MCCXXII

Administrative Office: University of Padova

Department of General Psychology

Doctoral School in Psychological Sciences

XXVIII CYCLE

Learning Regularities from the Visual World

School Director: Professor Francesca Peressotti

Supervisor: Professor Lucia Regolin

Doctoral Student: Chiara Santolin

To Renzo, Lucia and Davide,
without whom I could not have come this far

To Mauro,
my dear colleague. In loving memory

TABLE OF CONTENTS

ABSTRACT

Patterns of visual objects, streams of sounds, and spatiotemporal events are just a few examples of the structures present in a variety of sensory inputs. Amid such variety, numerous regularities can be found. In order to handle the sensory processing, individuals of each species have to be able to rapidly track these regularities. Statistical learning is one of the principal mechanisms that enable to track patterns from the flow of sensory information, by detecting coherent relations between elements (e.g., A predicts B). Once relevant structures are detected, learners are sometimes required to generalize to novel situations. This process can be challenging since it demands to abstract away from the surface information, and extract structures from previously-unseen stimuli. Over the past two decades, researchers have shown that statistical learning and generalization operate across domains, modalities and species, supporting the generality assumption. These mechanisms in fact, play a crucial role in organizing the sensory world, and developing representation of the environment.

But when and how do organisms begin to track and generalize patterns from the environment? From the overall existing literature, very little is known about the roots these mechanisms. The experiments described in this thesis were all designed to explore whether statistical learning and generalization of *visual* patterns are fully available at birth, using the newborn domestic chick (*Gallus gallus*) as animal model. This species represents an excellent developmental model for the study of the ontogeny of several cognitive traits because it can be tested soon after hatching, and allows complete manipulation of pre- and post-natal experience.

In Chapter 2, four statistical learning experiments are described. Through learning-by-exposure, visually-naive chicks were familiarized to a computer-presented stream of objects defined by a statistical structure; in particular, transitional (conditional) probabilities linked together sequence elements (e.g., the cross predicts the circle 100% of the times). After exposure, the familiar structured sequence were compared to a random presentation (Experiment 1) or a novel, structured combination (Experiment 2) of the familiar shapes. Chicks successfully differentiated test sequences in both experiments. One relevant aspect of these findings is that the learning process is unsupervised. Despite the lack of reinforcement, the mere exposure to the statistically-defined input was sufficient to obtain a significant learning effect.

Two additional experiments have been designed in order to explore the complexity of the patterns that can be learned by this species. In particular, the aim of Experiments 3 and 4 was to investigate chicks' ability to discriminate subtle differences of distributional properties of the stimuli. New sequences have been created; the familiar one was formed by a pairs of shapes that always appear in that order whereas the unfamiliar stimulus was formed by shapes spanning the boundaries across familiar pairs (part-pairs). Unfamiliar part-pairs were indeed created by joining the last element of a familiar pair and the first element of another (subsequent) familiar pair. The key difference among pairs and part-pairs lied on the probabilistic structure of the two: being formed by the union of two familiar elements, part-pairs were experienced during familiarization but with a lower probability. In order to distinguish test sequences, chicks needed to detect a very small difference in conditional probability characterizing the two stimuli. Unfortunately, the animals were unable to differentiate test sequences when formed by 8 (Experiment 3) or

6 (Experiment 4) elements. My final goal would have been to discover whether chicks are effectively able to pick up transitional probabilities or whether they simply track frequencies of co-occurrence. In Experiments 1 and 2, since the frequency of appearance of each shape was balanced across stimuli, it was impossible to tell if chicks detected transitional probabilities (e.g., X *predicts* Y) or frequencies of co-occurrence (e.g., X and Y *co-occur* together, but any predictive relation characterize them) among elements. However, since the animals did not succeed in the first task, being unable to discriminate pairs vs. part-pairs, data are inconclusive as regards to this issue. Possible explanations and theoretical implications of these results are provided in the final chapter of this thesis.

In Chapter 3, the two studies described were aimed at testing newborn chicks' capacities of generalization of patterns presented as stings of visual tokens. For instance, the pattern AAB can be defined as "two identical items (AA) followed by another one, different from the formers (B)". Patterns were presented as triplets of simultaneously-visible shapes, arranged according to AAB, ABA (Experiment 5), ABB and BAA (Experiment 6). Using a training procedure, chicks were able to recognize the trained regularity when compared to another (neutral) regularity (for instance, AAB displayed as cross-cross-circle vs. ABA displayed as cross-circle-cross). Chicks were also capable of generalizing these patterns to novel exemplars composed of previously-unseen elements (AAB vs. ABA implemented by hourglass-hourglass-arrow vs. hourglass-arrow-hourglass).

A subsequent study (Experiment 6) was aimed at verifying whether the presence/absence of contiguous reduplicated elements (in <u>AA</u>B but not in ABA) may have facilitated

learning and generalization in previous task. All regularities comprised an adjacent repetition that gave the triplets asymmetrical structures (AAB vs. ABB and AAB vs. BAA). Chicks discriminated pattern-following and pattern-violating novel test triplets instantiating all regularities employed in the study, suggesting that the presence/absence of an adjacent repetition was not a relevant cue to succeed in the task.

Overall, the present research provides new data of *statistical learning* and *generalization* of visual regularities in a newborn animal model, revealing that these mechanisms fully operate at the very beginning of life. For what concerns statistical learning, day-old chicks performed better than neonates but similar to human infants. As regards to generalization, chicks' performance is consistent to what shown by neonates in the linguistic domain. These findings suggest that newborn chicks may be predisposed to track visual regularities in their postnatal environment. Despite the very limited previous experience, after a mere exposure to a structured input or a 3-days training session, significant learning and generalization effects have been obtained, pointing to the presence of early predispositions serving the development of these cognitive abilities.

RIASSUNTO

Il mondo sensoriale è composto da un insieme di regolarità. Sequenze di sillabe e note musicali, oggetti disposti nell'ambiente visivo e sequenze di eventi sono solo alcune delle tipologie di *pattern* caratterizzanti l'input sensoriale. La capacità di rilevare queste regolarità risulta fondamentale per l'acquisizione di alcune proprietà del linguaggio naturale (ad esempio, la sintassi), l'apprendimento di sequenze di azioni (ad esempio, il linguaggio dei segni), la discriminazione di eventi ambientali complessi come pure la pianificazione del comportamento. Infatti, rilevare regolarità da una molteplicità di eventi permette di anticipare e pianificare azioni future, aspetti cruciali di adattamento all'ambiente. Questo meccanismo di apprendimento, riportato in letteratura con il nome di *statistical learning,* consiste nella rilevazione di distribuzioni di probabilità da input sensoriali ovvero, relazioni di dipendenza tra i suoi diversi componenti (ad esempio, X predice Y). Come illustrato nell capitolo introduttivo della presente ricerca, nonostante si tratti di uno dei meccanismi responsabili dell'apprendimento del linguaggio naturale umano, lo *statistical learning* non sembra essersi evoluto in modo specifico per servire questa funzione. Tale meccanismo rappresenta un processo cognitivo generale che si manifesta in diversi domini sensoriali (acustico, visivo, tattile), modalità (temporale oppure spaziale-statico) e specie (umana e non-umane). La rilevazione di *pattern* gioca quindi un ruolo fondamentale nell'elaborazione dell'informazione sensoriale, necessaria ad una corretta rappresentazione dell'ambiente. Una volta apprese le regolarità e le strutture presenti nell'ambiente, gli organismi viventi devono saper generalizzare tali strutture a stimoli nuovi da un punto di vista percettivo, ma rappresentanti le stesse

regolarità. L'aspetto cruciale della generalizzazione è quindi la capacità di riconoscere una regolarità familiare anche quando implementata da nuovi stimoli. Anche il processo di generalizzazione ricopre un ruolo fondamentale nell'apprendimento della sintassi del linguaggio naturale umano. Ciò nonostante, si tratta di un meccanismo dominio-generale e non specie-specifico.

