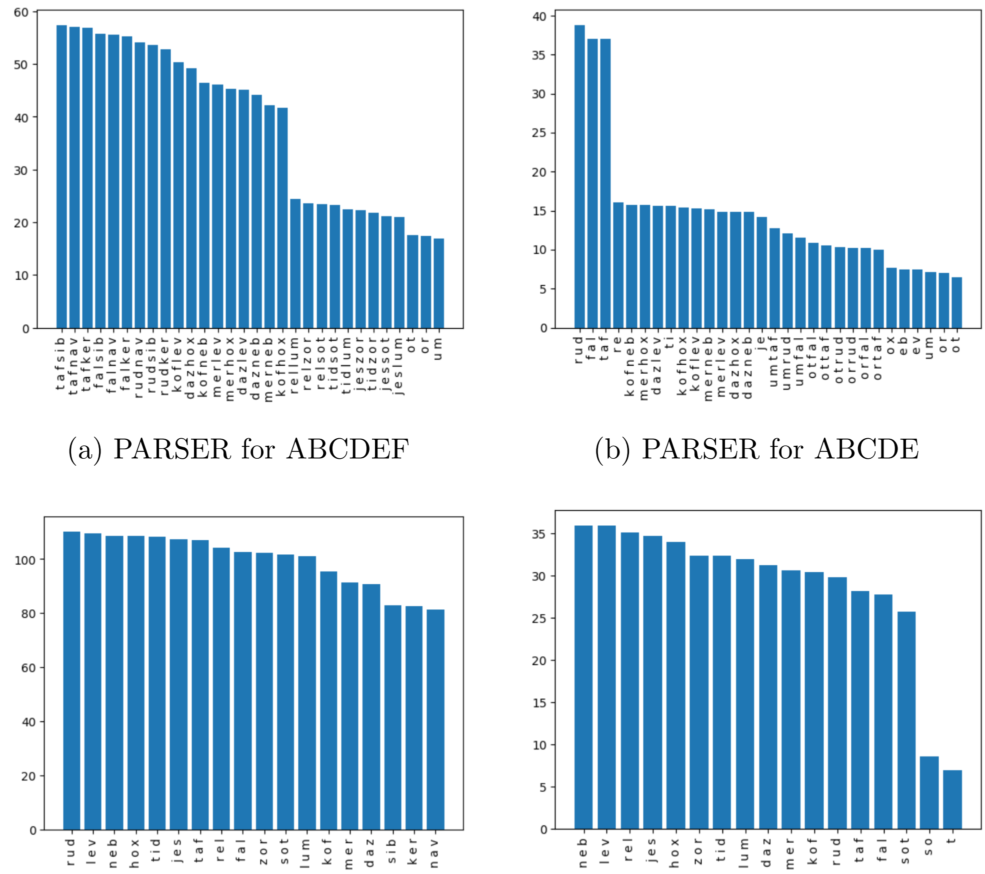
4.4 Illustrative Simulations

In this section, we discuss, from a computational point of view, some of the mechanisms related to known issues in ISL and Computational Creativity literature. Our aim is that of discussing, by giving computational interpretations of, these issues more than trying to find biological or neurological foundations for these behaviors. The hope is to provide functional insights regarding these mechanisms and their interplay.

4.4.1 Segmentation or Chunking

In the ISL realm, chunking and segmentation are two distinct mechanisms over which the debate on their roles and interplay is unsettled yet [159]. Here, we don't want to discuss this particular issue, instead, we want to draw some computational reflections about the generic case of sequence learning. For our purpose, however, we wanted to investigate at the computational level the benefit of both mechanisms. The first consideration is about the effect of chunking: in Figure 4.2a and Figure 4.2b the entire percept, formed by units, is stored as well, causing longer words (in fact, joint words) to be remembered better then effective words. Another point regards the limit of assuming syllables as atomic units: in Figure 4.2b, sequences, as well as words, have odd lengths. Thus, a key aspect regards the potential of ISL (compared to pure chunking methods) to aid segmentation, achieving similar or better performance in chunk recognition too.

This experiment is the simplest one to demonstrate the effects of TPs used as cues instead of relying on syllables (bi-gram only) in a PARSER, language-like scenario. In their paper, the authors used languages composed of syllabic words (i.e., even-length words) for their experiments: this is an impractical and simplifying assumption for modeling other, general real-life scenarios and languages such as music or movements. Contrarily, TPs as cues are a more general and robust mechanism for this purpose (note that TPs vary with language complexities and transitions). To explain this difference, we choose another well-studied artificial grammar used to explore whether learners can use transitional probability as a cue to phrase structure [219]. The language is formed by 6 classes (A, B, C, D, E, F), each comprising three



(a) PARSER for ABCDEF

(b) PARSER for ABCDE

(c) TPs + memory, for ABCDEF

(d) TPs + memory, for ABCDE

Figure 4.2: Memory of PARSER and (TPs + memory) for ABCDE and ABCDEF languages. TPs + memory does not use chunking: that is, perceived units are stored separately, without storing the joined percept

words formed with the pattern consonant–vowel–consonant (so of length 3, in contrast to Saffran even-length words). To stress further the concept of even-length words, we applied both models also on a slightly modified version of the language where sequences are composed using the ABCDE pattern instead of the ABCDEF one.

In Figure 4.2 are shown the active units in memory at the end of learning. The PARSER memory is compared against the memory formed with TPs and the same attentional mechanism. As said, PARSER relies on grouping together syllables and on frequent chunks meanwhile TPs rely on segmentation, i.e., finding chunk boundaries. However, if PARSER, as a chunking module, uses only simple mechanisms, such as associative learning, tracking requires a more sophisticated one (tracking at least bi-grams and unigrams to calculate transitional frequencies). Nevertheless, using TPs (with attentional-memory mechanisms) seems more appropriate for a general domain class of phenomena (not only language). Moreover, this TPs approach supports the hypothesis of [73], which suggests that people could simultaneously acquire knowledge about concrete chunks and abstract structures of the temporal sequence, meanwhile chunking mechanisms do not account for temporal structure.

4.4.2 TPs and Memory: A Matter of Complexity

In this example, we wanted to show a straightforward scenario involving memory usage (active chunks) compared to the exploitation of statistical cues. For doing this, the action taken by the algorithm are plotted at each cycle in the learning phase. As could be expected the system relies on memory whenever the language to learn is a simple one. With more complex languages (in structure and vocabulary) the units in memory continue swapping, as a result of decay and interference interplay, and the chunks are no longer a reliable resource. In this case, the exploitation of TPs appears to be a more reliable mechanism. However, relying on memory depends most on chunk frequencies in the short term, not on the dynamics of transitions.

It is worth noting that we are considering the differences in relying on memory instead of segmenting the input. This scenario is similar to the exploit-explore scenario: that is, whatever the system has to rely on memory (exploiting) or explores novel information in the input (segmentation, TPs). This could be related also to the dual-route view of imitation [217]: supposing

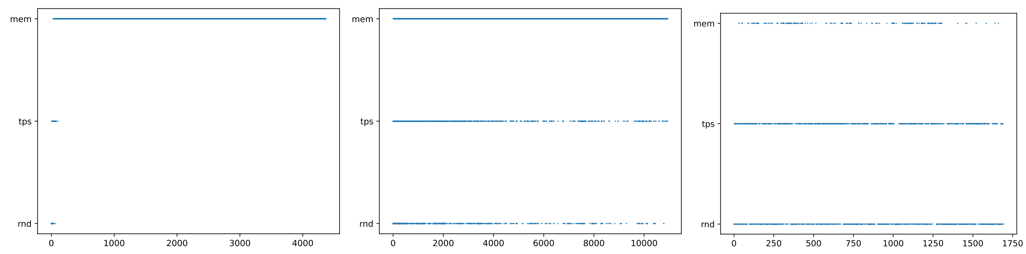
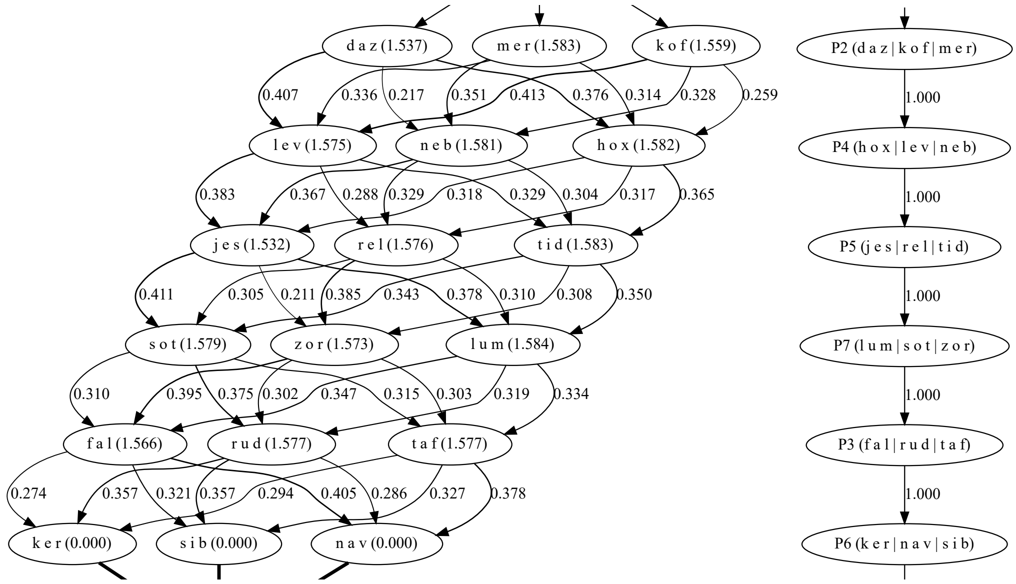


Figure 4.3: Perception mechanisms used at each cycle in learning ABCDEF language (4.3a), irish songs (4.3b) and Bach's preludes (4.3c).

that what could be in (short-term) memory (compared to long-term, procedural, TPs) is also related to retrieval (and also rehearsal) mechanism, so related to known information. Typically imitation and other implicit sequence learning tasks involve Reaction Times (RTs) to measure implicit learning. The latter, as for other more sophisticated indicators (e.g., ERPs), could be compared, for example, with the present computational model behavior, aiding to more clear statements and precise experiments in such contexts. As a hypothesis to test, which supports this human behavior, shifts in one way or the other could be reflected in the use between memory and statistical computation modules, respectively.

4.4.3 The Generalization Mechanism

This step is inspired by studies on incubation: a time when unconscious behavior takes over, which allows for unique connections to be made without consciously trying to make logical order out of the problem. Taking as an example the language used in the previous section, we discuss the effects of the abstraction step. Figure 4.4 shows the learned TPs and the graph yielded after this step for an archetypal scenario. Building from what is grasped by the learning procedure, which contains words and transition probabilities, the generalization phase looks at node similarity to compute the form classes. Then it parses the generated sequences to build the new generalized



(a) (b)

Figure 4.4: The effect of abstraction. (a) The learned graph (TPs graph), and (b) the generalized one (GG) where the nodes represents the classes of the grammar: ABCDEF

graph that outputs form classes and transitions between them. Of course, this is an oversimplified example: where the grammar of the language learned is straightforward. In more real-world cases, the learning procedure yields errors (i.e., it grasps non-words) that propagate throughout generalization. Nevertheless, there are many possible solutions to this issue: reiterating the generalization procedure (e.g., as successive incubation steps) to refine the generalized graph, which also relates to simulating replay mechanisms [231, 127, 121], or using more explicit knowledge, or using task-driven or error-driven procedures [141, 150], for example. While generalization is convenient in the learning phase to grasp higher-level relations–such as relations between classes of chunks–its effect on productions is twofold: i) it brings errors, as the generalized graph encodes new possible, and might unfeasible, transitions (and combination) between elements of the created classes, ii) it brings structure to produced sequences. If with (i) it allows for new possibilities at the cost of making errors, with (ii) it allows for the potential discovery of new higher constraints that could also reflect on the evaluation of creativity: if one considers this higher-level knowledge, some variations would have no creativity because not all nodes have multiple outgoing edges in the graph, and so the probability of that single edge (representing the variation) is maximal (1.0), that is, it's constrained, and this means that it brings no novelty. This mechanism is like thinking about a particular production (idea or solution) from another (higher) perspective.

4.5 Experiments

While the previous section discussed some pivotal issues by showing exemplifying scenarios, in this case, the focus is on the cognitive aspects the model entails. In the end, the ultimate goal of this dissertation is to develop an approach that can be useful not only for robotics but also for psychological and social studies, in line with the CC philosophy. Hence, here we discuss some experiments concerning: (i) the divergent and (ii) convergent abilities of the model, on a typical language used in psychological tests conceived for language acquisition; and then, (iii) the comparison with a psychological, related prototype for language learning. In particular, for convergent and divergent phases, we used a language generated by the grammar of Table A.1, used in [219], where we explored the role of the previously introduced mechanisms and the effects of generalization. Then, the comparison is carried out in the shallow parsing task of mother-child speech corpora, considering two other models, CBL and CogComp Chunker–as the comparison and the baseline model, respectively–in the same vein as the experiment conducted in [133].

4.5.1 Divergence

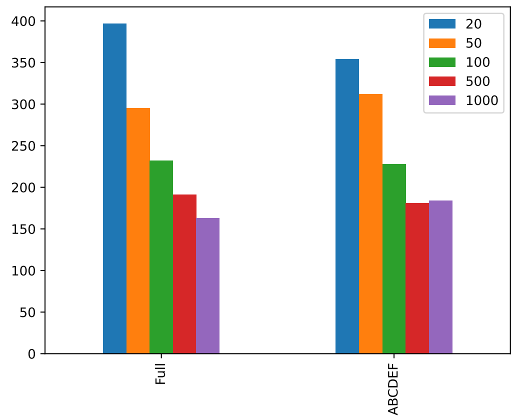
To explore the divergent characteristics of the system, in this section, we discuss two experiments where, more than on the quality of solutions, the focus is on the number of variations the model can produce. In particular, we tested three different methods for boundary detection, diverse decay parameters for the memory, and the effect of the incubation period, seen as the number of sequences created (with the TPs graph) to obtain the generalized graph. Even if the chunking mechanism in memory is the primary drive for word discovery, the effects of its interplay with the ISL module are not trivial: the segmented units and the method used for boundary detection influence the attentional mechanism in memory, that in turn affects the encoding of the successive transitions between chunks. Another important investigated factor, mainly for generalization, is the role of context. For these reasons, we decided to use two similar grammars using the words and classes in Table A.1: the first language is built using the ABCDEF structure, and the latter using ABCDEF + ABCD + ABEF + CDEF (also called hereafter the *full* grammar). Then we discuss the differences between the TPs graph and the generalized one. The hypotheses, as stated in Section 4.4.3, are various. In particular, the generalized graph should be able, at least in the divergent phase, to produce more alternatives, with the additional hypothesis that the "incubation period" should be more effective for convergent behavior than the divergent one. Moreover, given that the language is quite simple, the system will be relying more on memory; thus, memory decay should have a role also in the divergent phase (a light forgetting implies, in chunking, the memorization of multi-unit words, encoded, in turn, in the graphs, constraining generalization and, in general, variations). Finally, given the intricate interplay between the two modules, even if the system had to rely more on memory, the method used for boundary discovery should influence learning and, as a consequence, production.