Ciò che non risultava chiaro dalla letteratura era l'ontogenesi di entrambi i meccanismi, specialmente nel dominio visivo. In altre parole, non era chiaro se le abilità di *statistical learning* e generalizzazione di strutture visive fossero completamente sviluppate alla nascita. Il principale obbiettivo degli esperimenti condotti in questa tesi era quindi quello di approfondire le origini di *visual statistical learning* e generalizzazione, tramite del pulcino di pollo domestico (*Gallus gallus*) come modello animale. Appartenendo ad una specie precoce, il pulcino neonato è quasi completamente autonomo per una serie di funzioni comportamentali diventando il candidato ideale per lo studio dell'ontogenesi di diverse abilità percettive e cognitive. La possibilità di essere osservato appena dopo la nascita, e la completa manipolazione dell'ambiente pre- e post- natale (tramite schiusa e allevamento in condizioni controllate), rende il pulcino un'ottimo modello sperimentale per lo studio dell'apprendimento di regolarità.

La prima serie di esperimenti illustrati erano allo studio di *statistical learning* (Chapter 2). Tramite un paradigma sperimentale basato sull'apprendimento per esposizione (*imprinting* filiale), pulcini neonati *naive* dal punto di vista visivo, sono stati esposti ad una video-sequenza di elementi visivi arbitrari (forme geometriche). Tale stimolo è definito da una struttura "statistica" basata su *transitional (conditional) probabilities* che determinano l'ordine di comparsa di ciascun elemento (ad esempio, il quadrato predice la croce con una probabilità del 100%). Al termine della fase di esposizione, i pulcini riuscivano a riconoscere tale sequenza, discriminandola rispetto a sequenze non-familiari che consistevano in una presentazione *random* degli stessi elementi (ovvero nessun elemento prediceva la comparsa di nessun altro elemento; *Experiment 1*) oppure in una ricombinazione degli stessi elementi familiari secondo nuovi *pattern* statistici (ad esempio, il quadrato predice la T con probabilità del 100% ma tale relazione statistica non era mai stata esperita dai pulcini; *Experiment 2*). In entrambi gli esperimenti i pulcini discriminarono la sequenza familiare da quella non-familiare, dimostrandosi in grado di riconoscere il struttura statistica alla quale erano stati esposti durante la fase d'imprinting. Uno degli aspetti più affascinanti di questo risultato è che il processo di apprendimento è non-supervisionato ovvero nessun rinforzo era stato dato ai pulcini durante la fase di esposizione.

Successivamente, sono stati condotti altri due esperimenti (*Experiments 3 and 4*) con l'obbiettivo di verificare se i pulcini fossero in grado di apprendere regolarità più complesse di quelle testate in precedenza. In particolare, il compito che dovevano svolgere i pulcini consisteva nel differenziare una sequenza familiare strutturata similmente a quella appena descritta e una sequenza non-familiare composta da *part-pairs* ovvero coppie di figure composte dall'unione dell'ultima figura componente una coppia familiare e la prima figura componente un'altra coppia familiare. Essendo formate dall'unione di elementi appartenenti a coppie familiari, le *part-pairs* venivano esperite dai pulcini durante la fase di familiarizzazione ma con una probabilità più bassa rispetto alle *pairs*. La difficoltà del compito risiede quindi nel rilevare una sottile differenza caratterizzante la distribuzione di probabilità dei due stimoli. Sfortunatamente i pulcini non sono stati in grado di discriminare le due sequenze ne quando composte da 8 elementi (*Experiment 3*) ne da 6 (*Experiment 4*). L'obbiettivo finale di questi due esperimenti sarebbe stato quello di scoprire il tipo di regolarità appresa dai pulcini. Infatti, negli esperimenti 1 e 2 i pulcini potrebbero aver discriminato sequenze familiari e non familiari sulla base delle frequenze di co-occorrenza delle figure componenti le coppie familiari (ad esempio, co-occorrenza di X e Y) piuttosto che sulle probabilità condizionali (ad esempio, X predice Y). Tuttavia, non avendo superato il test presentato negli esperimenti 3 e 4, la questione riguardante quale tipo di cue statistico viene appreso da questa specie rimane aperta. Possibili spiegazioni e implicazioni teoriche di tale risultato non significativo sono discusse nel capitolo conclusivo.

Il secondo gruppo di esperimenti condotti nella presente ricerca riguarda l'indagine del processo di generalizzazione di regolarità visive (*Chapter 3*). Le regolarità indagate sono rappresentate come stringhe di figure geometriche organizzate spazialmente, i cui elementi sono visibili simultaneamente. Ad esempio, la regolarità definita come AAB viene descritta come una tripletta in cui i primi due elementi sono identici tra loro (AA), seguiti da un'altro elemento diverso dai precedenti (B). I *pattern* impiegati erano AAB, ABA (*Experiment 5*) ABB e BAA (*Experiment 6*) e la procedura sperimentale utilizzata prevedeva addestramento tramite rinforzo alimentare. Una volta imparato a riconoscere il *pattern* rinforzato (ad esempio, AAB implementato da croce-croce-cerchio) da quello non rinforzato (ad esempio, ABA implementato da croce-cerchio-croce), i pulcini dovevano riconoscere tali strutture rappresentate da nuovi elementi (ad esempio, clessidra-clessidra-freccia vs. clessidra-freccia-clessidra). Gli animali si dimostrarono capaci di generalizzare tutte le regolarità a nuovi esemplari delle stesse. L’aspetto più importante di questi risultati è quanto dimostrato nell’esperimento 6, il cui obbiettivo era quello di indagare le possibili strategie di apprendimento messe in atto dagli animali nello studio precedente. Infatti, considerando il confronto AAB vs. ABA, i pulcini potrebbero aver riconosciuto (e generalizzato) il *pattern* familiare sulla base della presenza di una ripetizione consecutiva di uno stesso elemento (presente in AAB ma non in ABA, dove lo stesso elemento A è ripetuto e posizionato ai due estremi della tripletta). Nell’esperimento 6 sono state quindi confrontate regolarità caratterizzate da ripetizioni: AAB vs. ABB e AAB vs. BAA. I pulcini si mostrarono comunque in grado di distinguere le nuove regolarità e di generalizzare a nuovi esemplari, suggerendo come tale abilità non sia limitata a un particolare tipo di configurazione.

Complessivamente, i risultati ottenuti nella presente ricerca costituiscono la prima evidenza di *statistical learning* e generalizzazione di regolarità visive in un modello animale osservato appena dopo la nascita. Per quanto riguarda lo *statistical learning*, i pulcini dimostrano capacità comparabili a quelle osservate in altre specie animali e agli infanti umani ma apparentemente superiori a quelle osservate nel neonato. Ipotesi e implicazioni teoriche di tali differenze sono riportate nel capitolo conclusivo. Per quanto riguarda i processi di generalizzazione, la *performance* dei pulcini è in linea con quanto dimostrato dai neonati umani nel dominio linguistico. Alla luce di questi risultati, è plausibile pensare che il pulcino si biologicamente predisposto ad rilevare regolarità caratterizzanti il suo ambiente visivo, a partire dai primi momenti di vita.

CHAPTER 1 INTRODUCTION

Patterns of visual objects, streams of sounds, and spatiotemporal events are just a few examples of the structures present in a variety of sensory inputs. Amid such variety, numerous regularities can be found. In order to handle the sensory processing, a fundamental ability for the survival of any living organism, individuals of any species have to be able to rapidly track these regularities. *Statistical learning* is one of the core processes that enable to track regularities from the flow of sensory information. Once relevant structures are acquired, learners are sometimes required to generalize to novel situations. The *generalization* can be challenging since it demands to abstract beyond the surface information, and extract regularities underlying perceptually-novel stimuli. Recently, generalization of regularities has been defined as the *extension of learned [...] structure to unseen stimuli, typically from within the same modality or stimulus domain* (Frost, Armstrong, Siegelman & Christiansen, 2015). From now on, this mechanism will be defined as such.

The studies I am going to present in this thesis are organized into two distinct lines of investigation. The first series of experiments was aimed at studying the origins of statistical learning. From the literature indeed, it was not clear whether statistical learning is fully available at the onset of life thus I decided to investigate this issue using a novel newborn animal model: the domestic chick (*Gallus gallus*). Findings are described in Chapter 2. The second line of research explores the origins of the generalization of visual regularities, reporting novel evidence in the same species (Chapter 3).

In both series of studies behavioral methodologies have been used, allowing to perform almost direct comparisons between human newborns and newborn chicks. As we will see throughout this thesis, chicks proved to be an excellent model for the study of the ontogeny of these cognitive traits. As regards to Tinbergen's questions, the four types of problems faced by scientists when explaining behavior of living organisms (Tinbergen, 1963), the present work offers new insights about the *ontogeny* and the *phylogeny* of statistical learning and generalization in chicks. The ontogeny points to the individual development of a given trait (statistical learning or generalization) in a species (domestic chick). The phylogeny instead refers to the evolution of this trait throughout the history of the species (see Bateson & Laland, 2013 for a recent review). Furthermore, indirect evidence regarding plausible *functions* and *mechanisms* of these processes in chicks are discussed.

Why should it be interesting to explore the roots of learning and generalization of regularities? As we will see in the next paragraph, processing regularities is a fundamental mechanism to organize and interpret disparate sensory information provided by the environment. From spoken languages to animals' vocalizations, to visual scenes and sequences of events, many living creatures are consistently bombarded by streams of information. Therefore, a problem faced by these organisms is to identify all possible structures of these inputs, generalize to new realities and, eventually, form an internal representation of the environment (see Frost et al., 2015 for a recent review). But when and how do learners begin to deal with it? The research illustrated in this thesis provides novel data supporting the domain-generality of statistical learning and generalization, and showing that both mechanisms fully function at the very beginning of life.

1.1 STATISTICAL LEARNING

State of the art

Over the past two decades, a powerful and general learning process facilitating the extraction of probabilistic structures from the environment, has been investigated in different contexts and species (see Krogh, Vlach & Johnson, 2013; Frost et al., 2015, for review). Through this process, statistical coherence can be detected in a variety of domains such as objects arranged in visual scenes (e.g., Fiser & Aslin, 2001; 2002a; 2002b; 2005), sounds occurring in vocal communication and music (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin & Newport, 1999), and sequences of statistically-ordered events (e.g., Kirkham, Slemmer & Johnson, 2002; Kirkham, Slemmer, Richardson & Johnson, 2007; Bulf, Johnson & Valenza, 2011). Besides being robust enough to operate in multiple domains and modalities, statistical learning has also been reported in some animal species such as rodents (Toro & Trobalòn, 2005), nonhuman primates (Saffran, Hauser, Seibel, Kapfhamer, Tsao & Cushman, 2008; Wilson, Slater, Kikuchi, Milne, Marslen-Wilson, Smith et al., 2013; Hauser, Newport & Aslin, 2001; Grainger, Dufau, Montant, Ziegler & Fagot, 2012; Goujon & Fagot, 2013; Meyer & Olson, 2011; see Conway & Christiansen, 2001 for review) and songbirds (e.g., Takahasi, Yamada & Okanoya, 2010; Abe & Watanabe, 2011).

These results led scientists to describe this phenomenon as a domain-general mechanism that plays a fundamental role in organizing the sensory world and, eventually, elaborating internal representations of the environment. In the next section, I am going to outline some pioneering works that allowed a better understanding of this process.

In the linguistic domain, before starting to comprehend and produce words, infants face the problem of word segmentation: identifying isolated words from a continuous stream of sounds with no physical cues consistent with word boundaries (e.g., Cole & Jakimik, 1980). In a seminal study, Saffran and colleagues (1996) showed that 8-month-old infants are promptly able to detect boundaries across words by tracking the distribution of sounds combined to form those words. After only 2 minutes of exposure, infants were capable of recognizing individual words from an artificial language played as a fluent stream of sounds with no pauses in between words (orthographically, *golabupadotitupiro....*). Infants discriminated sequences containing higher transitional probabilities within words (e.g., *golabu*) from sequences containing lower transitional probabilities spanning word boundaries (part-words, e.g., *bupado,* formed by assembling *"bu"* from *gola<u>bu</u>* and *"pado"* from *<u>pado</u>ti*) and from novel combinations of familiar syllables (non-words, e.g., *tilado*). Further investigations showed that infants did not simply track frequencies of co-occurrence of adjacent syllables, but instead detected conditional probabilities (Aslin, Saffran & Newport, 1998). This distinction is fundamental. The transitional probability between two consecutive syllables represents a type of conditional probability defined as:

$$\textit{probability of } Y|X = \textit{frequency of } XY / \textit{frequency of } X$$

(Miller & Selfridge, 1950)[1]. Therefore, in Saffran et al. (1996), the transitional probability that linked all syllables composing a word (e.g., *golabu*) was 1.0, meaning that these 3 syllables predicted each other, and were always experienced in that

---

[1] There are at least two types of transitional probabilities documented in the literature: forward and backward (e.g., Pelucchi, Hay & Saffran, 2009). In this thesis I will only refer to forward transitional probabilities, described by the formula reported in the main text. On the contrary, backward transitional probabilities can be defined as $P(X|Y)=\textit{frequency }(XY)/\textit{frequency }(Y)$.

sequential order. On the contrary, the transitional probability between syllables that occurred consecutively but in between word boundaries was lower. Following the same example, in the part-word *bupado*, formed by *golabu#padoti*[2] joined together, transitional probabilities were 0.33 between *bu* and *pa*, and 1.0 between *pa* and *do* and between *do* and *ti*. The key aspect of transitional probabilities is that syllables predict each other, and predictions are based on individual frequencies of appearance of one of the syllables. For instance, in the *golabu* case, the presence of the second syllable *la* strictly depended on the presence of the preceding syllable *go*.

In contrast, the frequency of co-occurrence is a simpler type of statistical information that does not involve predictions but just the co-occurrence of elements. Computing frequencies of co-occurrence is thus easier to do with regards to transitional probabilities, because it requires to notice what co-occurs with what, and how often does this happen.

Statistical learning goes beyond the discovery of words boundaries, supporting the acquisition of grammatical categories and syntactic structures (e.g., Frost & Monaghan, 2015). In spite of being able to segment the speech stream into word units, language learners need to discover how these words are connected together; some of these words might be adjacent to one another, some others interspersed by 1 or more lexical items (see Santelmann & Jusczyk, 1998; Gómez, 2002; Gómez & Maye, 2005 for classic studies). In natural languages, non-adjacent dependencies characterize, for instance, subject-verb agreement like in the case of *Chuck quickly runs to the store.* Clearly, the presence of a subject (*Chuck*) predicts the presence of a verb ending with the

---

[2] In the statistical learning literature, the symbol pound (#) is conventionally used to signal the gap between two syllable sequences spoken with no pauses in between them.

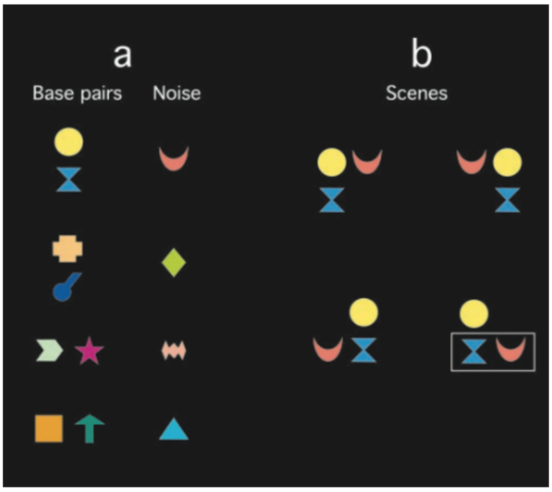
inflectional morpheme *-s,* somewhere along the sentence. This level of sentence processing appears to be available way at 18 months of age, as showed by Santelmann & Jusczyk (1998). The authors demonstrated that infants learn non-adjacent dependencies involving, for example, auxiliaries and verb endings such as *is playing,* where the statistical dependency involves non-adjacent syllables within the sentence. Learning this type of regularities requires to keep track of statistical coherence of two items interspersed among others.