Experiment 1

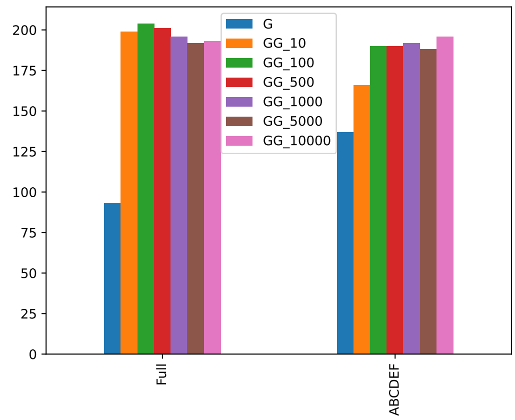
In this experiment, we wanted to evaluate the potential the divergent mechanism has, to reach unknown solutions. Given the focus on sequentiality, creativity should be in how units are ordered. Therefore, from the corpora of generated languages, we removed all the sequences that contain the specific subsequence **"neb-rel-sot"**. Accordingly, we removed 30 sequences from the ABCDEF + ABCD + ABEF + CDEF corpora (972 in total), resulting in a training set of 942 strings, and 27 sequences from the ABCDEF corpora (729 sequences in total), resulting in the second training set composed of 702 elements. Using these two training sets, we explored diverse decays (C in Equation 3.8, where tested values are: 20, 50, 100, 500, 1000), the three different formulas (equations 4.1, 4.2, and 4.3) for boundary discovery, and a different number of sequences produced to construct the generalized graph (called **repetitions**: 10, 100, 500, 1000, 5000, 10000). For each training set, once learned, we use the TPs graph to generate a variable number of *repetitions* to build the generalized one. Finally, we generated 1000 output sequences, with each graph, counting the number of produced strings that contain the sub-sequence *"nebrelsot"* (the number of hits). Extensive results are shown in the tables of Section A.3, which report the number of hits.

Results As expected, the decay parameter plays an important role. For these two simple languages, the values for decay that maximize the number of hits are 20 and, to a lesser extent, 50, as shown in Figure 4.5a. The lesser the decay, the greater the length of recognized words: this steers the system to store multi-words and the TPs module to store transitions between them (instead of between actual words) that, in turn, constrains the generation.

Moreover, as expected, the generalized graphs yield better divergent results than the TPs one, confirming that generalization could be useful for reaching unexplored areas of the given solution (or conceptual) space: see Figure 4.5b. However, contrary to predictions, the incubation period, simulated with the number of sequences used in the generalization step, does not



(a)



(b)

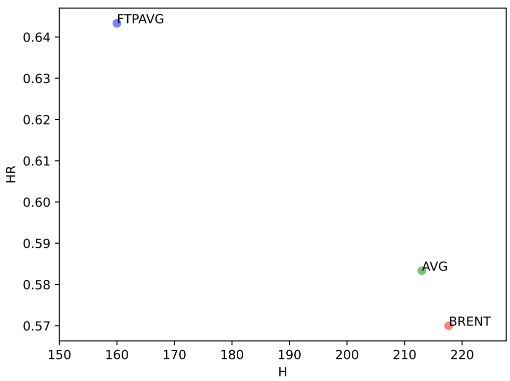
Figure 4.5: Results for divergence Experiment 1. (a) Aggregate number of hits for memory values for each grammar. (b) Aggregate number of hits for graph types (TPs graph and the various generalized ones) for each grammar.

influence the generation of novel sequences. An explanation for this behavior may lie in the (low) complexity of the two languages: in both cases, few repetitions are enough (for generalization) to improve the results and saturate the solution space. Perhaps for the same reason, even if slight differences could be noted, the structure of the language seems to be not so influential for the divergent generation of such solutions. One hypothesis is that a less structured language needs a longer incubation period for discovering specific structures; this is observable in the ABCDEF language, in Figure 4.5b, where the difference between the incubation period of 10 sequences and the others is evident. However, in the more structured (full) language, generalization has a discernible effect on the divergence.

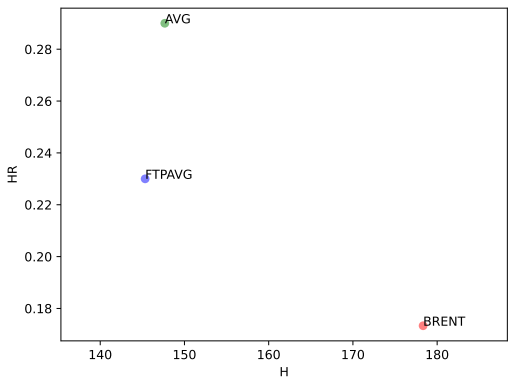
Experiment 2

Given the results of the first experiment, we conceived a second trial were we focused on smaller set of parameters values. Moreover, if in the previous experiment we tested the ability of the model to find a novel specific combination respect to the given corpora, in this simulation we want to explore the general divergent potential of the model to find novel sequences. It is worth noting that in the previous test we did not account for unfeasible solutions: in that case, even a ungrammatical sequence that contains the *"nebrelsot"* subsequence counted as an hit. In other words, in this case the task is conceived for a broader exploration of the actual solution space. Thus, we built the training set (547 sequences) using only the ABCDEF structure from which we have randomly removed 1/4 of the sequences for the test set (182 sequences); therefore, an hit is counted whenever the system produced a sequence of the test set.

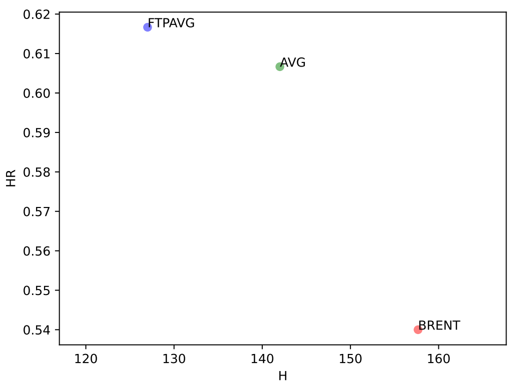
Results Complete results are shown in the tables of Section A.4, where the number of hits are shown: averaged (H), the hit rate calculated as the number of distinct hits (i.e., sequences) divided by the total number of hits (HR), maximum (M), and minimum (m). In Figure 4.6 we show the average number of hits (x axis) and hit ratios (y axis) of the TPs graph and three generalized graphs (GG) obtained with 10, 100, and 1000 repetitions. As it can be noted, the methods for boundary founding yields different behaviors. The BRENT technique have the highest number of hits (on average) but a scarce hit ratio: this means that it yields feasible solutions but tends to produce always the same strings. On the other hand, the FTPAVG seems to have the best hit ratio that means this method produces a set of more diverse solutions but struggles more in producing acceptable ones. Therefore, from these results, the AVG method seems to yield the best balance between the number of acceptable solutions and their diversification.



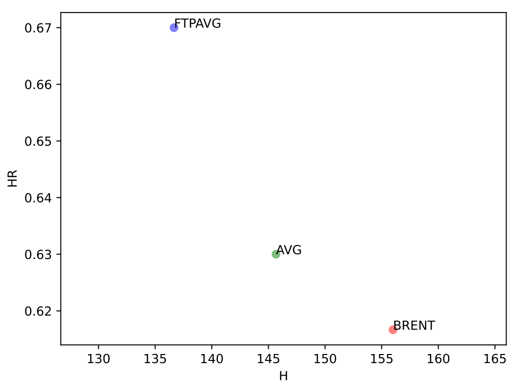
(a) TPs graph



(b) GG, rep = 10



(c) GG, rep = 100



(d) GG, rep = 500

Figure 4.6: Results for divergence Experiment 2. The three methods for boundary discovery are shown for each used graph.

4.5.2 Convergence

In the convergence process, the utility certainly plays a key role as it defines the solution space, and it is seldom easy to specify one that assures the expected results; this is particularly true in artistic and creative contexts. For these tests, we used the ABCDEF grammar of the experiment in Section 4.5.1 (without the sequences containing the substring *"nebrelsot"*) as the training set. We remind that the test set comprises all the feasible sequences containing *"nebrelsot"*. Then we conceived a utility function that assigns 1.0 if the generated solution matches a sequence in the test set, or, otherwise, returns a value that represents the halved similarity with *"nebrelsot"*, as follows:

$$utility(s) = \begin{cases} 1.0 & \text{if s } in \text{ test set} \\ similarity(s, \text{“}nebrelsot\text{”}) & \text{otherwise} \end{cases} \tag{4.6}$$

where $similarity(s, \text{“}nebrelsot\text{”})$ uses the Levenshtein Distance to compute the ratio of the most similar substring in s compared to *"nebrelsot"*, divided by 2. In this manner, even if the sequence is not in the test set, we reward solutions that contain the searched substring. As an example to clarify the second case, if s= ***"nebrelsot"*** or s= *"xxx**nebrelsot**yyyy"*–two solutions that are not in the test set–the utility value for both is 0.5. This choice is because this utility is more precise than, for example, the value calculated as the similarity with the entire test set (cf. Section 3.1), and it rewards also divergent behavior.

As previously discussed, it has been suggested that creativity is not only a combination of novelty and value but also involves a third factor: surprise. It is worth noting that the definition of a creative artifact requires that it had never been produced (and seen) before. Thus, each creative solution should be viewed as such only the first time it occurs; that is, an idea, or product, can not be surprising, novel, and thus creative, two times. From this perspective, the hypotheses for the role of surprise are manifolds. While it is fairly accepted as an element of evaluation of the creative artifact –paying attention in discerning it from novelty–its role in the creative process is less clear. For this reason, we conceived two diverse functions for evaluating creativity, used to generate novel sequences: one is Simonton's formula, comprising surprise $(1-v)$, and the other one is the *standard* formulation of creativity without this third element.

$$\textit{Simonton's creativity}: (1-p)*u*(1-v) \tag{4.7}$$

$$\textit{standard creativity}: (1-p)*u \tag{4.8}$$

These functions compute the creativity of the edges in the graph; these values represent the weights of the Monte Carlo walk that passes through the graph to generate new sequences. Therefore, we examined (i) their roles during the generation process and (ii) their efficacy as measures for convergence, analyzing, as done in the previous section, the number of final hits they led to. Following BVSR, we employ these formulas in a generation-and-test (or trial-and-error) routine. In Simonton's formulation, the v value takes the form of *sightedness*, and its counterpart, $(1-v)$ is what is called *blindness*. Stated in this manner it should be more clear that the Blind in BVSR is not randomness [202, 203]; following Simonton, as a consequence of blindness, the creative process should exhibit nonmonotonicity, seen as deviations from gradual improvement.

The procedure employs 1000 iterations, where for each iteration, the TPs graph and the generalized ones (considering repetitions) are used to generate 100 sequences (using the Equations 4.7 and 4.8). Then, these sequences are evaluated using the utility (Equation 4.6), and finally, the parameters of the edges (u and v) are updated, for each graph. This is a noteworthy aspect: if utility–and surprise, defined as the blindness on that utility– are known only in the moment of evaluation of the entire sequence, how can we model the creative choice along the train of decisions, that ends up with that solution? Trying to model this aspect, even if the utility applies on the entire sequence, the update procedure acts on each edge using formulas 4.4 and 4.5: thus, especially for v parameter, this factor introduces a nonobvious aspect of variability.

Results

The displayed results are obtained using the AVG method (Equation 4.2) and averaged over memory decays (C = 20, 50, 100, 500). Note that the figures show the creative value of solutions (C) using Simonton's formula (Equation 4.7). However, in this case, what is evaluated is the final artifact: that is the generated sequence, instead of the single edge. Hence, the p and v of a sequence, are calculated respectively as the product of the probabilities (p) and the average of sightedness (v) of the edges–in the graph–traversed to produce it.

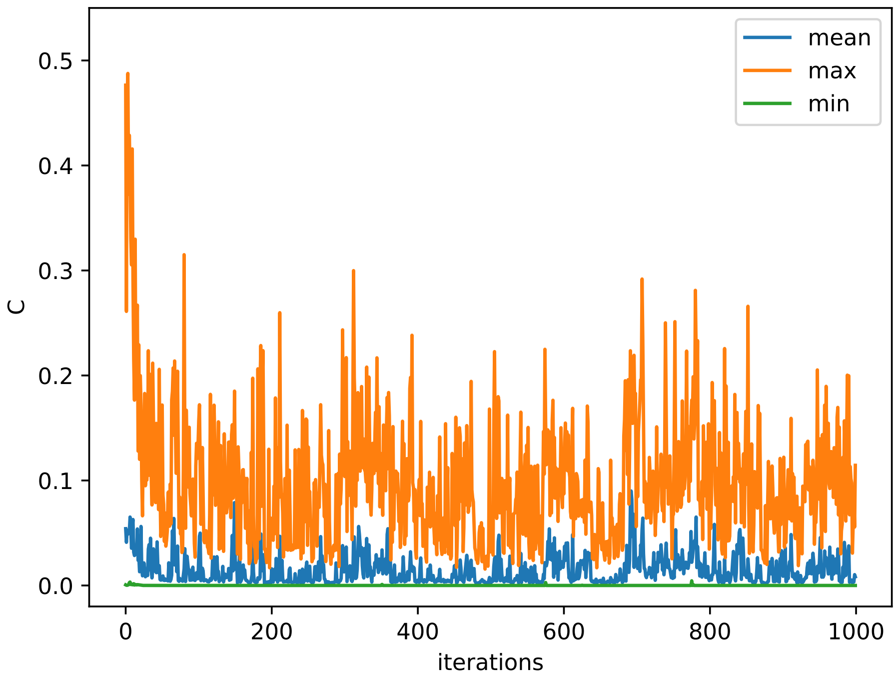


Figure 4.7: Results for convergence with TPs graph, using Simonton's formula for creativity.

The first detectable effect of using surprise within the generative process is to enhance exploration, as it penalizes already seen solutions (both good and bad ones). The variability of the max value, and especially, of the mean in each generation is more evident in the process using v. However, the best creative values seem to be yielded by the standard formulation (without v), which exploits the utility in the choice at each node and can exploit each edge multiple times. These aspects are noticeable comparing the figures 4.7 and 4.8 that show the creativity values (C, using Equation 4.7) of sequences generated by the TPs graph using the two diverse creativity formulations in the generate-and-test phase. In Figure 4.7, the expected behavior of Simonton's formula is evident: a peak of creativity, caused by the initial exploitation of creative paths, is followed by a subsequent decrement caused by the exploration (induced by v) of other paths.

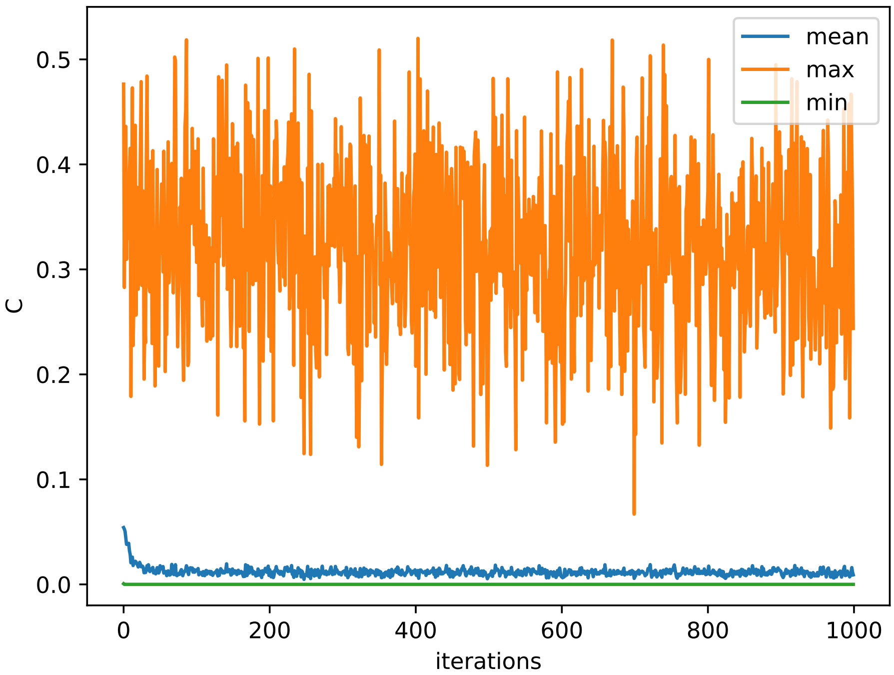


Figure 4.8: Results for convergence with TPs graph, using the *standard* formula for creativity.