The acquisition of word classes is also supported by a statistical learning mechanism (Saffran, 2001; 2002). During the first year of life, infants begin to acquire that words are organized in grammatical categories such as determiners, nouns, verbs, etc., and arranged according to hierarchical structures. In discovering word classes, infants learn that these classes can comprise more than one lexical item, and that there is no transparent evidence of that in the linguistic input (e.g., in English, determiners could be both *a* and *the*). Infants are thus required to abstract beyond the input and recognize which words pertain to a given class. In addition, infants learn that the presence of a member of a word class predicts the presence of a member of another class. For example, they acquire that a determiner always predicts a noun in English, whereas a noun can stand alone without any determiner. Learning predictive dependencies amongst word classes is fundamental for adequately processing the syntactic input (Misyak & Christiansen, 2007; Frost & Monaghan, 2015).

Furthermore, statistical learning has been broadly reported as a key mechanism involved in many other crucial aspects of language acquisition, enabling infants to extract prosodic patterns (e.g., Mattys, Jusczyk, Luce, & Morgan, 1999), discover phonotactic regularities (e.g., Mattys & Jusczyk, 2001) and link sounds to meanings (Graf-Estes, Evans, Alibali & Saffran, 2007).

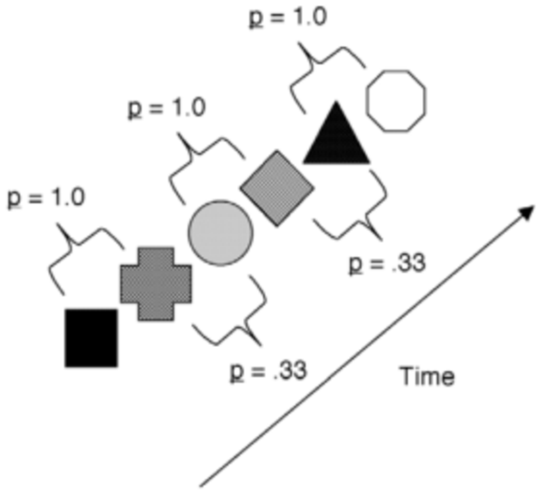
Statistical learning is not limited to language, operating also in auditory non-linguistic contexts as well as visual and tactile domains (e.g., Saffran et al., 1999; Fiser & Aslin, 2001; 2002a; 2002b; 2005; Conway & Christiansen, 2005). Eight-month-old infants and adults are capable of segmenting continuous streams of musical tones defined by a statistical structure (Saffran et al., 1999). Similar segmentation abilities have been shown in the visual domain through two different modalities of stimuli presentation. Fiser & Aslin (2002b) conducted a series of experiments where 9-month-old infants were habituated to visual scenes formed by several shapes grouped into spatially-organized pairs i.e., the circle was always located above the hourglass (Figure 1.1). After habituation, infants were presented with test pairs, some of them followed the familiar spatial pattern, some others did not. Infants proved to be sensitive both to lower-level (frequencies of co-occurrence of element pairs) and higher-order dependencies (conditional probabilities that confer higher predictability of the elements embedded in pairs). Infants did not just learned that, for instance, the circle and the hourglass co-occurred more often as a pair with respect to other shapes, but they also learned that one item predicted the other in precise spatial positions (i.e., the circle was placed always above the hourglass). The authors demonstrated that infants learned the structure of the visual scene by actually tracking conditional probabilities, by equating frequencies of co-occurrence and making conditional probability the only relevant statistical information.

Figure 1.1. Representation of the stimuli used in Fiser & Aslin (2002b). On the left side (a), shape-pairs (“Base pairs” column) and distractors (“Noise” column). On the right side (b), four possible test scenes created by a pair and a distractor, showing the strong spatial relation between the two elements (e.g., the yellow circle is always located above the blue hourglass) no matter what surrounds them.



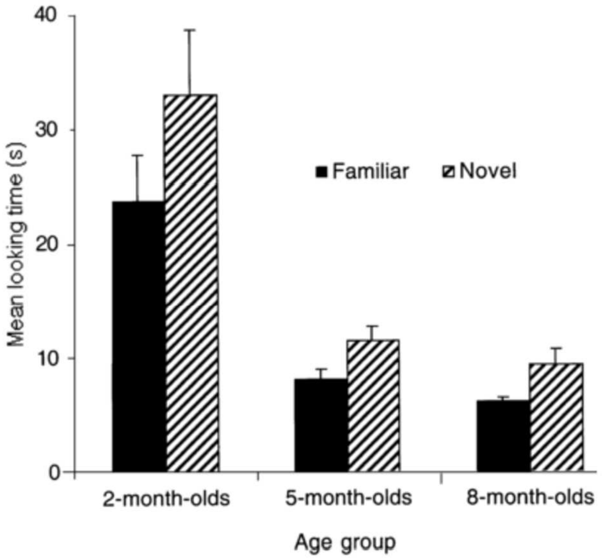
These findings suggest that statistical learning allows learners to extract salient cues from visual inputs, and eventually leads learners to develop internal representations of the (visual) environment (see also Fiser & Aslin, 2005). Interestingly, neither instruction nor feedback was provided to participants, underscoring the striking power of this mechanism as a form of unsupervised learning (for a detailed explanation see Chapter 2). An alternative way to investigate visual statistical learning was adopted by Kirkham et al. (2002) who created sequences of geometrical shapes presented one at a time, whose ordering was determined by statistical regularities. More precisely, higher transitional probabilities defined shape-pairs so that the first element of a pair always predicted the second element of the same pair (probability=1.0). As represented in Figure 1.2, the square always predicted the cross, the circle always predicted the diamond and the triangle always predicted the hexagon. Conversely, lower transitional probabilities defined relationships *between* pairs (probability=0.33).

Figure 1.2. Graphic representation of the familiar sequence of Kirkham et al. (2002). Numbers correspond to transitional probabilities within (p=1.0) and between (p=0.33) pairs.



Since the familiar sequence was presented as a continuous stream, with no pauses between pairings, the only available cue to segment the sequence was its probabilistic structure. The unfamiliar sequence was instead a semi-random presentation of the same familiar shapes, defined by any statistical dependency between them, with the only constraint that the same shape never appeared consecutively. After habituation, 2-, 5- and 8-month-old infants were equally able to extract the sequence structure, distinguishing the familiar, statistically-defined, sequence from the semi-random stimulus. This evidence is represented by the fact that all infants looked significantly longer to the unfamiliar stimulus (Figure 1.3, “novel” columns), meaning that they recognized the habituation sequence. Please note that in the habituation-dishabituation paradigm, the expected direction of the preference is toward the novel, unfamiliar stimulus.

Figure 1.3. On the Y axis, looking times during the test phase are represented; on the X axis, the three groups of infants divided on the basis of their age.



An extension of this work is represented by data collected in a subsequent research on 1- to 3-day old neonates. Bulf and collaborators (2011) used an habituation-dishabituation procedure during which they played a fluent stream of 4 (LDC=low-demand condition) or 6 (HDC=high-demand condition) shapes, similar to the stimuli employed by Kirkham et al. (2002), (Figure 1.4). Once habituation was reached, neonates were presented with the familiar sequence and a semi-random generation of the familiar items. As shown in Figure 1.5, neonates discriminated test sequences only in the low-demand condition, failing when 6-element sequences were used.

Figure 1.4. Graphic representation of habituation sequences used in the HDC (on the left) and LDC (on the right). Numbers correspond to transitional probabilities within (p=1.0 in both conditions) and between (p=0.33, in the LDC; p=0.5, in the HDC) pairs.