Although less visible, the same trend is also noticeable in the case employing generalized graphs (in figures 4.9 and 4.10, see also figures in A.5).

Nonetheless, in the generalized graph, the stochastic (or the analogical) de-

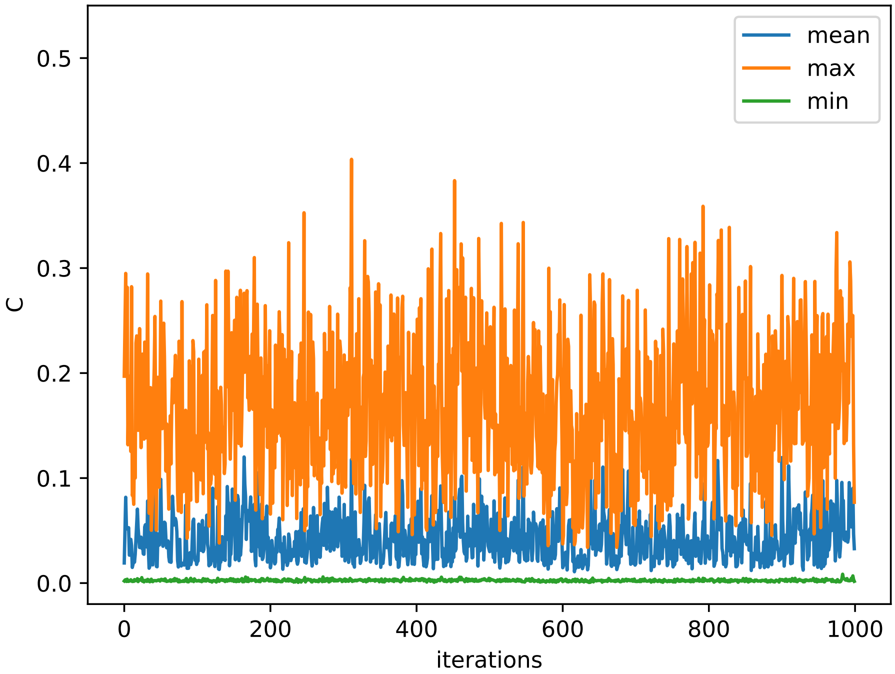


Figure 4.9: Results for convergence with generalized graph with 100 repetitions, using Simonton's formula for creativity.

cision made at each *choice node*, which are not affected at all by the utility, should have the same effect: that of favoring exploration. This variability induced by the "surprise" factor $(1 - v)$, contrarily to prediction, is reflected also in the number of appropriate, feasible solutions produced (i.e., final hits). However, this trend is favored by the utility. In particular, the second term, *similarity*$(s$, *"nebrelsot"*$)$, rewarding "divergent" solutions, increases the utility value of those sequences that are not in the test set but that contain a subpath similar to *"nebrelsot"*. From Figure A.2 in Section A.5, the more steady behavior produced by the *standard* formula is visible, and this steadiness is manifested also across generalization.

Another important aspect involves the use of surprise as an explorative

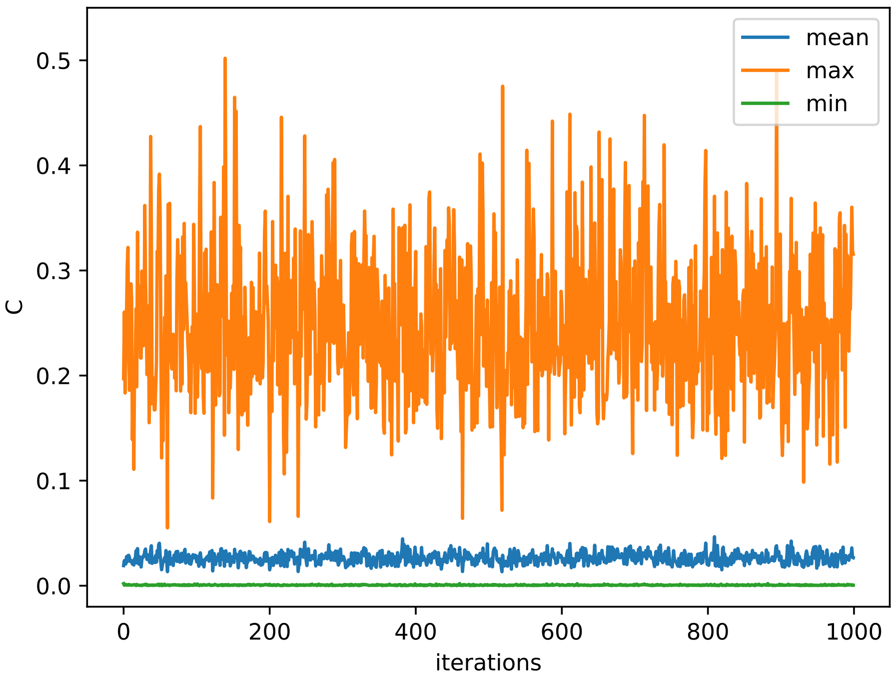
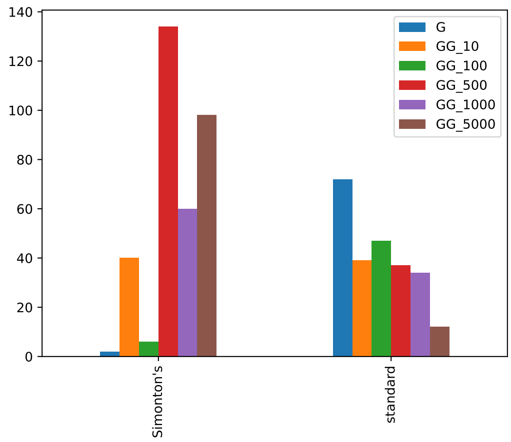


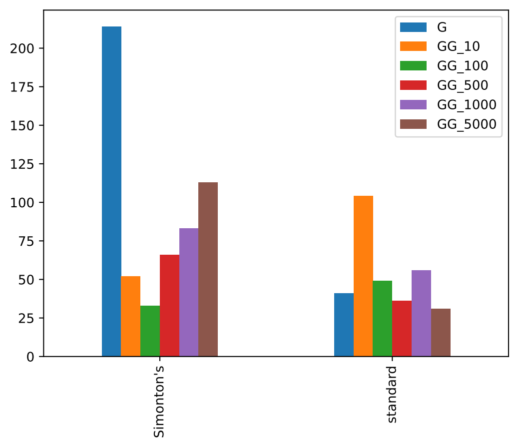
Figure 4.10: Results for convergence with generalized graph with 100 repetitions, using the *standard* formula for creativity.

factor compared to the role of decay and generalization. In Figure 4.11 are plotted the results of using the model with decay factors C = 20 and C = 50. In the case of C = 50, the memory stores composite (multi-) words that constrain the generation (see the effect on the results using the more convergent *standard* creativity formula). However, both generalization and surprise seem to ameliorate this tendency. In particular, surprise has opposite effects: where actual, right words are stored (C = 20), its effect is counterproductive (since negates the exploitation of good but already seen edges) while, in the case of C = 50, it solicits exploration of alternatives.

In conclusion, the effect of surprise, in the generative process, is to enhance exploration of the whole solution space (both the feasible and the unfeasible). However, its role is different from that of novelty (that deals



(a)



(b)

Figure 4.11: Comparison between models with diverse memory decays. Aggregate number of hits with (a) C = 20 and (b) C = 50.

with the similarity between produced artifacts, cf. Section 3.2) since surprise deals with the agent's knowledge about the utility of that solution.

On the other hand, generalization, which enables analogies, maintains a "natural" rate of variability regardless of the formula used in the process (Simonton's or *standard*). However, the effect of incubations, so the number of repetitions used to build the graph, has a twofold effect, similar to surprise: they enhance exploration that is useful in some cases but vain in other more exploitative ones. These aspects, however, need more in-depth studies and experiments.

4.5.3 Shallow Parsing

It has been suggested that in language acquisition humans form representations which are merely "good enough" for the communication task at hand [65]. That study, which analyzed event-related potentials, suggested that sentence meaning is constructed using simple heuristics arguing that this approach to language comprehension is similar to the use of heuristics for decision-making. In [133] the author maintain that shallow processing based on local information is a pervasive and widespread phenomenon in psycholinguistic research and the computational model devised in that work, the Chunk Based Learner, was made to emulate this behavior in children. As said, the CBL model[2] is the most similar one to the present approach, so we compared the two model in the shallow parsing task using the CHILDES [129] database,[3], which collects transcribed dialogues between mother and child; from the database we took Belfast, Lara, Manchester and Thomas transcripts (a total of 1236 files) and we used the CogComp Chunker [110], considered instead as the reference model, i.e., the baseline[4]. To mimic the constant decay of the PARSER memory used in [133] for comparisons, in these simulations, we use S = 1000 and C = 1000 in the formula 3.8.

Results

Figure 4.12 shows the result obtained. We calculated the F-Score using the formula in [133], for each model, taking the results of CogComp Chunker as the ground truth, where we considered: a true positive when the model placed a boundary in the right position, a false positive as a boundary placed when there should be not, and a false negative as a boundary not placed where there should be.

The major issue concerns applying the model to unsegmented or segmented data: that is, processing a stream of symbols in contrast to processing a sequence of words. The problem resides in how PARSER works (exploiting syllables as minimal units to be concatenated to form words). When neither memory nor TPs are exploitable for input processing, a random number of units is taken to parse (and to chunk) the input. However, this particular mechanism is conceived more for unsegmented data, especially within the shallow parsing task. This workaround is conceived to address all the other phenomena that may take part in this process (e.g., the agent's vigilance)

[2]Available at: `https://github.com/StewartMcCauley/CBL`

[3]Available at: `https://childes.talkbank.org/`

[4]urlhttps://github.com/CogComp/cogcomp-nlp/tree/master/chunker

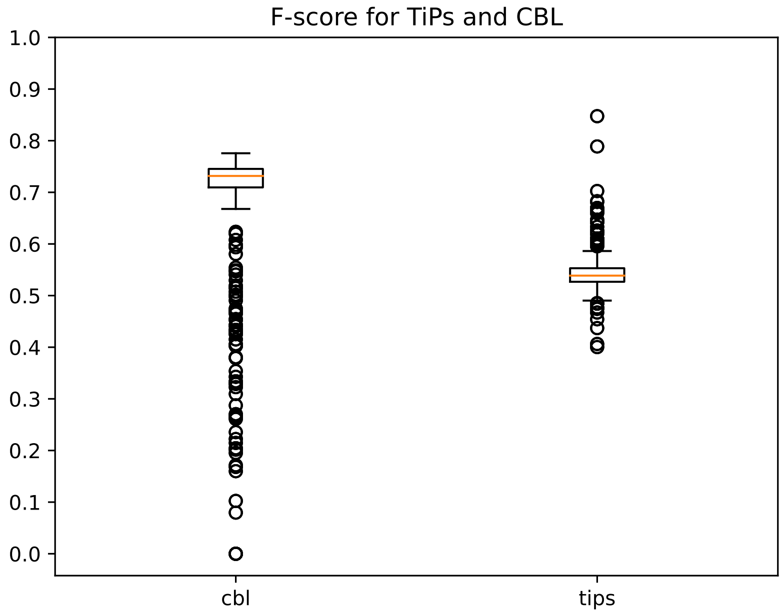


Figure 4.12: Shallow parsing results for CBL and TiPS

but it plays a role in the behavior of the system. See for instance Figure A.1 that shows the results obtained with the same set of parameters as those in Figure 4.12, but where instead of randomly picking a number in [1,3], a single unit (in this case a word) is chosen. Thus, this issue must be investigated more deeply.

	Min	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**
CBL	**0.0**	0.701	0.732	0.686	0.745	**0.776**
TiPS	**0.4**	0.527	0.539	0.542	0.553	**0.848**

Table 4.1: Summary results for CBL and TiPS

In addition, the two models are moderately inversely correlated (Pearson correlation coefficient = -0.5). Therefore, in general, they have a good performance in different cases (even if to varying degrees). The graph of sorted values (Figure 4.13) confirms these observations (see Min. and Max. columns in Table 4.1) and shows that TiPS follows the trend of CBL. In conclusion, TiPS does not perform badly, but in some respects, it may be preferable to or at least combined with, the CBL mechanism, to cover those cases where CBL fails.

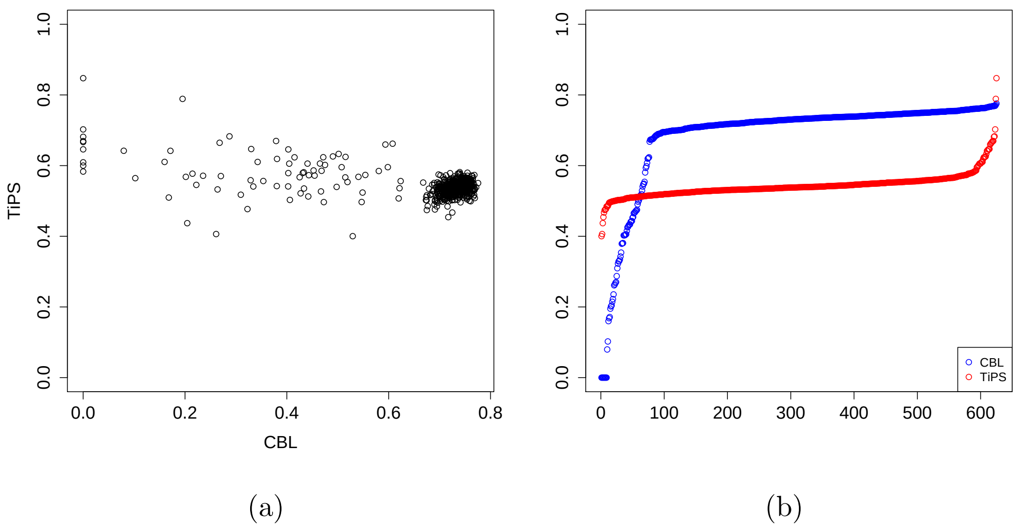
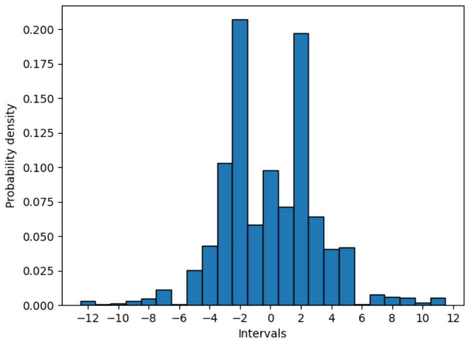


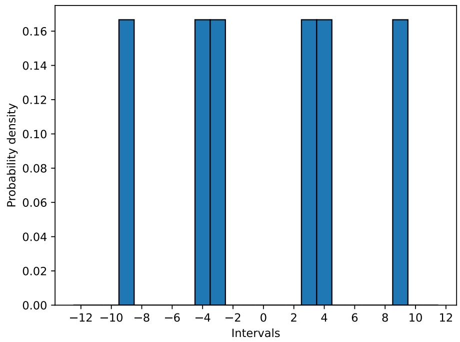
Figure 4.13: Performance of TiPS compared to CBL. F-Scores correlations (a) and sorted values (b).