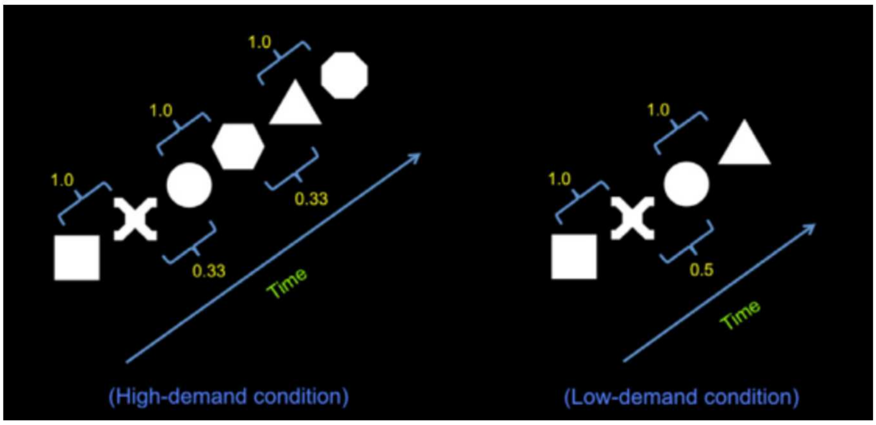
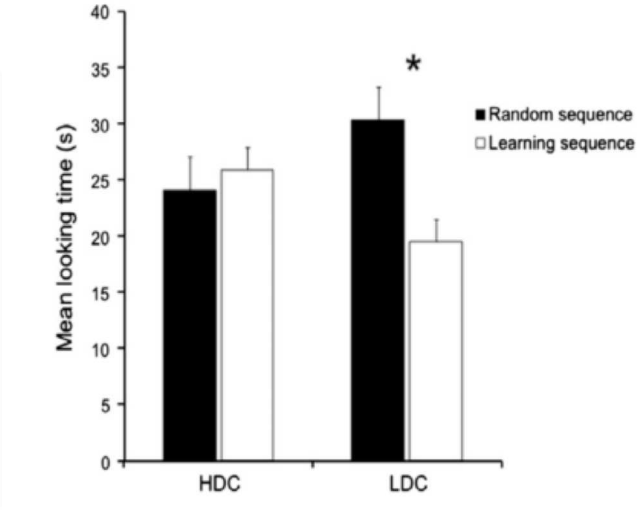


Figure 1.5. On the Y axis, mean looking times during the test phase are represented; on the X axis, the two groups of infants divided on the basis of the condition. Asterisk indicates significant difference in looking times between the familiar (learning) sequence (mean=19.49±8.56 seconds) and the random one (30.26±16.31 seconds) in the LDC (stimuli formed by 4 shapes).



There are many reasons that may explain the findings obtained by Bulf et al. (2011). Cognitive limitations of the human neonate may have reduced the amount of information acquired as well as their limited selective attention capacities. Considering these natural restrictions, a stimulus composed by 4 sequentially-presented items becomes easier to master with respect to 6-element sequences. Furthermore, as pointed out by the authors, in the low-demand condition neonates may have had increased opportunities to detect the habituation pattern as opposed to the high-demand condition. This is due to the fact that 4 items have higher frequency of appearance than 6 shapes within the same time window. Another plausible explanation of this evidence regards the primary essence of the sensory world. In the auditory environment, sounds coming from a same source usually appear one at a time, and do not persist in time. Visual scenes are instead likely formed by simultaneously-visible objects, with rare patterns of sequentiality (e.g., sign language, sequences of actions). Acoustic regularities are indeed better learned when presented as a temporal sequence, and visual patterns are better acquired when elements are visible all at once (Saffran, 2001; 2002; Gebhart, Newport & Aslin, 2009). It might be the case that, when learning requires further cognitive effort like in the high-demand condition, and stimuli are not displayed coherently to the natural laws, perceptual aspects such as modality of presentation function as constraints to the learning process. Overall, results of this study revealed that visual statistical learning *can* operate at birth but it is highly constrained by developmental factors and perceptual aspects of the input.

Statistical learning is not limited to humans. Different levels of complexity in auditory and visual statistical learning have been demonstrated in some animal species including rodents (Toro & Trobalòn, 2005), nonhuman primates (Saffran et al., 2008;

Wilson et al., 2012; Hauser et al., 2001; Grainger et al., 2012; Goujon & Fagot, 2013; Meyer & Olson, 2011) and songbirds (e.g., Takahasi et al., 2010; Abe & Watanabe, 2011). For example, rats are capable of segmenting the same speech stream used by Saffran et al. (1996) however they do so by tracking frequencies of co-occurrences rather than transitional probabilities (Toro & Trobalòn, 2005). Through a visual search task, baboons learned to identify a target stimulus presented along with some distractors, taking advantage of contextual regularities of the visual scene (i.e., spatial relations among items; Goujon & Fagot, 2013). To date, some songbirds species demonstrate the highest level of statistical learning being able to detect complex statistical regularities resembling those of the human syntax, implemented by their own song syllables (e.g., Abe & Watanabe, 2011). Even more interesting, when learning songs from adult tutors, juveniles of Bengalese finches (*Lonchura striata*) select several segments of notes sung by different individuals. The interesting aspect of this process is that the birds do not select randomly, but extract the pieces of songs with the highest internal statistical coherence (Takahasi et al., 2010).

Research questions

The first aim of the present research arose from previous evidence of visual statistical learning in the human newborn. How does statistical learning function at the onset of life? As mentioned above, neonates failed to recognize the structured pattern when was compared to a random stimulus involving the same shapes in the case that both stimuli comprised more than 4 temporally-presented elements. Limited cognitive and attentional resources may have been responsible for neonates' difficulty to keep track of many elements over time. Therefore, in the present research I investigated whether statistical learning is effective at the very beginning of learners' life. In order to explore this issue, the newborn domestic chick (*Gallus gallus*) has been used and validated as behavioral model for the study of visual statistical learning, for the first time. As illustrated in Paragraph 1.3, this species provides an excellent developmental model for the study of the ontogeny of several cognitive traits. I developed a paradigm based on previous studies of visual statistical learning in infancy (Bulf et al., 2011; Kirkham et al., 2002), allowing to explore similarities and differences between the two species.

The second research question concerned the extent to which chicks can track statistical patterns from streams of information. One way to investigate this issue is to test whether chicks can perform complex computations like, for instance, detecting small differences in transitional probabilities between elements, as in the words vs. part-words comparison described earlier. To date, this contrast has been investigated in speech segmentation tasks in two different species, infants and rodents (e.g., Saffran et al., 1996; Aslin et al., 1998; Toro & Trobalòn, 2005). Learners were presented with a stream of syllables formed by words, non-words and part-words and the key difference among them was that syllables composing words had higher probability to appear in that order with respect to non-words or part-words. Crucially, since part-words were formed by assembling together the last syllable of a word and the first syllable of another word, they did appear in the familiarization stream but with lower probability. Recognizing the structural difference between words and part-words was thus more challenging than the computation needed to differentiate words and non-words, whose syllable order had never been experienced before. I created a visual version of the words vs. part-words comparison, named pairs vs. part-pairs. Part-pairs were constructed by joining the last and the first shapes of two separate familiar pairs. The goal was to investigate whether chicks could differentiate strings of shapes linked by high transitional probabilities (pairs) from strings of shapes linked by low transitional probabilities (part-pairs).

Once established that, my final goal would have been testing whether chicks are effectively able to pick up transitional probabilities or whether they simply track frequencies of co-occurrence. As described earlier, computing frequencies of co-occurrence requires detecting what *co-occurs* with what, and how often does this happen. On the contrary, computing transitional probabilities requires detecting what *predicts* what, based on different levels of probability. Since both statistical cues can coexist within the same input, the only way to investigate this issue is by making transitional probabilities the only available information to learn the structure of the input. Considering how part-words are typically created in this kind of tasks, and given the fact that word frequencies are usually balanced across stimuli (so that each word appears a same number of times), in Saffran et al. (1996) words appeared twice as often as part-words. Therefore, infants could have easily distinguished words and part-words just by tracking their frequencies. This can be avoided by manipulating word frequencies so that two of the words appear twice as often (high-frequency words) than the remaining two words (low-frequency words). As a consequence, part-words composed by syllables forming the two high-frequency words have identical absolute frequency of the two low-frequency words, but they differ in transitional probabilities (Aslin et al., 1998; Toro & Trobalòn, 2005). A similar manipulation could then be performed with visual sequences, creating frequency-matched stimuli in order to make transitional probabilities the only relevant cue to differentiate test stimuli. This would have allowed to a neat understanding of which statistical cues chicks can detect. Unfortunately, chicks did not succeed in the first task, the pairs vs. part-pairs comparison, being unable to detect subtle differences in transitional probabilities. Data are therefore inconclusive as regards whether chicks are sensitive to transitional probabilities or whether they simply track frequency of co-occurrences.