4.6 An Application on Irish Songs

This section concerns a possible application of the previously discussed model as an instance to explore its potential in creative realms. Again, we have chosen the context of music: besides representing a more complex language (with respect to the previously investigated ones) in the creative domain, it is a compelling area of investigation for Statistical Learning and cognitive research. In addition, for this fulfillment, we introduce and discuss two utility functions to steer the generation toward the desired characteristics. Moreover, this is an example of how to use a utility unrelated to human or expert judgment. It is based on the intervals between adjacent notes, computed as a numerical difference in semitones, and calculates the frequency of these differences. For Irish music, used as the learning corpus, these intervals do not exceed an octave, so their probability distribution, in the input set, represented using histograms, remains in the range [-12,+12] (see Figure 4.14a).



(a)



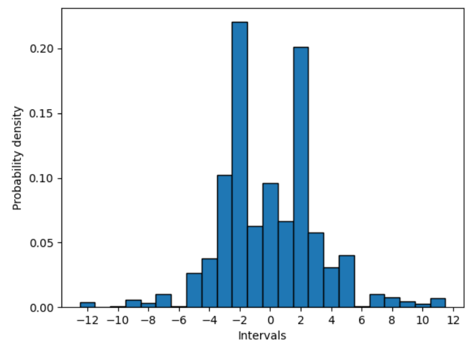
(b)

Figure 4.14: Histograms of Irish music: note intervals distribution for Irish input corpora (a) and allowed one by the utility function(b)

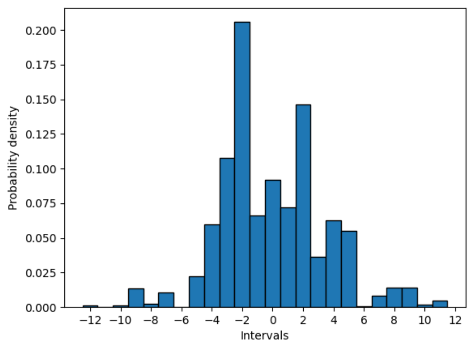
With the conceived utility, we want to induce a distribution of intervals different from that of the source corpus; specifically, in this case, we decided to favor melodies containing thirds (both minor and major) and sixths (major). To this aim, we take a reference histogram characterized by low frequencies for all the intervals, except for thirds and sixths (see Figure 4.14b). We defined two utilities: one as the similarity between the two vectors of interval frequencies using Euclidean distance, and the other using Kullback–Leibler divergence (KL) [105] by interpreting these histograms as probability distributions.

4.6.1 Results

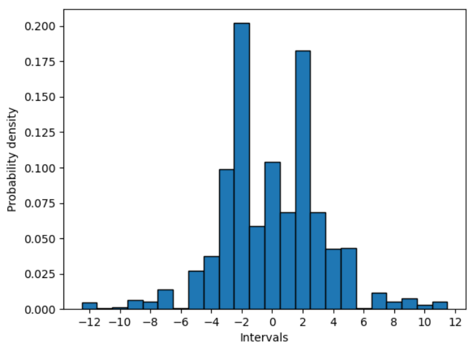
The model used to generate the results employ the AVG method for boundary discovery and C = 1000 for memory decay. The generalized graph was obtained using 1000 repetitions for its construction. The resulting histograms in Figure 4.15 and Figure 4.16 show that the KL utility works better because it has a more comprehensive view; with Euclidean distance, there is compensation between the various elements, while the KL one distinguishes these features.



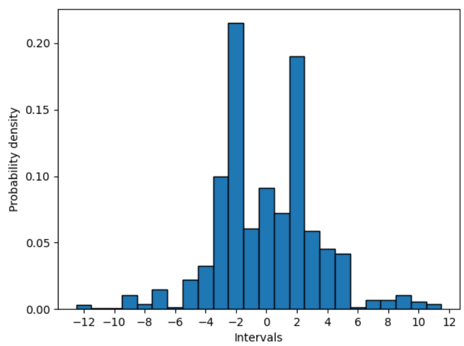
(a) Simonton's C + euclidean U



(b) Simonton's C + KL U



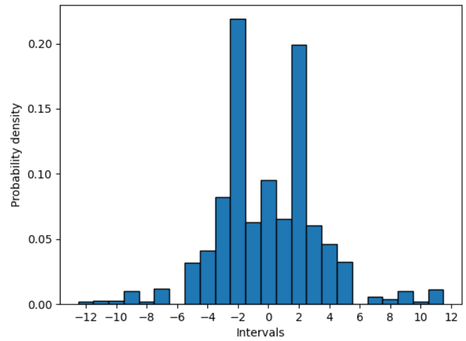
(c) standard C + euclidean U



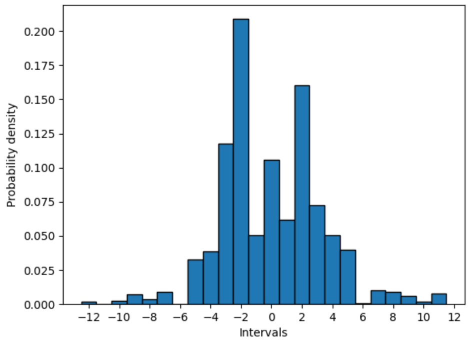
(d) standard C + KL U

Figure 4.15: Result for TPs graph with euclidean (a, c) and KL utilities (b, d), using Simonton's (a, b) or standard creativity (c, d).

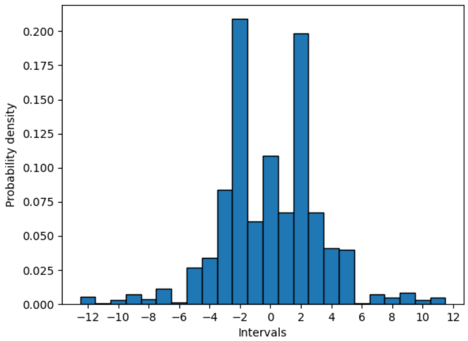
Concerning the quality of productions, a higher similarity to the musical structures of the original corpus is perceived in the music generated with the generalized graph (Figure 4.16). From the obtained histograms, a greater ability to approach the reference histogram is observed when Simonton's creativity is employed: this is most likely because surprise introduces diversification in the search.



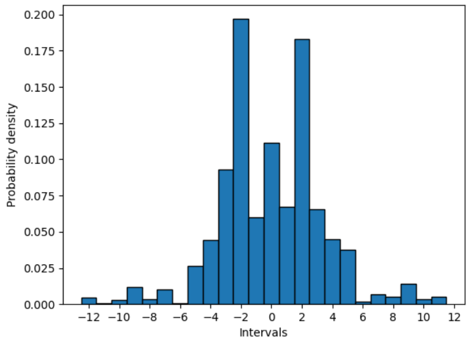
(a) Simonton's C + euclidean U



(b) Simonton's C + KL U



(c) standard C + euclidean U



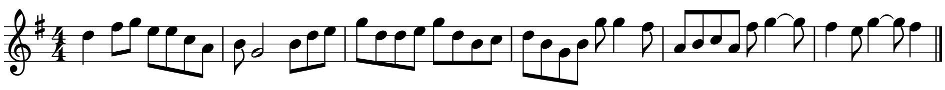
(d) standard C + KL U

Figure 4.16: Result for TPs graph with euclidean (a, c) and KL utilities (b, d), using Simonton's (a, b) or *standard* creativity (c, d).

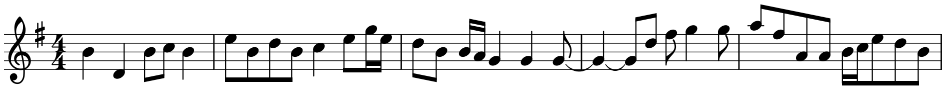
In conclusion, the surprise (1-v), in Simonton's formula, helps to diversify search, a characteristic of divergent behavior typical of creativity. However, even in such controlled experiments, the dynamics and interplay of surprise and utility, the effect of generalization, and their role in both the divergent and convergent phases that lead to creativity, are intricate. Hence, further extensive and in-depth studies will be needed to confirm this hypothesis.

4.6.2 Representative Examples

Below are shown the most creative sequences, one for each model and for each of the two functions of creativity (Simonton's and *standard*). These Irish series are generated using the KL function because it yields more significative values for the designed utility than the Euclidean one. We can observe that, at varying degrees across the generating graphs and creativity function, all the melodies contain arpeggios by thirds and a considerable amount of (both ascending and descending) thirds and sixths.

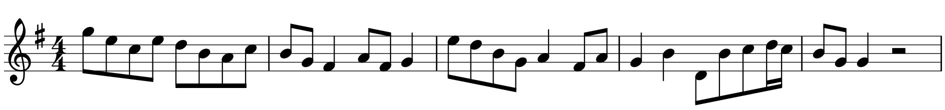


(a)



(b)

(c)



(d)

Figure 4.17: Representative examples of Irish songs produced with KL utility, using the TPs graph (a, b) and the generalized one with 500 repetitions (c, d). Both formulas of creativity were employed: Simonton's (a, c), and *standard* (c, d).

4.7 Discussion and Future Work

In this chapter, we presented the cognitive aspects related to the conceived computational model. We confronted and tested the present model with a related psychological model on a well-studied task on language acquisition in the literature. In addition, we tested the potential the model has in generating new sequences: both in the task of producing variations and novel solutions, and in the opposite one, to converge to task-appropriated ones. Finally, we applied this model to Irish music, discussing the produced quality of the model, and its potential in an artistic domain. Further experiments with human evaluation are planned to assess the quality of the pieces of music produced and to estimate to what extent evaluation by means of a creativity function correlates with human evaluation.

Compelling aspects regard mainly the role of surprise and that of generalization. These experiments highlighted also the role of short-term memory. That is, to facilitate the acquisition and subsequent computation of learned material, as well as that of attention. Moreover, besides chunk formation, ISL is essential for encoding sequentiality and could be the basis for more abstract and explicit knowledge formation: we explored some aspects of these mechanisms with the studies on generalization. These suggested continuing to investigate those cognitive aspects from a computational perspective meanwhile enhancing the model following these brain capacities (cf. [48]).

Despite considering cognitive issues, the present studies focused on computational analyses. This approach highlighted some aspects, at an algorithmic level, that could shed light and suggest refinements of the computational hypotheses that regard ISL, the formation of chunks, the convergent/divergent phases, and the role of surprise and generalization on creative generation. In the hope that these discussed aspects could favor cognitive research and interdisciplinary approaches to creativity, further computational experiments will continue following this vision.

Chapter 5

Final Remarks

The presented model could be useful in several ways, and it is general enough to lend itself well to many domains by construction. This characteristic is inherited mainly from the ISL mechanism it implements, which, in turn, is based on TPs computation. For instance, it can model subjectivity, and, even more importantly, developmental mechanisms, as the model is unsupervised and learns online and incrementally, so its evolution depends on the input (environmental) material. Moreover, ISL and TPs are being ever extensively studied in neuroscience and psychology; ISL mechanism is related to the hippocampus and other brain areas activity [191, 107]. It had been investigated via Reaction Rimes (RTs) and two-alternative forced choice (2-AFC) tasks [111, 224], and through Event-Related Potentials (ERPs), Electroencephalography (EEG) and magnetoencephalography (MEG) analyses [210, 112, 209, 113, 114], for example. In these respects, and besides the applicative realm, one ultimate end of this model is also that of being a cognitive-modeling tool for better understanding (or sharing research on) and simulating neurocognitive mechanisms (c.f., [36]). Hence, before wrapping up and concluding, the remainder of this chapter discusses ongoing work, further developments, and the possible connection to other major recent approaches in the cognitive literature.

5.1 Ongoing Work

A major challenge in the research field of robotics is developing robots that can engage in long-term human interaction. Therefore, a robot ought to be able to learn languages (e.g., spoken, movements, or implicit interactions) in an unsupervised manner and to be able to produce them, similarly, to become integrated with society. Therefore, it is crucially important to computationally understand how humans can learn and obtain skills through their autonomous development and (multimodal) interactions with the environment, in order to develop a robot that can emulate those behaviors. Hence, it is fundamentally important to understand systems that change dynamically in a constructive manner and build, and shape, their subjective experience.

Thus, besides the cognitive realm–which requires in-depth, structured experiments and resources–the aim is also to explore IT metrics both to analyze cognitive mechanisms–seeing the brain as an information processor–and to automatize these behaviors on robots or machines, for aiding machine creativity. These measures, as for TPs and ISL, are related to brain activity and a the same time are easily employable in robots.

5.1.1 IT Metrics Exploration

We are conducting more in-depth analyses on the possible exploitation of the metrics from Information Theory (IT) to validate and investigate–but also to computationally enhance– the creative production (e.g., Set-Based Complexity and Normalized Compression Distance). From a computational perspective, these metrics help automatize the processes of learning, generation, and validation (assessment) for creativity. For instance, they provide the means to apply Ritchies's criteria [179]: for attributing creativity to a computer program, or, on the other hand, if embedded in the program itself, to confer to the agent more (computational) autonomy.

Furthermore, IT provides a convenient framework to model brain activity, cognition, and consciousness [221, 35, 69]. For example, entropy has been recently adopted to associate brain activities with divergent thinking [199]. Furthermore, neurophysiological studies have revealed that musical sequences with higher entropy are learned through higher-order TPs, while series with lower entropy are learned by using lower-order TPs, and that certain brain regions perform entropy computations independently of TPs [43, 44, 45].

Essentially, these measures constitute a functional substrate to evaluate surprise, uncertainty, free energy, and other information-related metrics applicable (cf. KL-divergence), for example, to assist the choice of what cue is the more helpful or what mechanism (chunking or statistical computation) to adopt depending on the case. The same metrics could be employed to evaluate, in the same fashion, the complexity of the input and to describe the chosen behavior followed.

5.1.2 Robots at Theatre

We are currently embedding the model in a controller for humanoid robots (e.g., the Nao robot) to acquire, reproduce (imitate), and create sequences of movements (creativity). The entire pipeline includes the acquisition of human poses using Google MediaPipe, the use of inverse kinematics (IK), and motion planning software (MoveIt) to control the robot's movements, based on the (spatial) position of its end-effectors, acquired from the camera. The overall goal is to exploit the presented model for the learning and production of captured movements. With such a model, the robot, in addition to memorizing, can also learn the internal structure of these sequences (i.e., movements): discovering internal patterns and dependencies. Moreover, the identification of "constituent elements" enables the robot to recombine them and thus generate new sequences (through creative recombination). The goal is to test the robot on a theater stage in a two-stage process: first, improvisers are allowed to perform some movements in front of the robot to acquire such choreographies, then the robot is allowed to create new material for the actors to judge so that the robot learns aesthetic judgments to improve the creative outcome. Again, the goal has been twofold: in addition to the creative realm, these experiments are intended as a proof of concept for the more general context of implicit interactions in social robotics.

5.2 Further Developments

In this dissertation, we explored some essential mechanisms for encoding sequentially and discovering structured information employed by the brain for learning procedural knowledge. The aim was to describe a general-enough computational approach that should be able to account for computational, creative applications (conceived to be embedded in robots' controllers) and for computational investigations of brain functioning. However, every aspect mentioned throughout this work deserves more in-depth analyses, especially considering the related cognitive factors.

From a computational perspective, for example, the parameters of the model touched upon in the thesis could be subjected to adaptive procedures, such as Machine Learning algorithms that might exploit IT metrics (related to the input streams, or the complexity of the built graphs) to conceive systems that can work alongside humans (and artists). This relation accounts for another creative modality: the one emerging from the interaction between the human and the machine.

5.2.1 Enhancing the Sequential Module

Chunking mechanisms as PARSER, are sufficient to account for word discovery; the present studies confirmed the literature [159]. However, these results are true in language acquisition, where the structure of natural languages themselves follow a Zipfian distribution that facilitates the emergence of frequent words favoring chunking [119]; this is probably not true for all the sequential phenomena that could be addressed by this model (i.e., music and movements). In addition, the major tenet of this thesis is the encoding of sequentiality and the search for structure and hierarchical organization in sequential phenomena and broadly referring to human abilities, of procedural knowledge (i.e., knowing "how" not "what"). For this reason, TPs gain a fundamental role here: they represent the constituent part [27] for the emergence of sequential knowledge at increasing degrees of abstraction. Therefore, the aim is to enhance the (cognitive) mechanisms used to encode sequential phenomena: thus, following [48], the next step features the implementation of an algebraic module. If the generalization helped discover (form) classes, this step will enable a very powerful kind of abstraction that lead to pattern discovery of consecutive perceived stimuli, to capture the relationships between successive stimuli or stimulus categories. Consider, for instance, experiencing the succession of (symbolic) events such as —"totobu" "mimika" "paparo"—. It has been studied that few minutes of such exposure appears sufficient for a baby to recognize that all such sequences obey a similar pattern that may be denoted as AAB. When this pattern is violated, e.g., by an ABB item, the baby perceives this change [48]. As one can imagine, this module allows the implementation of a series of new mechanisms for creativity (and cognition) that leads to analogies, bisociations, and the discovery of new (higher-order and hidden) constraints on the given conceptual space. Moreover, as for generalization, all this reified knowledge (i) could serve as new attentional cues, (ii) could be used in other domains or conceptual spaces (cf. transfer of knowledge), and (iii) can be applied over features as well, in a multimodal scenario.

5.2.2 Attentional Mechanisms and Cues

A compelling aspect regards the exploitation of learned statistical cues in perception, which represented another motivation we had to add the memory mechanisms in our model. In the same way learned TPs could be used as cues for boundary-finding and segmentation, and in general, to discover structure in the input, several alternatives could be explored. In doing this, we envisage at least two possible scenarios. An option is exploiting TPs in a retrieval mode: with the current perceived chunk as a "retrieval cue" or, in the same way, following the traced, perceived path with TPS, the next expected unit could be "loaded" in memory (or using PARSER definition: the perception shaper) to enhance memory hits. That is a diverse approach, for example, to the other conventional possibility: that is, adopting an inferential, error-driven approach, where this mechanism could serve to generate expectations instead of attentional cues (see Section 5.3.2).

5.2.3 Effects of Generalization on Learning

In this dissertation, the efforts to investigate generalization had been focused more on the production side. However, they could be influential also for the learning phase in several ways. As stated above, the generalized entity and associations could be exploited by the attentional memory mechanism, in a retrieval or inferential mode. In addition, this higher-order, discovered information could be employed internally, by the agent, to ("mentally") evaluate or compare newly created sequences: as a utility function or as a different model (concerning the learned TPs) useful for comparisons–and potentially discover new constraints or patterns.

5.2.4 Segmented vs Unsegmented Input

The model can be applied both on unsegmented and segmented corpora. In the literature, however, there is a lack of computational experiments that address differences and similarities between segmented and unsegmented data; for instance, CBL currently works only on segmented input. Nevertheless, even if the underlying mechanisms might be the same (using chunking and statistical computations), the experiments conducted in this thesis suggest that the way (and the parameters) in which they could be applied seems very different. In this scenario, also the roles of TPs between symbols or between learned chunks (and their interplay) gain particular importance.

One hypothesis is that there could be some kind of additional adaptational mechanisms underlying this behavior. For example, when multiple cues were available, infants' looking behavior seemed to track with the strength of the strongest one [59]: this is in line with theories that suggest the brain acts to reduce uncertainty and to reduce error, as in predictive brain theories and Free Energy Principle (see Section 5.3.2).

5.2.5 Embracing Multimodality

An enthralling development, especially for the quality of the produced artifacts, concerns modeling multimodal (or multi-features) input. Even if ISL is considered a domain-general approach, some studies revealed modality and stimulus specificity [72]. These investigations raised the question of how statistical learning mechanisms could also account for these aspects. In particular, this issue concerns the differences in ISL between different modalities (audio, visual, sensorimotor) and between diverse stimuli in the same modality: for example, in music (auditory), the distinction between tone, pitch, or timbre.

Instead, within the computational perspective (applied CC), the modeling of multimodality enriches the possibilities for creative generations and their quality. Moreover, this approach is interesting for engineering applications in another creative context: that is, that of multimodal translation. In particular, by integrating statistical models derived from different contexts: i.e., stories and movements, which are specifically interesting for Noh drama.

5.2.6 Behavioural and Neuroscientific Experiments

A compelling experimental scenario of this model regards also the investigation of creativity in humans. There is, in fact, an ever-growing interest in the cognitive science of Creativity [13], yet, it is in an early stage, and research endeavors have often been undertaken separately by researchers of isolated sub-disciplines (e.g., neuroscience, psychology, education) [223]. The idea is to introduce novel experimental paradigms, building upon the studies in the literature that investigate ISL and combining behavioral, electrophysiological, and (most importantly) computational methods to examine human behavior along with the neural correlates of creativity, as, for example, in [233]. In this perspective, Computational models at least favor interdisciplinary collaborations, as recently reported in [232].

Moreover, if one considers embedding the model in robots or apps (e.g., for tablets), these studies (and applications) could also be carried out in educational (e.g., for cognitive enhancement) settings for humans [168]. For example, following the principles in [135], two unexplored possibilities (also in literature) refer to the use of these models to support learning-by-teaching tasks or serious games. All in all, this model is a generative system capable of unsupervised learning.

5.3 Connections to Other Approaches

TPs computation forms the basis for all the other statistical approaches. One of the main reasons, as stated previously, is the possibility to turn this recognition-based approach into an inference-based (or error-driven) one. The other key factor is how these TPs are computed: that is, through incrementally counting associations, and based solely on perceived occurrences (in contrast with batch descriptive statistics). Particularly, the relation to the free-energy principle and Schmidhuber's curiosity or intrinsic motivation [192, 194, 193], is in the use of information gain and other IT-related quantities (e.g., uncertainty, risk) in selection steps (features, cues, and learned higher-order knowledge) as attentional mechanisms, for instance, aiming at discovering "less-uncertain models" for encoding perceived phenomena. The exploration of various mechanisms gives the chance to learn and compare different aspects (using Bayesian model comparison, for example) of the phenomenon which is being learned, favoring the abstraction of the knowledge.

5.3.1 The Bayesian Approach

Under the veil of CC, the Bayesian approach provides a compelling option to address several aspects (e.g., surprise, commonsense, preferences, and beliefs) and represents an option to investigate statistical learning [200, 157, 60]. Moreover, unlike more traditional statistical inference, it preserves uncertainty and allows hypothesis testing. In a broad sense, if the present model and ISL account for the "how", the Bayesian framework could tell us what it requires and why it works [158]. If ISL models assume that the learner is passive and implicitly learns in an associative way in reaction to stimuli delivered by the environment, Bayesian models, on the other hand, describe how the learner should actively probe the environment to learn optimally. Along these lines, ISL and TPs recognition (using Markov Chains) form the proper basis to enable Bayesian approaches: that is, they could also be combined to represent different mechanisms (see [93] for an old but gold proposal). If TPs are learned implicitly with practice or exposition (like procedural memory), they could represent the prior knowledge for Bayesian modeling. Moreover, sequences can be stored at several levels of detail [48]: currently, the presented model build three levels of encoding. That is, it forms three levels of probability distributions: (i) TPs between symbols, learned from raw input; (ii) TPs between chunks, learned from segmented input that had passed through the memory; (iii) TPs between (form) classes, learned from the generalization phase. Therefore, a compelling case regards the employment of Bayesian models for *overhypothesis* [109][1] operating on those learned distributions. Moreover, this enables the distinction and combination of implicitly learned statistics (experience) with (more explicit) subjective belief and commonsense, for example. Thus, in this way, the Bayesian approach could enable Creative MetaCognition, which refers to the combination of self-knowledge and contextual knowledge used to make decisions about an

[1] Even if Goodman's definition of "overhypothesis" is more precise, we use the term more generally, as in the cited paper, to refer to any form of abstract knowledge that forms a hypothesis space (i.e., probability distributions) at a less abstract level.

agent's creative efforts and accomplishments [108] (cf. autonomy and intentionality in creativity [226]), representing a promising way to achieve genuine transformational (or meta-) creativity.

5.3.2 The Predictive Brain

Even if not considered, the present model is a solid math base for studying inference mechanisms. From this perspective, the brain not merely relies on bottom-up information but also adds expectations (cf. the Bayesian framework, discussed above). Several empirical studies have provided evidence for a brain that predicts its sensory input [171]. This perspective builds upon the notion that attention is not only a selective process necessary to avoid sensory overload but rather a mechanism that assigns weights to predictions and prediction errors [69]. Similarly, long-term memory, traditionally viewed as a device that stores past information, has been reconsidered as a system that provides the prerequisites for the simulation of future events [222]. Even in this case, sequence learning and ISL are two underlying processes for procedural learning [209], and procedural (implicit) memory is a type of long-term memory (involved in the performance of different actions and skills). In this perspective, the learned TPs graph stands for the long-term procedural memory learned from experience that could serve to explore such approaches.

Conclusions

In this dissertation, we explored computational mechanisms for generating creative sequences and investigating the cognitive basis of sequence learning and creative productions. In particular, the role of TPs as statistical means to favor interdisciplinarity, which enables cognitive studies in AI and CC settings. We explored some applications in evolutionary and artificial arts for robot movements and music. Then we discussed the cognitive mechanisms implemented, and we tested the model on characteristic tasks in language acquisition, which is the paradigmatic context for TPs for addressing human (and animal) learning and behavior. Finally, we discussed possible enhancements of the presented model that could improve creative procedures, on the one hand, and could aid the research of human behaviors, on the other. The last section described the connections with related cognitive approaches at the forefront of the literature.

In conclusion, we have presented some computational ideas on the nature of creativity in sequential processing as an ensemble of many different continuous processes combining the notions of experience, surprise, contextual value, novelty, and generalization. We believe this may open new possibilities for research on creativity in the domain of cognitive science. The benefits of adopting such an interdisciplinary approach are the positive implications for a broader cross-pollination across learning, robotic, and brain behaviors.

Appendix A

Supplementary Material

A.1 Thompson and Newport grammar

Classes					
A	B	C	D	E	F
kof	hox	jes	sot	fal	ker
daz	neb	rel	zor	taf	nav
mer	lev	tid	lum	rud	sib

Table A.1: Thompson and Newport's grammar. Sequences can have ABCDEF (baseline), ABCD, ABEF, and CDEF structure.

A.2 Algorithms

Algorithm 1: Pseudo code for GA

```
eval ← mono
for #iters do
    eval(pop)
    offspring, elite ← pop
    offspring crossover and mutation
    pop ← offspring + elite
    eval ← select_objective(pop)
    archive ← archive_assessment(elite)
```

Algorithm 2: Pseudo code for *select_objective*() function

```
eval = { mono          // the Markov Score
       { biobjective   // Pareto(MarkovScore(), Novelty())
prevFit ← bestFit
bestFit ← selBest(pop)
if eval == mono then
    if prevFit ≃ bestFit then
        counter ← counter - 1
        if counter == 0 then
            reset(counter)
            lastAvg ← avg(pop)
            eval ← biobjective
    else
        restart(counter)
else
    bestAvg ← avg(pop)
    if lastAvg ≃ bestAvg then
        eval ← mono
```

Algorithm 3: Pseudo code for the learning algorithm

```
foreach input sequence s do
    units_to_perceive ← rnd(1,3)
    while len(s) > 0 do
        units_list ← [] while len(s) > 0 & units_to_perceive > 0 do
            active_mem ← select unit above threshold in memory
            if s starts with a unit u in active_mem then
                // memory shaped perception
                next_unit ← u
            else if TPs founds a boundary in s at pos i then
                // found TPs cue
                next_unit ← s[0 : i]
            else
                // random segmentation
                next_unit ← s[0 : rnd(1, 3)]
            s ← s[0:len(next_unit)]
            units_to_perceive ← units_to_perceive - 1
            append next_unit to units_list
        whole_percept ← join(units_list)
        encode units_list and whole_percept in memory
        encode TPS between symbols in whole_percept
        encode TPS between chunks in units_list
        apply decay and interference
```

Algorithm 4: Pseudo code for the generalization step

```
initialize ggraph
form_classes ← cluster nodes using (in- and out-) SimRank for each
  pair of nodes in the graph of TPs between chunks
gen_paths ← generate n sequences with the graph of TPs between
  chunks
foreach gen_paths do
    start_node ← 'start'
    foreach chunk in gen_paths do
        end_node ← form_classes[chunk]
        if the arch (start_node, end_node) in ggraph then
            add 1 to weight of ggraph(start_node, end_node)
        else
            add (start_node, end_node) arch to ggraph

        start_node ← end_node
normalize the weight of outgoing edges for each node
return ggraph
```

A.3 Alternative results of shallow parsing

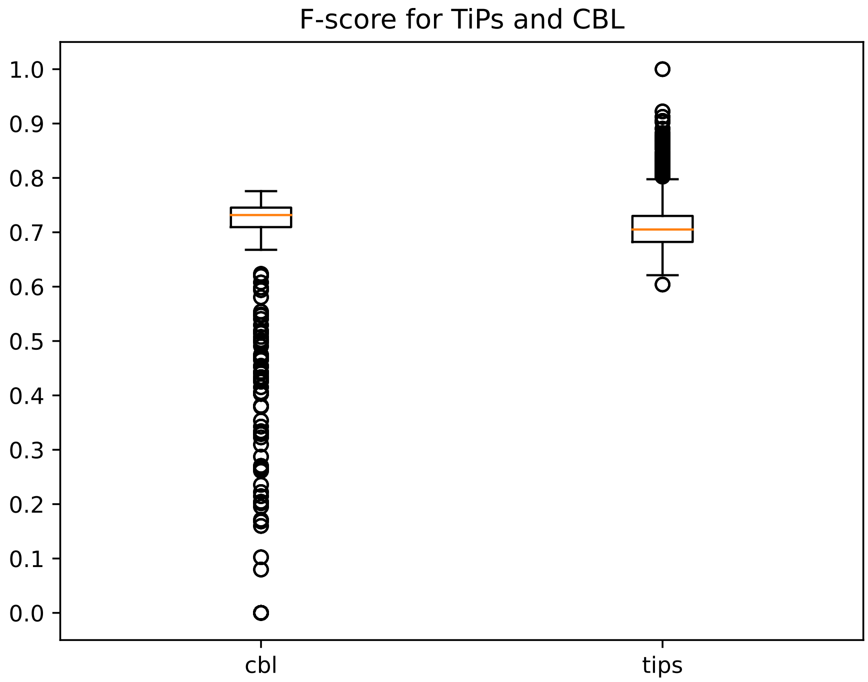


Figure A.1: Shallow parsing results for CBL and TiPS selecting a single word instead of employing a random choice.