Finally, my third research aim came from a deep examination of the animals' literature on statistical learning. The existing nonhuman evidence seems to be incomplete due to the lack of inclusion of animals distantly related to humans. Nonhuman primates may share precursors of communicative as well as statistical learning capacities with our species, and the only non-mammalian model studied to date has been the songbirds who are vocal learners. Indeed, humans and songbirds are both classified as high- and complex- vocal learners (see Petkov & Jarvis, 2012, for a review) based on their ability to produce structured sequences of sounds, and learn novel vocalizations. Therefore, it is perhaps not surprising that songbirds are sensitive to statistical properties defining songs produced by conspecifics (e.g., Beckers, Bolhuis, Okanoya, & Berwick, 2012). The comparative investigation of statistical learning could greatly benefit from evidence obtained in species that are at the same time 1) phylogenetically more distant from humans than nonhuman primates, and 2) not endowed with vocal learning capacities resembling those of humans. The domestic chick fits this profile. Cross-species evidence unrelated to language and communicative skills could thus come from avian species without complex vocalization systems or vocal learning abilities. Do such species possess robust statistical learning capacities, like those shown by the humans and nonhumans tested so far? Looking at the avian phylogenetic tree combined with the complex-vocal learning phenotype, *galliformes* such as chickens and quails do not belong to the vocal-learners category (Petkov & Jarvis, 2012). Thus, any statistical learning abilities shown by chicks should be unrelated to vocal learning or singing abilities.

1.2 GENERALIZATION OF REGULARITIES

State of the art

A fascinating approach adopted to deeply investigate regularity learning is based on exploring one of its key components, the *generalization.* As mentioned in the introduction of this chapter, generalization requires to detect the structure of the input without relying on its surface information.

In a seminal study, Marcus, Vijayan, Bandi Rao & Vishton (1999) familiarized 7-month-old infants with a set of syllable strings arranged according to two patterns: ABB and ABA, where A and B could be implemented by any given syllable. Strings created according to these regularities were, for example, *ga-ti-ti* and *ga-ti-ga*. Infants were then able to recognize the familiar pattern by discriminating novel triplets that followed the pattern from those that did not i.e., *wo-fe-fe* and *wo-fe-wo*. Importantly, new exemplars did not share perceptual features with the familiar input thus infants could not have simply relied on the syllable strings to succeed in the task. They were required to detect the relation underlying string elements, independently of the syllable identity. For instance, the pattern defining ABB can be described as “one token followed by two identical tokens, but different from the former”.

Despite being one of the core mechanisms involved in language acquisition, generalization of regularities can be conceived as a domain-general process that operates across domains and modalities (see ten Cate & Okanoya, 2012 for review). However, when it comes to nonlinguistic stimuli, the generalization process appears to be constrained by perceptual aspects of the input. When the elements implementing a pattern are particularly salient for the learner, the pattern seems to be better acquired. For example, in infancy, acoustic regularities presented as sequences of spoken syllables (Marcus et al., 1999) are more efficient in driving the learning process than musical tones. Marcus, Fernandes & Johnson (2007) revealed that infants acquired and generalized AAB, ABB and ABA instantiated by nonlinguistic sounds (piano notes, musical tones differing only in timbres and animal calls) only if they were previously exposed to the same patterns presented as strings of syllables (Marcus et al., 2007; see Thiessen, 2012 and Dawson & Gerken, 2009 for further discussion). This evidence has been interpreted in the light of previous findings showing that infants prefer to listen to speech stimuli over non-speech sounds and non-speech analogues that retain some temporal and spectral characteristics of the speech signal (Vouloumanos & Werker, 2004). It has been thus suggested that infants' listening bias toward the speech may have influenced their capabilities of tracking regularities from the non-speech sounds used in Marcus et al., (2007).

A slightly-different and broader explanation has been further proposed by Thiessen (2012). As opposed to a special preference for the speech, it might be the case that the speech represents just a very interesting and familiar sound for the infants. Other findings can be explained under the same light, supporting a more general perspective. In the visual domain, Saffran Pollak, Seibel & Shkolnik (2007) created visual triplets of pictures of dog and cat breeds according to ABA, AAB and ABB (Figure 1.6). Triplets were presented as spatial-temporal sequences where each token appears one at a time resulting in a full-visible triplet of images.

Figure 1.6. Examples of regularities used in Exp. 1 and 2 in Saffran et al. (2007).

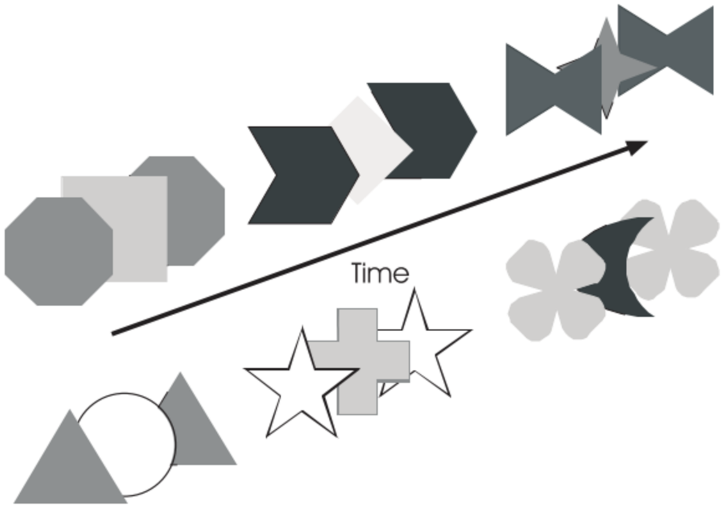
Test patterns (Exp. 1) implementing ABA

Test patterns (Exp. 2) implementing ABA

Seven-month-old infants were not only capable of tracking the habituation pattern, they also recognized the same pattern when instantiated by novel images representing different breeds of dogs and cats. For instance, if the habituation pattern was ABA displayed as *Malamute -Cattle Dog - Malamute*, infants could differentiate pattern-following (e.g., *Shiba Inu - Caanan Dog - Shiba Inu*) and pattern-violating (*Shiba Inu - Caanan Dog - Caanan Dog*) test triplets formed by perceptually-novel images.

In a subsequent study, Johnson, Fernandes, Frank, Kirkham, Marcus, Rabagliati & Slemmer (2009) familiarized infants with triplets of geometrical shapes constructed according to ABB, AAB and ABA (Figure 1.7). Triplets were presented as temporal sequences whose elements appeared one by one on a screen, and never simultaneously-visible. Results revealed a lower level of discrimination and generalization compared to what obtained by Saffran et al. (2007). In fact, 8-month-old infants did not distinguish ABB from AAB whereas 11-month-olds successfully differentiated these patterns. On the contrary, both groups of infants failed in the ABA vs. ABB comparison. This latter finding is inconsistent with Marcus et al. (1999) in which 7-month-olds infants discriminated and generalized ABA vs. ABB represented as strings of syllables as well as other type of patterns.

Figure 1.7. Schematic representation of triplets sequences used in Johnson et al. (2009), instantiating the ABA pattern.



Taken together these findings showed that triplets of dog/cat images are more efficient than geometrical shapes in driving infants' attention to the regularity, promoting its learning. Thiessen grouped all these evidence together (Marcus et al., 1999; Marcus et al., 2007; Saffran et al., 2007; Johnson et al., 2009), and put forward the hypothesis that the perceptual saliency of the stimuli implementing a pattern may contribute to enhance learning and generalization. In the above-mentioned works, infants may have been particularly attracted by spoken syllables and pictures of dogs/cats due to the high-familiarity represented by these stimuli.