Tables for divergent tests 1

C = 20							
graph:	G	GG					
		10	100	500	1000	5000	10000
ABCDEF + ABCD + ABEF + CDEF							
BRENT	13	6	14	18	15	13	15
AVG	14	42	43	43	35	39	38
FTPAVG	8	6	8	7	6	7	7
ABCDEF							
BRENT	16	6	22	15	15	16	16
AVG	18	19	22	25	24	25	25
FTPAVG	16	12	14	11	16	10	14

Table A.2: Results for C=20. Number of generated sequences that contain the sub-sequence *"nebrelsot"*

C = 50							
graph:	G	GG					
		10	100	500	1000	5000	10000
ABCDEF + ABCD + ABEF + CDEF							
BRENT	8	4	14	15	14	12	11
AVG	12	6	17	18	17	14	13
FTPAVG	10	15	20	17	20	19	19
ABCDEF							
BRENT	12	8	18	18	14	12	15
AVG	13	24	10	15	18	17	18
FTPAVG	11	4	13	20	18	16	18

Table A.3: Results for C=50. Number of generated sequences that contain the sub-sequence *"nebrelsot"*

C = 100							
graph:	**G**	**GG**					
		10	**100**	**500**	**1000**	**5000**	**10000**
ABCDEF + ABCD + ABEF + CDEF							
BRENT	2	5	5	6	9	5	8
AVG	5	23	14	12	14	16	16
FTPAVG	3	10	15	12	13	19	20
ABCDEF							
BRENT	9	11	6	6	6	5	8
AVG	11	21	18	14	14	17	19
FTPAVG	12	9	7	6	10	11	12

Table A.4: Results for C=100. Number of generated sequences that contain the sub-sequence *"nebrelsot"*

C = 500							
graph:	**G**	**GG**					
		10	**100**	**500**	**1000**	**5000**	**10000**
ABCDEF + ABCD + ABEF + CDEF							
BRENT	3	2	6	8	5	5	4
AVG	3	39	14	9	10	9	10
FTPAVG	3	8	9	11	12	11	10
ABCDEF							
BRENT	2	2	4	5	5	6	6
AVG	5	14	18	13	16	17	14
FTPAVG	3	9	10	8	8	8	8

Table A.5: Results for C=500. Number of generated sequences that contain the sub-sequence *"nebrelsot"*

C = 1000							
graph:	**G**	**GG**					
		10	**100**	**500**	**1000**	**5000**	**10000**
ABCDEF + ABCD + ABEF + CDEF							
BRENT	4	2	4	4	6	5	6
AVG	3	5	12	9	10	9	6
FTPAVG	2	26	9	12	10	9	10
ABCDEF							
BRENT	3	9	10	9	7	7	7
AVG	3	8	10	14	12	13	12
FTPAVG	3	10	8	11	12	8	8

Table A.6: Results for C=1000. Number of generated sequences that contain the sub-sequence *"nebrelsot"*

A.4 Tables for tests on divergent tests 2

TPs graph												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	229	0.57	243	219	223	0.57	241	212	201	0.6	214	183
AVG	239	0.55	252	231	214	0.58	222	200	186	0.6	207	158
FTPAVG	101	0.73	127	68	188	0.6	210	162	191	0.6	208	165

Table A.7: Results for TPs graph.

Generalized Graph (GG) with rep=10												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	157	0.33	249	92	160	0.11	253	98	218	0.08	260	168
AVG	138	0.4	244	63	179	0.17	225	138	126	0.3	186	63
FTPAVG	89	0.31	169	32	238	0.09	365	136	109	0.29	188	48

Table A.8: Results for generalized graph built with 10 repetitions.

Generalized Graph (GG) with rep=100												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	152	0.62	185	130	168	0.51	206	116	153	0.49	198	83
AVG	145	0.63	240	82	185	0.52	205	145	96	0.67	114	75
FTPAVG	84	0.72	132	52	176	0.49	201	160	121	0.64	156	67

Table A.9: Results for generalized graph built with 100 repetitions.

Generalized Graph (GG) with rep=500												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	149	0.66	193	104	176	0.59	207	134	143	0.6	190	120
AVG	147	0.68	262	79	175	0.58	206	145	115	0.63	145	83
FTPAVG	81	0.76	112	48	182	0.59	199	159	147	0.66	186	84

Table A.10: Results for generalized graph built with 500 repetitions.

Generalized Graph (GG) with rep=1000												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	164	0.61	199	134	169	0.62	190	124	115	0.68	130	99
AVG	147	0.66	253	83	182	0.58	228	124	134	0.66	200	100
FTPAVG	75	0.74	104	48	179	0.61	192	173	167	0.61	208	137

Table A.11: Results for generalized graph built with 1000 repetitions.

Generalized Graph (GG) with rep=10000												
C:	**20**				**50**				**100**			
	H	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**	**H**	**HR**	**M**	**m**
BRENT	155	0.64	193	117	185	0.6	243	121	146	0.67	182	120
AVG	136	0.65	227	86	185	0.59	217	133	126	0.69	158	91
FTPAVG	79	0.77	125	55	178	0.61	191	161	124	0.68	181	61

Table A.12: Results for generalized graph built with 10000 repetitions.

A.5 Additional results for convergence

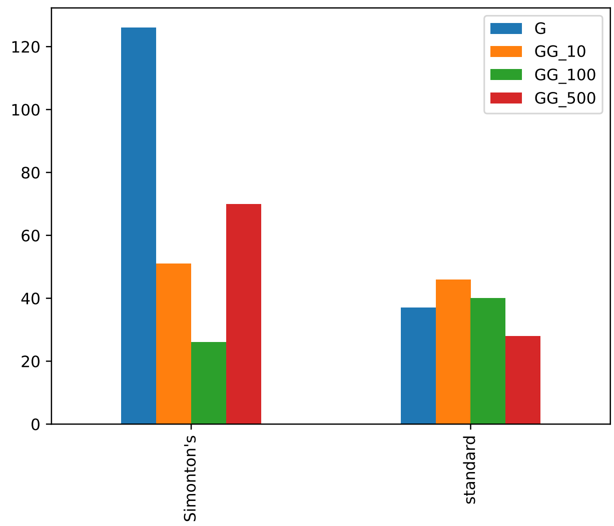
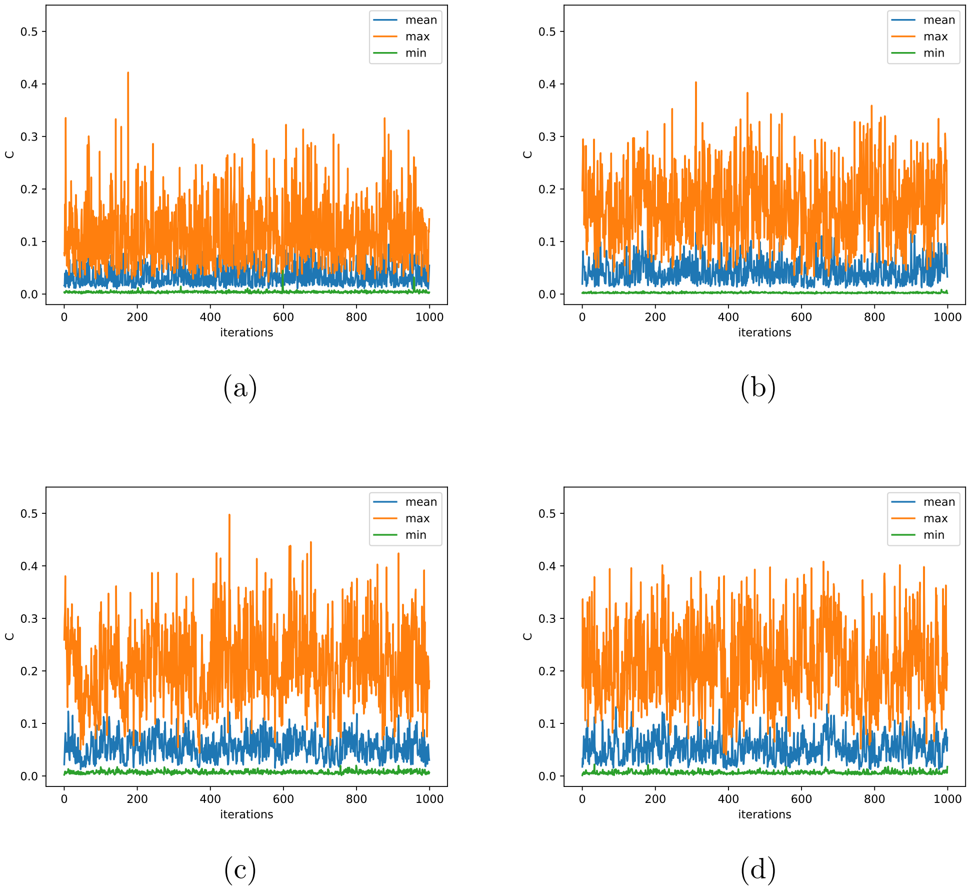


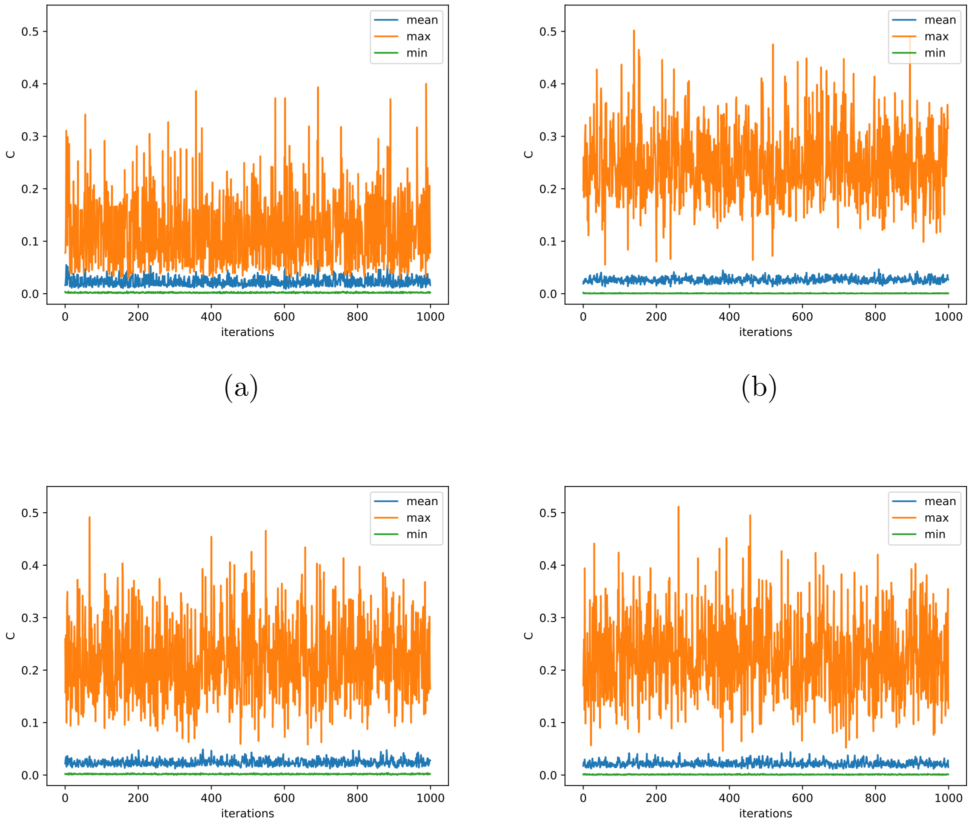
Figure A.2: Aggregate number of hits of both formulas for creativity



(a) (b)

(c) (d)

Figure A.3: Additional results of convergence tests for generalized graphs with repetitions (10,100,1000,10000) using Simonton's creativity to steer the generation



(a) (b)

(c) (d)

Figure A.4: Additional results of convergence tests for generalized graphs with repetitions (10,100,1000,10000) using *standard* creativity to steer the generation

Bibliography

[1] Y. Adini, Y. S. Bonneh, S. Komm, L. Deutsch, and D. Israeli. The time course and characteristics of procedural learning in schizophrenia patients and healthy individuals. *Frontiers in human neuroscience*, 9:475, 2015.

[2] M. Allen and K. J. Friston. From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 195(6):2459–2482, 2018.

[3] G. T. Altmann. 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(1711):20160060, 2017.

[4] C. Anderson, A. Eigenfeldt, and P. Pasquier. The generative electronic dance music algorithmic system (gedmas). In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2013.

[5] B. C. Armstrong, R. Frost, and M. H. Christiansen. The long road of statistical learning research: past, present and future, 2017.

[6] R. Aslin, E. Newport, and P. Reeder. The role of distributional information in linguistic category formation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 31 of number 31, 2009.

[7] R. N. Aslin and E. L. Newport. Distributional language learning: mechanisms and models of category formation. *Language Learning*, 64(s2):86–105, 2014.

[8] R. N. Aslin and E. L. Newport. Statistical learning: from acquiring specific items to forming general rules. *Current directions in psychological science*, 21(3):170–176, 2012.

[9] M. Barbaresi, S. Bernagozzi, and A. Roli. Robot choreographies: artificial evolution between novelty and similarity. In A. Finzi, A. Castellini, L. Buoncompagni, and S. Anzalone, editors, *Proceedings of the 7th Italian Workshop on Artificial Intelligence and Robotics (AIRO@AIxIA2020)*, pages 17–21, 2020.

[10] M. Barbaresi and A. Roli. Evolutionary music: statistical learning and novelty for automatic improvisation. In J. J. Schneider, M. S. Weyland, D. Flumini, and R. M. Füchslin, editors, *Artificial Life and Evolutionary Computation*, pages 172–183, Cham. Springer Nature Switzerland, 2022. ISBN: 978-3-031-23929-8.

[11] M. Barbaresi and A. Roli. Machine improvisation through generalized transition probability graphs, 2022.

[12] L. J. Batterink, K. A. Paller, and P. J. Reber. Understanding the neural bases of implicit and statistical learning. *Topics in cognitive science*, 11(3):482–503, 2019.

[13] R. E. Beaty, P. Seli, and D. L. Schacter. Network neuroscience of creative cognition: mapping cognitive mechanisms and individual differences in the creative brain. *Current opinion in behavioral sciences*, 27:22–30, 2019.

[14] H. Behrens. Constructivist approaches to first language acquisition. *Journal of Child Language*, 48(5):959–983, 2021.

[15] L. Benjamin, A. Fló, M. Palu, S. Naik, L. Melloni, and G. Dehaene-Lambertz. Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Biorxiv*, 2021.

[16] J. M. Berg. When silver is gold: forecasting the potential creativity of initial ideas. *Organizational Behavior and Human Decision Processes*, 154:96–117, 2019.

[17] J. Biles. Improvizing with genetic algorithms: genjam. In *Evolutionary Computer Music*, pages 137–169. Springer, 2007.

[18] J. Biles et al. Genjam: a genetic algorithm for generating jazz solos. In *ICMC*, volume 94, pages 131–137, 1994.

[19] C. Blum and A. Roli. Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM computing surveys*, 35(3):268–308, 2003.

[20] M. Boden. Artificial intelligence and natural man. *Synthese*, 43(3), 1980.

[21] M. A. Boden. Creativity in a nutshell. *Think*, 5(15):83–96, 2007.

[22] M. A. Boden. *The creative mind: Myths and mechanisms*. Routledge, 2004.

[23] C. R. Bowman, T. Iwashita, and D. Zeithamova. Tracking prototype and exemplar representations in the brain across learning. *elife*, 9, 2020.

[24] A. Brandt. Defining creativity: a view from the arts. *Creativity Research Journal*, 33(2):81–95, 2021.

[25] M. R. Brent. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34(1):71–105, 1999.

[26] P. F. Brown, V. J. Della Pietra, P. V. Desouza, J. C. Lai, and R. L. Mercer. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–480, 1992.

[27] I. Bybee. Sequentiality as the basis. *The evolution of language out of pre-language*, 53:109, 2002.

[28] J. Bybee. *Language, usage and cognition.* Cambridge University Press, 2010.

[29] D. T. Campbell. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review*, 67(6):380, 1960.

[30] F. Carnovalini and A. Rodà. Computational creativity and music generation systems: an introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3:14, 2020.

[31] P. F. Carvalho, C.-h. Chen, and C. Yu. The distributional properties of exemplars affect category learning and generalization. *Scientific reports*, 11(1):1–10, 2021.

[32] M. H. Christiansen. Implicit statistical learning: a tale of two literatures. *Topics in Cognitive Science*, 11(3):468–481, 2019.

[33] R. Cilibrasi and P. Vitányi. Clustering by compression. *IEEE Transactions on Information theory*, 51(4):1523–1545, 2005.

[34] A. Cleeremans and Z. Dienes. Computational models of implicit learning. *Cambridge handbook of computational psychology*:396–421, 2008.

[35] G. Collell and J. Fauquet. Brain activity and cognition: a connection from thermodynamics and information theory. *Frontiers in psychology*, 6:818, 2015.

[36] C. M. Conway. How does the brain learn environmental structure? ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, 112:279–299, 2020.

[37] C. M. Conway and M. H. Christiansen. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1):24, 2005.

[38] L. E. Coursey, R. T. Gertner, B. C. Williams, J. B. Kenworthy, P. B. Paulus, and S. Doboli. Linking the divergent and convergent processes of collaborative creativity: the impact of expertise levels and elaboration processes. *Frontiers in Psychology*, 10:699, 2019.

[39] M. Csikszentmihalyi. Flow and the psychology of discovery and invention. *HarperPerennial, New York*, 39, 1997.

[40] D. j. grossman's j.s. bach page. URL: `http://www.jsbach.net` (visited on 08/01/2022).

[41] D. Dahan and M. R. Brent. On the discovery of novel wordlike units from utterances: an artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, 128(2):165, 1999.

[42] T. Daikoku. Depth and the uncertainty of statistical knowledge on musical creativity fluctuate over a composer's lifetime. *Frontiers in Computational Neuroscience*, 13:27, 2019.

[43] T. Daikoku. Entropy, uncertainty, and the depth of implicit knowledge on musical creativity: computational study of improvisation in melody and rhythm. *Frontiers in Computational Neuroscience*, 12:97, 2018.

[44] T. Daikoku. Musical creativity and depth of implicit knowledge: spectral and temporal individualities in improvisation. *Frontiers in computational neuroscience*:89, 2018.

[45] T. Daikoku. Neurophysiological markers of statistical learning in music and language: hierarchy, entropy and uncertainty. *Brain sciences*, 8(6):114, 2018.

[46] T. Daikoku. Time-course variation of statistics embedded in music: corpus study on implicit learning and knowledge. *PLoS One*, 13(5):e0196493, 2018.

[47] T. Daikoku, G. A. Wiggins, and Y. Nagai. Statistical properties of musical creativity: roles of hierarchy and uncertainty in statistical learning. *Frontiers in Neuroscience*, 15:640412, 2021.

[48] S. Dehaene, F. Meyniel, C. Wacongne, L. Wang, and C. Pallier. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1):2–19, 2015.

[49] G. de Hollander, B. U. Forstmann, and S. D. Brown. Different ways of linking behavioral and neural data via computational cognitive models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(2):101–109, 2016.

[50] D. C. Dennett. Darwin's dangerous idea. *The Sciences*, 35(3):34–40, 1995.

[51] A. Dijksterhuis and T. Meurs. Where creativity resides: the generative power of unconscious thought. *Consciousness and cognition*, 15(1):135–146, 2006.

[52] H.-W. Dong, W.-Y. Hsiao, and Y.-H. Yang. Pypianoroll: open source python package for handling multitrack pianorolls. In *Late-Breaking Demos of the 19th International Society for Music Information Retrieval Conference*, 2018.

[53] Y. Du and J. E. Clark. New insights into statistical learning and chunk learning in implicit sequence acquisition. *Psychonomic bulletin & review*, 24(4):1225–1233, 2017. ISSN: 1531-5320. DOI: `10.3758/s13423-016-1193-4`.

[54] W. Duch. Idyot architecture-is this how minds operate? *Physics of Life Reviews*, 34:54–56, 2020.

[55] S. Edelman. *Computing the mind: How the mind really works*. Oxford University Press, 2008.

[56] A. Eigenfeldt and P. Pasquier. Realtime generation of harmonic progressions using controlled markov selection. In *Proceedings of ICCC-X-Computational Creativity Conference*, pages 16–25, 2010.

[57] J. L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2):195–225, 1991.

[58] A. D. Endress and A. Langus. Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive psychology*, 92:37–64, 2017.

[59] A. D. Endress, L. K. Slone, and S. P. Johnson. Statistical learning and memory. *Cognition*, 204:104346, 2020.

[60] L. C. Erickson and E. D. Thiessen. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37:66–108, 2015.

[61] S. Esser, C. Lustig, and H. Haider. What triggers explicit awareness in implicit sequence learning? implications from theories of consciousness. *Psychological Research*, 86(5):1442–1457, 2022.

[62] J. S. B. Evans. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59:255–278, 2008.

[63] S. Farrell and S. Lewandowsky. Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, 19(5):329–335, 2010.

[64] J. D. Fernández and F. Vico. Ai methods in algorithmic composition: a comprehensive survey. *Journal of Artificial Intelligence Research*, 48:513–582, 2013.

[65] F. Ferreira and N. D. Patson. The 'good enough'approach to language comprehension. *Language and linguistics compass*, 1(1-2):71–83, 2007.

[66] C. Francois and D. Schön. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21(10):2357–2365, 2011.

[67] M. C. Frank, J. B. Tenenbaum, and E. Gibson. Learning and long-term retention of large-scale artificial languages. *PloS one*, 8(1):e52500, 2013.

[68] R. M. French, C. Addyman, and D. Mareschal. Tracx: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological review*, 118(4):614, 2011.

[69] K. Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[70] C. D. Frith and U. Frith. Implicit and explicit processes in social cognition. *Neuron*, 60(3):503–510, 2008.

[71] R. Frost, B. C. Armstrong, and M. H. Christiansen. Statistical learning research: a critical review and possible new directions. *Psychological Bulletin*, 145(12):1128, 2019.

[72] R. Frost, B. C. Armstrong, N. Siegelman, and M. H. Christiansen. Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, 19(3):117–125, 2015.

[73] Q. Fu, H. Sun, Z. Dienes, and X. Fu. Implicit sequence learning of chunking and abstract structures. *Consciousness and cognition*, 62:42–56, 2018.

[74] L. Gabora. An analysis of the blind variation and selective retention theory of creativity. *Creativity Research Journal*, 23(2):155–165, 2011.

[75] D. Galas, M. Nykter, G. Carter, N. Price, and I. Shmulevich. Biological information as set-based complexity. *IEEE Transactions on Information Theory*, 56(2):667–677, 2010.

[76] A. C. Gilbert, V. J. Boucher, and B. Jemel. The perceptual chunking of speech: a demonstration using erps. *Brain Research*, 1603:101–113, 2015.

[77] K. J. Gilhooly. Incubation and intuition in creative problem solving. *Frontiers in psychology*, 7:1076, 2016.

[78] K. J. Gilhooly. *Incubation in problem solving and creativity: Unconscious processes*. Routledge, 2019.

[79] F. Gobet, P. C. Lane, S. Croker, P. C. Cheng, G. Jones, I. Oliver, and J. M. Pine. Chunking mechanisms in human learning. *Trends in cognitive sciences*, 5(6):236–243, 2001.

[80] F. Gobet, M. Lloyd-Kelly, and P. C. Lane. What's in a name? the multiple meanings of "chunk" and "chunking". *Frontiers in psychology*, 7:102, 2016.

[81] A. Goel. Looking back, looking ahead: symbolic versus connectionist ai. *AI Magazine*, 42(4):83–85, 2022.

[82] M. H. Goldstein, H. R. Waterfall, A. Lotem, J. Y. Halpern, J. A. Schwade, L. Onnis, and S. Edelman. General cognitive principles for learning structure in time and space. *Trends in cognitive sciences*, 14(6):249–258, 2010.

[83] R. L. Gómez, R. R. Bootzin, and L. Nadel. Naps promote abstraction in language-learning infants. *Psychological science*, 17(8):670–674, 2006.

[84] A. Goujon, A. Didierjean, and S. Poulet. The emergence of explicit knowledge from implicit learning. *Memory & cognition*, 42(2):225–236, 2014.

[85] S. Gu, M. Gao, Y. Yan, F. Wang, Y.-y. Tang, and J. H. Huang. The neural mechanism underlying cognitive and emotional processes in creativity. *Frontiers in Psychology*, 9:1924, 2018.

[86] O. Guest and A. E. Martin. How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4):789–802, 2021.

[87] J. P. Guilford. The nature of human intelligence. 1967.

[88] H. norbeck's `abc` tunes. URL: `http://www.norbeck.nu/abc/` (visited on 02/01/2022).

[89] J. Haar. Lassus, orlande de, 2021. DOI: `10/g5nj`. Grove Music Online, accessed on 12 Nov. 2021.

[90] Y. C. Han, K. D. Schmidt, E. Grandoit, P. Shu, C. P. McRobert, and P. J. Reber. Cognitive neuroscience of implicit learning. *The Cognitive Unconscious: The First Half Century*:37, 2022.

[91] D. M. Harrington. On the usefulness of "value" in the definition of creativity: a commentary. *Creativity research journal*, 30(1):118–121, 2018.

[92] J. F. Hay, B. Pelucchi, K. G. Estes, and J. R. Saffran. Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive psychology*, 63(2):93–106, 2011.

[93] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: the combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

[94] S. Hélie and R. Sun. Implicit cognition in problem solving. *The Psychology of Problem Solving: An Interdisciplinary Approach*, 2012.

[95] S. Hélie and R. Sun. Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological review*, 117(3):994, 2010.

[96] S. Hélie, R. Proulx, and B. Lefebvre. Bottom-up learning of explicit knowledge using a bayesian algorithm and a new hebbian learning rule. *Neural Networks*, 24(3):219–232, 2011.

[97] D. Herremans, C.-H. Chuan, and E. Chew. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)*, 50(5):1–30, 2017.

[98] P. Hitzler and M. Sarker. Neuro-symbolic ai= neural+ logical+ probabilistic ai. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, 342:173, 2022.

[99] D. R. Hofstadter. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* Basic books, 1995.

[100] A. Horner and D. Goldberg. *Genetic algorithms and computer-assisted music composition*, volume 51. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 1991.

[101] A. Hupbach, R. L. Gomez, R. R. Bootzin, and L. Nadel. Nap-dependent learning in infants. *Developmental science*, 12(6):1007–1012, 2009.

[102] E. S. Isbilen and M. H. Christiansen. Statistical learning of language: a meta-analysis into 25 years of research. *Cognitive Science*, 46(9):e13198, 2022.

[103] E. S. Isbilen, S. M. McCauley, and M. H. Christiansen. Individual differences in artificial and natural language statistical learning. *Cognition*, 225:105123, 2022.

[104] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 538–543, 2002.

[105] J. M. Joyce. Kullback-leibler divergence. In *International encyclopedia of statistical science*, pages 720–722. Springer, 2011.

[106] M. A. Kaliakatsos-Papakostas, M. G. Epitropakis, and M. N. Vrahatis. Weighted markov chain model for musical composer identification. In *European conference on the applications of evolutionary computation*, pages 334–343. Springer, 2011.

[107] E. A. Karuza, E. L. Newport, R. N. Aslin, S. J. Starling, M. E. Tivarus, and D. Bavelier. The neural correlates of statistical learning in a word segmentation task: an fmri study. *Brain and language*, 127(1):46–54, 2013.

[108] J. C. Kaufman, R. A. Beghetto, and C. Watson. Creative metacognition and self-ratings of creative performance: a 4-c perspective. *Learning and Individual Differences*, 51:394–399, 2016.

[109] C. Kemp, A. Perfors, and J. B. Tenenbaum. Learning overhypotheses with hierarchical bayesian models. *Developmental science*, 10(3):307–321, 2007.

[110] D. Khashabi, M. Sammons, B. Zhou, T. Redman, C. Christodoulopoulos, V. Srikumar, N. Rizzolo, L. Ratinov, G. Luo, Q. Do, et al. Cogcompnlp: your swiss army knife for nlp. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[111] R. Kim, A. Seitz, H. Feenstra, and L. Shams. Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience letters*, 461(2):145–149, 2009.

[112] A. Kóbor, K. Horváth, Z. Kardos, Á. Takács, K. Janacsek, V. Csépe, and D. Nemeth. Tracking the implicit acquisition of nonadjacent transitional probabilities by erps. *Memory & Cognition*, 47:1546–1566, 2019.

[113] A. Kóbor, Á. Takács, Z. Kardos, K. Janacsek, V. Csépe, and D. Nemeth. Erps differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biological psychology*, 135:180–193, 2018.

[114] S. Koelsch, T. Busch, S. Jentschke, and M. Rohrmeier. Under the hood of statistical learning: a statistical mmn reflects the magnitude of transitional probabilities in auditory sequences. *Scientific reports*, 6(1):1–11, 2016.

[115] A. Koestler. *The act of creation.* In *Brain Function, Volume IV: Brain Function and Learning.* D. B. Lindsley and A. A. Lumsdaine, editors. University of California Press, Berkeley, 1967, pages 327–346.

[116] O. Kolodny, S. Edelman, and A. Lotem. Evolved to adapt: a computational approach to animal innovation and creativity. *Current Zoology*, 61(2):350–368, 2015.

[117] O. Kolodny, A. Lotem, and S. Edelman. Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science*, 39(2):227–267, 2015.

[118] I. Koumarelas, A. Kroschk, C. Mosley, and F. Naumann. Experience: enhancing address matching with geocoding and similarity measure selection. *Journal of Data and Information Quality*, 10(2):1–16, 2018.

[119] C. Kurumada, S. C. Meylan, and M. C. Frank. Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, 127(3):439–453, 2013.

[120] C. Lacaux, T. Andrillon, C. Bastoul, Y. Idir, A. Fonteix-Galet, I. Arnulf, and D. Oudiette. Sleep onset is a creative sweet spot. *Science Advances*, 7(50):eabj5866, 2021.

[121] P. A. Lewis, G. Knoblich, and G. Poe. How memory replay in sleep boosts creative problem-solving. *Trends in cognitive sciences*, 22(6):491–503, 2018.

[122] M. Li, X. Chen, X. Li, B. Ma, and P. M. Vitányi. The similarity metric. *IEEE transactions on Information Theory*, 50(12):3250–3264, 2004.

[123] C. X. Ling and M. Marinov. A symbolic model of the nonconscious acquisition of information. *Cognitive Science*, 18(4):595–621, 1994.

[124] C. X. Ling and M. Marinov. Answering the connectionist challenge: a symbolic model of learning the past tenses of english verbs. *Cognition*, 49(3):235–290, 1993.

[125] L. Litman and A. S. Reber. *Implicit Cognition and Thought.* Cambridge University Press, 2005.

[126] C.-H. Liu and C.-K. Ting. Computational intelligence in music composition: a survey. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 1(1):2–15, 2016.

[127] Y. Liu, R. J. Dolan, Z. Kurth-Nelson, and T. E. Behrens. Human replay spontaneously reorganizes experience. *Cell*, 178(3):640–652, 2019.

[128] D. J. MacKay, D. J. Mac Kay, et al. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[129] B. MacWhinney. *The CHILDES project: The database*, volume 2. Psychology Press, 2000.

[130] M. Maheu, S. Dehaene, and F. Meyniel. Brain signatures of a multiscale process of sequence learning in humans. *elife*, 8:e41541, 2019.

[131] D. Mareschal and M. S. Thomas. Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation*, 11(2):137–150, 2007.

[132] D. Marr. *Vision: A computational investigation into the human representation and processing of visual information.* MIT press, 2010.

[133] S. M. McCauley and M. H. Christiansen. Language learning as language use: a cross-linguistic model of child language development. *Psychological review*, 126(1):1, 2019.

[134] V. Mekern, B. Hommel, and Z. Sjoerds. Computational models of creativity: a review of single-process and multi-process recent approaches to demystify creative cognition. *Current Opinion in Behavioral Sciences*, 27:47–54, 2019.

[135] A. N. Meltzoff, P. K. Kuhl, J. Movellan, and T. J. Sejnowski. Foundations for a new science of learning. *science*, 325(5938):284–288, 2009.

[136] A. Mesoudi. Cultural evolution. In *Cultural Evolution.* University of Chicago Press, 2011.

[137] T. Meulemans and M. Van der Linden. Implicit learning of complex information in amnesia. *Brain and Cognition*, 52(2):250–257, 2003.