Another perceptual factor that may constrain the generalization of regularities is the modality of presentation. Acoustic regularities are better learned when presented sequentially whereas visual patterns are better acquired when elements are simultaneously-visible (Saffran, 2001; 2002; Gebhart et al., 2009). Looking at the results reported earlier, modality-specificity may have also influenced infants' performance, at least in the visual studies. Dog/cat images forming the triplets used in Saffran et al. (2007) were visible all at once, whereas the shapes employed in Johnson et al. (2009) were not. It is thus possible that the combination of the two perceptual aspects (saliency of the stimuli implementing the patterns and modality-specificity) has contributed to a better performance in one case (Saffran et al., 2007) rather than the other (Johnson et al., 2009).

The third perceptual aspect of the input that may affect learning and generalization is the presence/absence of adjacent reduplicated elements, that confers an asymmetrical structure to the pattern. For instance, in AAB two identical As are adjacent to one another (asymmetrical structure), as opposed to ABA where the same item is repeated in non-adjacent positions within the triplet (symmetrical structure). Research focused on the role of adjacent reduplicated items in boosting the generalization of acoustic patterns, revealed that repetitions matter. Gervain, Macagno, Cogoi, Peña & Mehler (2008) showed learning differences when ABB syllable sequences (e.g., *mu-ba-ba*) were presented to neonates with respect to ABA sequences (e.g., *ba-mu-ba*), both contrasted with random (ABC) stimuli (e.g., *mu-ba-ge*). This study showed greater activation of temporal and left frontal regions only when ABB stimuli were presented, revealing that neonates could discriminate those patterns from random stimuli. A perceptual mechanism that privileges adjacent repetitions in acoustic stimuli had been previously proposed in the literature. Endress, Dehaene-Lambertz & Mehler (2007) showed that human adults could generalize familiar tone sequences to novel ones, only when adjacent repetitions characterized the structure of the sequence. Moreover, participants proved to be sensitive to different positions of the repetition within the sequence, showing successful generalization only when repetitions were located at the very final edge of the string (e.g., ABCDE*FF* vs. ABCD*EE*F; Endress, Scholl & Mehler, 2005). Even though the role of repetitions has been mostly investigated in the auditory domain, learning and generalization of visual structures appear to be similarly constrained. Visual regularities comprising adjacent repetitions (AAB and ABB) seem to be better acquired than those with non-adjacent identical items (ABA), (e.g., Johnson et al., 2009).

In sum, although human beings can easily track regularities from different domains, the nature of the input affects learning outputs. Perceptual constraints seem to be i) learners' familiarity with the input, ii) modality-specificity and, iii) input symmetrical vs. asymmetrical structure (see also, Saffran & Thiessen, 2007).

In order to further investigate the phylogenetic origins of regularity learning, a handful of studies have explored this process in several nonhuman species. Among mammals, rats' ability to detect regularities dates back to the 1970s when Rescorla & Wagner (1972) discovered the capacity to associate events following one another. More pertinent with the current topic is rats' ability to learn sequential patterns of tones following structures such as XYX vs. XXY and XYX vs. YXX, showing generalization to novel exemplars (Murphy, Mondragon & Murphy, 2008). Moreover, rats can learn similar regularities implemented by sequences of consonants and vowels, and generalize to new sequences of words (de La Mora & Toro, 2013). Similar evidence has been found in rhesus monkeys (*Macaca mulatta*; Hauser & Glynn, 2009) which can discriminate short strings of species-specific vocalizations displayed as ABB or AAB, and generalize to perceptually-new strings. As regards to avian species, songbirds (European starlings and Zebra finches) have been extensively studied in this field because of their rich repertoire of vocalizations (see ten Cate & Okanoya, 2012 for a review), revealing strong learning of complex patterns defining their own songs. In spite of that, songbirds' generalization capacities seem to be controversial. In a seminal study conducted by Gentner, Fenn, Margoliash & Nusbaum (2006), starlings (*Sturnus vulgaris*) generalized regularities instantiated by song motifs like rattles and warbles (e.g., rattle-warble-rattle-warble representing ABAB) to stimuli composed of novel rattles and warbles. As generalization sequences were formed by elements belonging to the same category, it was unclear whether starlings generalized based on phonetic similarity or whether they really detected the structures. A subsequent study by Comins & Gentner (2013) demonstrated that starlings' generalization might be perceptually-constrained. The birds could learn sequences of song motifs organized as XXYY or XYXY, but they could not generalize to novel exemplars if the novel sounds do not follow natural acoustic categories. Zebra finches (*Taeniopygia guttata*) also showed a limited degree of generalization. In a recent study, Chen, van Rossum & ten Cate (2015) trained the finches to discriminate sequences of song elements arranged according to ABA or AAB. All birds revealed discrimination of the reinforced pattern whereas none of them could identify it when implemented by novel song items.

As regards to the visual domain, the most important evidence of regularity learning comes from two diverse species: pigeons and honeybees. The investigations conducted on these models were guided by different research questions than those addressed by my research, nevertheless these works provided fundamental evidence supporting the domain-generality of these mechanisms. Pigeons (*Columba livia*) are well-known for their impressive capacity of visual object recognition and category learning (see Zentall, Wasserman, Lazareva, Thompson & Ratterman, 2008; Soto & Wasserman, 2014, for review). A huge body of data indeed demonstrated that pigeons extract invariant properties across groups of objects depicted as photographs, being able to identify objects belonging to the same category based on perceptual features (e.g., flowers and cars are categorized as members of different categories; Lazareva, Freiburger & Wasserman, 2004). The aspect of interest for the present thesis is that these birds, after learning to discriminate members of a given category, can recognize new members of the same category showing strong generalization.

An invertebrate organism such as the honeybee (*Apis mellifera*) is capable of forming the concepts of *sameness* and *difference* with visual stimuli. Through a matching-to-sample task, bees learned to find a test stimulus that matched a training stimulus for some visual properties (vertical and horizontal configurations). The most impressive aspect of this study regards the generalization phase, during which bees succeeded at transferring the learning to novel visual stimuli (colors), and to a different sensory modality (olfaction; Giurfa, Zhang, Jenett, Menzel & Srinivasan, 2001).

Research questions

Cross-species evidence showed that the generalization of regularities is a quite strong mechanism that operates in different domains and organisms. However, very little is known about the earliest stages of this process in the animal kingdom. The vast majority of the studies illustrated in the literature have focused on nonhuman adult learners, leaving open the issue of how generalization of patterns such as strings of sounds or visual objects functions at the onset of life.

Do newborn organisms with very limited experience of the sensory world show similar generalization of patterns to what found in adult learners? This was the first research question addressed by this second series of studies, aimed at shedding light on the ontogeny of this trait. In a comparative perspective this knowledge would allow researchers to compare young human and nonhuman learners, providing new information on cross-species differences in the development of generalization abilities. I decided to investigate this issue again employing the domestic chick as animal model. Newly-hatched chicks were reared in a controlled environment in order to control their visual experience prior to the experimental sessions, which were ran through an operant conditioning procedure.

The second research question focused on whether perceptual mechanisms such as the symmetrical vs. asymmetrical structure of the pattern constrain the generalization in nonhuman species. A repetition-detector mechanism appears to positively affect learning and generalization of regularities in humans from the earliest stages of development. I was thus interested in whether the presence/absence of adjacent repetitions played such a predominant role in the processing of visual inputs in this model.