[138] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[139] R. B. Millward. Models of concept formation. *Aptitude, learning, and instruction*:245–276, 2021.

[140] T. H. Mintz. Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30(5):678–686, 2002.

[141] I. Momennejad. Learning structures: predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, 32:155–166, 2020.

[142] C. Monroy, M. Meyer, S. Gerson, and S. Hunnius. Statistical learning in social action contexts. *PloS one*, 12(5):e0177261, 2017.

[143] B. Mor, S. Garhwal, and A. Kumar. A systematic literature review on computational musicology. *Archives of Computational Methods in Engineering*, 27(3), 2020.

[144] E. Muñoz, J. Cadenas, Y. Ong, and G. Acampora. Memetic music composition. *IEEE Transactions on Evolutionary Computation*, 20(1):1–15, 2014.

[145] Music excerpts for the paper titled "machine improvisation through generalized transition probability graphs" by m. barbaresi and a. roli. URL: `https://tinyurl.com/n4bc6x73` (visited on 11/01/2022).

[146] I. Nazli, A. Ferrari, C. Huber-Huber, and F. P. de Lange. Statistical learning is not error-driven. *bioRxiv*, 2022.

[147] B. Nettl, R. C. Wegman, I. Horsley, M. Collins, S. A. Carter, G. Garden, R. E. Seletsky, R. D. Levin, W. Crutchfield, J. Rink, P. Griffiths, and B. Kernfeld. Improvisation. In *Grove Music Online*. Oxford University Press, 2001. ISBN: 9781561592630. DOI: `10.1093/gmo/9781561592630.article.13738`. Retrieved 3 Oct. 2022.

[148] E. Norman and M. C. Price. Social intuition as a form of implicit learning: sequences of body movements are learned less explicitly than letter sequences. *Advances in Cognitive Psychology*, 8(2):121, 2012.

[149] D. Norris. Short-term memory and long-term memory are still different. *Psychological bulletin*, 143(9):992, 2017.

[150] R. C. O'reilly. Generalization in interactive networks: the benefits of inhibitory competition and hebbian learning. *Neural computation*, 13(6):1199–1241, 2001.

[151] K. Oberauer. Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):411, 2002.

[152] P. Olejarczuk, V. Kapatsinski, and R. H. Baayen. Distributional learning is error-driven: the role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard*, 4(s2), 2018.

[153] T. Opacic, C. Stevens, and B. Tillmann. Unspoken knowledge: implicit learning of structured human dance movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6):1570, 2009.

[154] G. Orbán, J. Fiser, R. N. Aslin, and M. Lengyel. Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, 105(7):2745–2750, 2008.

[155] F. Pachet. Interacting with a musical learning system: the continuator. In *International Conference on Music and Artificial Intelligence*, pages 119–132. Springer, 2002.

[156] M. Palmiero. The relationships between abstraction and creativity. In *Creativity and the Wandering Mind*, pages 73–90. Elsevier, 2020.

[157] L. Pearl and S. Goldwater. Statistical learning, inductive bias, and bayesian inference in language acquisition, 2016.

[158] A. Perfors and D. J. Navarro. What bayesian modelling can tell us about statistical learning: what it requires and why it works. *Statistical learning and language acquisition*, 1:383–408, 2012.

[159] P. Perruchet. What mechanisms underlie implicit statistical learning? transitional probabilities versus chunks in language learning. *Topics in cognitive science*, 11(3):520–535, 2019.

[160] P. Perruchet and S. Pacton. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in cognitive sciences*, 10(5):233–238, 2006.

[161] P. Perruchet and B. Poulin-Charronnat. Word segmentation: trading the (new, but poor) concept of statistical computation for the (old, but richer) associative approach, 2012.

[162] P. Perruchet and A. Vinter. Parser: a model for word segmentation. *Journal of memory and language*, 39(2):246–263, 1998.

[163] J. Pesnot Lerousseau and D. Schön. Musical expertise is associated with improved neural statistical learning in the auditory domain. *Cerebral Cortex*, 31(11):4877–4890, 2021.

[164] G. Pezzulo, F. Donnarumma, P. Iodice, D. Maisto, and I. Stoianov. Model-based approaches to active perception and control. *Entropy*, 19(6):266, 2017.

[165] S. T. Piantadosi. Zipf's word frequency law in natural language: a critical review and future directions. *Psychonomic bulletin & review*, 21(5):1112–1130, 2014.

[166] M. Pilichowski and W. Duch. Braingene: computational creativity algorithm that invents novel interesting names. In *2013 IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI)*, pages 92–99. IEEE, 2013.

[167] R. C. Pinkerton. Information theory and melody. *Scientific American*, 194(2):77–87, 1956.

[168] E. Plante and R. L. Gómez. Learning without trying: the clinical relevance of statistical learning. *Language, speech, and hearing services in schools*, 49(3S):710–722, 2018.

[169] J. Playford. *The Division violin: containing a collection of divisions upon several grounds for the treble-violin.* Henry Playford, London, UK, third edition, 1688.

[170] L. Polyanskaya. Cognitive mechanisms of statistical learning and segmentation of continuous sensory input. *Memory & Cognition*, 50(5):979–996, 2022.

[171] R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

[172] A. S. Reber. Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118(3):219, 1989.

[173] M. Redington, N. Chater, and S. Finch. Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive science*, 22(4):425–469, 1998.

[174] P. Reeder, E. Newport, and R. Aslin. Novel words in novel contexts: the role of distributional information in form-class category learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32 of number 32, 2010.

[175] P. A. Reeder, E. L. Newport, and R. N. Aslin. Distributional learning of subcategories in an artificial grammar: category generalization and subcategory restrictions. *Journal of memory and language*, 97:17–29, 2017.

[176] P. A. Reeder, E. L. Newport, and R. N. Aslin. From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cognitive psychology*, 66(1):30–54, 2013.

[177] G. Remillard and J. M. Clark. Implicit learning of first-, second-, and third-order transition probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2):483, 2001.

[178] M. Rhodes. An analysis of creativity. *The Phi delta kappan*, 42(7):305–310, 1961.

[179] G. Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007.

[180] M. Rohrmeier and P. Rebuschat. Implicit learning and acquisition of music. *Topics in cognitive science*, 4(4):525–553, 2012.

[181] A. Roli, A. Ligot, and M. Birattari. Complexity measures: open questions and novel opportunities in the automatic design and analysis of robot swarms. *Frontiers in Robotics and AI*, 6:130, 2019.

[182] M. Rucci, D. Bullock, and F. Santini. Integrating robotics and neuroscience: brains for robots, bodies for brains. *Advanced Robotics*, 21(10):1115–1129, 2007.

[183] M. A. Runco. Divergent thinking, creativity, and ideation. 2010.

[184] M. A. Runco. Introduction to the special issue: commemorating guilford's 1950 presidential address. *Creativity Research Journal*, 13(3-4):245–245, 2001.

[185] K.-J. Sachs and C. Dahlhaus. Counterpoint, 2001. DOI: `10/g5nk`. Grove Music Online, accessed on 12 Nov. 2021.

[186] E. Sadler-Smith. Wallas' four-stage model of the creative process: more than meets the eye? *Creativity Research Journal*, 27(4):342–352, 2015.

[187] J. R. Saffran, R. N. Aslin, and E. L. Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, 1996.

[188] J. R. Saffran and N. Z. Kirkham. Infant statistical learning. *Annual review of psychology*, 69:181–203, 2018.

[189] C. Santolin and J. R. Saffran. Constraints on statistical learning across species. *Trends in Cognitive Sciences*, 22(1):52–63, 2018.

[190] P. Sarkar, A. Chakrabarti, and J. Gero. Studying engineering design creativity. In *Workshop on Studying Design Creativity*, 2008.

[191] A. Schapiro and N. Turk-Browne. Statistical learning. *Brain mapping*, 3:501–506, 2015.

[192] J. Schmidhuber. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2):173–187, 2006.

[193] J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE transactions on autonomous mental development*, 2(3):230–247, 2010.

[194] J. Schmidhuber. Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity & creativity. In *International conference on discovery science*, pages 26–38. Springer, 2007.

[195] J. F. Schwab, K. D. Schuler, C. M. Stillman, E. L. Newport, J. H. Howard Jr, and D. V. Howard. Aging and the statistical learning of grammatical form classes. *Psychology and Aging*, 31(5):481, 2016.

[196] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(1,2):379–423, 623–656, 1948.

[197] I. Shapiro and M. Huber. Markov chains for computer music generation. *Journal of Humanistic Mathematics*, 11(2):167–195, 2021.

[198] B. E. Sherman, K. N. Graves, and N. B. Turk-Browne. The prevalence and importance of statistical learning in human cognition and behavior. *Current opinion in behavioral sciences*, 32:15–20, 2020.

[199] L. Shi, R. E. Beaty, Q. Chen, J. Sun, D. Wei, W. Yang, and J. Qiu. Brain entropy is associated with divergent thinking. *Cerebral cortex*, 30(2):708–717, 2020.

[200] N. Siegelman, L. Bogaerts, B. C. Armstrong, and R. Frost. What exactly is learned in visual statistical learning? insights from bayesian modeling. *Cognition*, 192:104002, 2019.

[201] D. K. Simonton. Creative ideas and the creative process: good news and bad news for the neuroscience of creativity. *The Cambridge handbook of the neuroscience of creativity*:9–18, 2018.

[202] D. K. Simonton. Creative problem solving as sequential bvsr: exploration (total ignorance) versus elimination (informed guess). *Thinking Skills and Creativity*, 8:1–10, 2013.

[203] D. K. Simonton. Creative thought as blind variation and selective retention: why creativity is inversely related to sightedness. *Journal of Theoretical and Philosophical Psychology*, 33(4):253, 2013.

[204] D. K. Simonton. Creativity and discovery as blind variation: campbell's (1960) bvsr model after the half-century mark. *Review of General Psychology*, 15(2):158–174, 2011.

[205] D. K. Simonton. Creativity as blind variation and selective retention: is the creative process darwinian? *Psychological Inquiry*:309–328, 1999.

[206] D. K. Simonton. Defining creativity: don't we also need to define what is not creative? *The Journal of Creative Behavior*, 52(1):80–90, 2018.

[207] D. K. Simonton. On praising convergent thinking: creativity as blind variation and selective retention. *Creativity Research Journal*, 27(3):262–270, 2015.

[208] D. K. Simonton. The blind-variation and selective-retention theory of creativity: recent developments and current status of bvsr. *Creativity Research Journal*:1–20, 2022.

[209] P. Simor, Z. Zavecz, K. Horváth, N. Éltető, C. Török, O. Pesthy, F. Gombos, K. Janacsek, and D. Nemeth. Deconstructing procedural memory: different learning trajectories and consolidation of sequence and statistical learning. *Frontiers in Psychology*, 9:2708, 2019.

[210] A. P. Soares, F.-J. Gutiérrez-Domínguez, M. Vasconcelos, H. M. Oliveira, D. Tomé, and L. Jiménez. Not all words are equally acquired: transitional probabilities and instructions affect the electrophysiological correlates of statistical learning. *Frontiers in Human Neuroscience*, 14:577991, 2020.

[211] Z. Solan, D. Horn, E. Ruppin, and S. Edelman. Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences*, 102(33):11629–11634, 2005.

[212] P. T. Sowden, A. Pringle, and L. Gabora. The shifting sands of creative thinking: connections to dual-process theory. In *Insight and Creativity in Problem Solving*, pages 40–60. Routledge, 2019.

[213] T. C. Stewart. Notes for the development of a philosophy of computational modelling. *Carleton University Cognitive Science, Tech. Rep*, 2005.

[214] T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh. Symbol emergence in robotics: a survey. *Advanced Robotics*, 30(11-12):706–728, 2016.

[215] T. Taniguchi, E. Ugur, M. Hoffmann, L. Jamone, T. Nagai, B. Rosman, T. Matsuka, N. Iwahashi, E. Oztop, J. Piater, et al. Symbol emergence in cognitive developmental systems: a survey. *IEEE transactions on Cognitive and Developmental Systems*, 11(4):494–516, 2018.

[216] H. S. Terrace. Chunking and serially organized behavior in pigeons, monkeys and humans. *Avian visual cognition*, 2001.

[217] A. Tessari, N. Canessa, M. Ukmar, and R. I. Rumiati. Neuropsychological evidence for a strategic control of multiple routes in imitation. *Brain*, 130(4):1111–1126, 2007.

[218] E. D. Thiessen, A. T. Kronstein, and D. G. Hufnagle. The extraction and integration framework: a two-process account of statistical learning. *Psychological bulletin*, 139(4):792, 2013.

[219] S. P. Thompson and E. L. Newport. Statistical learning of syntax: the role of transitional probability. *Language learning and development*, 3(1):1–42, 2007.

[220] H.-K. Tien, B.-L. Chang, and Y.-K. Kuo. Does experience stimulate or stifle creativity? *European Journal of Innovation Management*, 2018.

[221] G. Tononi, M. Boly, M. Massimini, and C. Koch. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461, 2016.

[222] S. Trapp, T. Parr, K. Friston, and E. Schröger. The predictive brain must have a limitation in short-term memory capacity. *Current Directions in Psychological Science*, 30(5):384–390, 2021.

[223] J. Vallverdú and A. Sans Pinillos. The foundations of creativity: human inquiry explained through the neuro-multimodality of abduction. *Handbook of abductive cognition. Cham: Springer*, 2022.

[224] M. Van Witteloostuijn, I. Lammertink, P. Boersma, F. Wijnen, and J. Rispens. Assessing visual statistical learning in early-school-aged children: the usefulness of an online reaction time measure. *Frontiers in Psychology*, 10:2051, 2019.

[225] Various authors. *A collection of the Most Celebrated Irish Tunes Proper for the Violin, German Flute or Hautboy*. John and William Neal, Dublin, Ireland, 1724.

[226] D. Ventura. Autonomous intentionality in computationally creative systems. In *Computational creativity*, pages 49–69. Springer, 2019.

[227] A. S. Vincent, B. P. Decker, and M. D. Mumford. Divergent thinking, intelligence, and expertise: a test of alternative models. *Creativity research journal*, 14(2):163–178, 2002.

[228] A. Vinhas, F. Assunção, J. Correia, A. Ekárt, and P. Machado. Fitness and novelty in evolutionary art. In *International Conference on Computational Intelligence in Music, Sound, Art and Design*, pages 225–240. Springer, 2016.

[229] R. W. Weisberg. Expertise, nonobvious creativity, and ordinary thinking in edison and others: integrating blindness and sightedness. *Psychology of Aesthetics, Creativity, and the Arts*, 9(1):15, 2015.

[230] E. Wilf. Semiotic dimensions of creativity. *Annual Review of Anthropology*, 43(1):397–412, 2014.

[231] L. Wittkuhn, S. Chien, S. Hall-McMaster, and N. W. Schuck. Replay in minds and machines. *Neuroscience & Biobehavioral Reviews*, 129:367–388, 2021.

[232] O. A. Wudarczyk, M. Kirtay, A. K. Kuhlen, R. Abdel Rahman, J.-D. Haynes, V. V. Hafner, and D. Pischedda. Bringing together robotics, neuroscience, and psychology: lessons learned from an interdisciplinary project. *Frontiers in Human Neuroscience*, 15:630789, 2021.

[233] I. Zioga, P. M. Harrison, M. T. Pearce, J. Bhattacharya, and C. D. B. Luft. From learning to creativity: identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. *NeuroImage*, 206:116311, 2020.

[234] W. Zuidema, R. M. French, R. G. Alhama, K. Ellis, T. J. O'Donnell, T. Sainburg, and T. Q. Gentner. Five ways in which computational modeling can help advance cognitive science: lessons from artificial grammar learning. *Topics in cognitive science*, 12(3):925–941, 2020.