1.3 THE DOMESTIC CHICK AS ANIMAL EXPERIMENTAL MODEL: BEHAVIORAL, COGNITIVE AND NEURAL ASPECTS OF *Gallus gallus domesticus*

Chicks offer several advantages as an animal model for behavioral experiments. First of all, the *Gallus gallus domesticus* is a precocial species meaning that the newborns i) hatch in an advanced state of development, ii) are independent of their mother's care, and iii) display full behavioral repertoire (Lorenz, 1937) and complete maturation of the visual pathways at birth (Deng & Rogers, 1998). This species, like other avian species, possesses two main visual pathways to the forebrain, the thalamofugal and the tectofugal pathway, considered the equivalents to the geniculostriate and extrageniculate visual pathways of mammals (Shimizu & Karten, 1993). In chicks, both pathways present complete decussation in the optic chiasm, even though only the thalamofugal one projects to the visual Wulst (e.g., Gunturkun, Miceli & Watanabe, 1993). Chicks are also characterized by a good visual acuity from the very first days of life. All these aspects make vision in chicks the predominant sensory modality (Schmid & Wildsoet, 1998).

Being a precocial species, chicks can be tested soon after hatching, controlling for the role of experience prior to exposure or training, and test procedures. In addition, their behavioral responses are discrete and easy to detect, making the behavioral measures quite objective. As a consequence, newborn and juvenile chicks have been extensively used in visual-perception studies (Rosa Salva, Daisley, Regolin & Vallortigara, 2010; Rosa Salva, Regolin & Vallortigara, 2010; Rosa Salva, Regolin, Mascalzoni & Vallortigara, 2012; Rosa Salva, Rugani, Cavazzana, Regolin & Vallortigara; 2013) and cognitive investigations, proving that this species is endowed with an impressive set of abilities. For example, chicks are sensitive to point-light animated sequences displaying biological motion (Regolin, Tommasi & Vallortigara, 2000), can perform transitive inferences (Daisley et al., 2010), and are capable of discriminating visual objects composing a series based on their ordinal placement, implying a sensitivity to positional properties of multi-element series (Rugani, Regolin & Vallortigara, 2007; Rugani, Kelly, Szelest, Regolin & Vallortigara, 2010; Rugani, Vallortigara, Vallini & Regolin, 2011). Another remarkable feature of this model is its learning skills, such as those involved in the filial imprinting phenomenon. Imprinting allows the young of precocial animals to actively learn the structure of their environment, in order to identify relevant objects such as mother and siblings. This happens with no reinforcement or feedback but just as the result of chicks' exposure to the environment during the first hours after hatching.

Chicks' neuroanatomy and neurophysiology are also very well known with respect to other species. The chick's brain is far from being completely homologous to the human one, especially the telencephalon and its pallial parts that were independently shaped by evolution in birds and mammals. However, homologies of brain structures between avian and mammalian species have been increasingly recognized in recent years (e.g., homologies between avian and mammalian hippocampus, Mayer, Watanabe & Bischof, 2013). For instance, although the chick's pallium is not characterized by the typical layered structure of the human neocortex but defined by a nuclear structure (e.g., Karten

& Shimizu, 1989; Butler & Hodos, 2005), there are areas which are partially homologous to the mammalian cortex (i.e., areas of pallial origins), and are functionally correspondent to humans' cortical areas (Jarvis, Gunturkun, Bruce, Csillag, Karten, Kuenzel et al., 2005).

CHAPTER 2 INVESTIGATING STATISTICAL LEARNING

Brief introduction

This chapter focuses on a learning mechanism that allows organisms to detect regularities from streams of information, namely statistical learning. As reported in Chapter 1, there is a lack of evidence in the literature about whether learning of visual patterns is fully available at the very beginning of the learner's experience with the world. Therefore, I decided to investigate this issue using the newborn chick as animal model (see Chapter 1 for information about this species).

Aims & Hypotheses

I conducted 4 experiments with a visual exposure (imprinting) procedure, aimed at i) verifying chicks' unsupervised learning of structured visual sequences (e.g., X always predicts Y) comprising different numbers of elements (Experiments 1 and 2), and ii) investigating the complexity of the patterns that can be learned by this species (Experiments 3 and 4).

I hypothesized that chicks should be capable of discriminating visual stimuli based on differences in their statistical structures. To follow, I am going to justify this hypothesis in details. From the existing evidence reported in Chapter 1, it seems very plausible that statistical learning evolved to serve other perceptual and cognitive processes besides language and communication thus, it would not be surprising to find evidence in another nonhuman species. But the reason why chicks should be capable of tracking statistical patterns is not limited to the fact that statistical learning is a domain-general process. As mentioned above, statistical learning is a mechanism that enables organisms to extract the underlying structure of the sensory input (e.g., Frost et al., 2015), and develop an internal representation of the environment (e.g., Fiser & Aslin, 2005). Chicks belong to a precocial species whose neonates are largely independent from their mothers' care, leading newborns to actively learn from their environment right after hatching. Such an early learning process benefits from social contexts. For example, chicks can identify edible items by themselves or by observing others feeding. Considering both these assumptions, it seems plausible that this species is able and motivated to track statistical coherence among stimuli in the environment, especially those defining salient objects such as mother and siblings. The adaptive value of statistical learning may especially be related to learning about the social environment (e.g., recognizing the familiar group, identifying individual siblings from single cues, etc.), an essential capacity for the survival of this species since very early in life. In addition, statistical learning may enable chicks to recognize not only relevant objects of the environment, but also complex series of events such as diadic interactions between individuals, primary aspects of the establishment of social hierarchies (e.g., Daisley et al., 2010). It is therefore possible that statistical learning works in tandem with filial imprinting, allowing newly-hatched animals to extract fundamental patterns identifying the social world.

My general prediction was that, if chicks are capable of tracking regularities from the visual stimuli used in this research, they would show a preference for one of them as a sign of discrimination (Experiments 1 and 2). For what concerns Experiments 3 and 4, I did not have clear predictions regarding the type of patterns that these animals can learn.

2.1 EXPERIMENT 1

2.1.1 Subjects

The participants were 74 (36 males) domestic chicks (*Gallus gallus*), hatched in the Laboratory of Comparative Psychology (Department of General Psychology, University of Padova). Fertilized eggs were obtained from a local commercial hatchery (Agricola Berica, Montegalda, Vicenza, Italy). On arrival, eggs were placed in an MG 70/100 incubator (45×58×43 cm, 100-egg capacity) until day 19 of incubation. Temperature was maintained at 37.5 °C and humidity was maintained at 55–60%, providing standard conditions for optimal incubation. From day 19, the eggs were placed in a hatchery (60×65×66 cm) with the same temperature as the incubator, but at a lower humidity, ideal conditions for hatching. The incubator, the hatchery, and the hatching room were maintained in complete darkness. Behavioral observations took place within the first day of life. At the end of the observations chicks were housed in pairs in standard metal cages (28x32x40 cm). The rearing room was kept at stable temperature (28–31 °C) and humidity (68%) and constantly illuminated by fluorescent lamps (36 W) located 45 cm above each cage. Food and water were available *ad libitum*, in transparent glass jars (5 cm in diameter, 5 cm in high), placed in the center of the each cage. Within a few hours from the end of the experimental session chicks were donated to local farmers.

2.1.2 Stimuli

The stimuli used in Experiment 1 are based on those used in Bulf et al. (2011) and Kirkham et al. (2002). The stimulus used in the familiarization phase was a continuous stream of 4 shapes presented one at a time in the center of a computer screen. Each shape was presented for 2 seconds and loomed from 2 to 10 cm in height, at a viewing distance of 50 cm (visual angle: 2.2°-11.2°). The familiarization sequence was composed of orange shapes (RGB = 205r, 48g, 17b; Color-Hex Code #cd3011), in order to attract chicks' attention toward the stimulus during the relatively brief exposure time of only two hours (Johnson, Bolhuis & Horn, 1985; Regolin & Vallortigara, 1995; Rosa Salva et al., 2010). The familiar sequence consisted of two shape-pairs defined by statistical dependencies within and between elements. Pair 1 consisted of a *square* always followed by an *X-shape* (Transitional Probability within-pair=1.0) and Pair 2 consisted of a *circle* always followed by a *triangle* (TPs[3] within-pair=1.0). The item that appeared after each pair was the first element of one of the two pairs (TPs between-pairs=0.5). As in prior studies of visual statistical learning, consecutive repetitions of the same pair were allowed. Since there were no pauses between pairs, the only cue available to segment the stream was the statistical structure of the sequence.

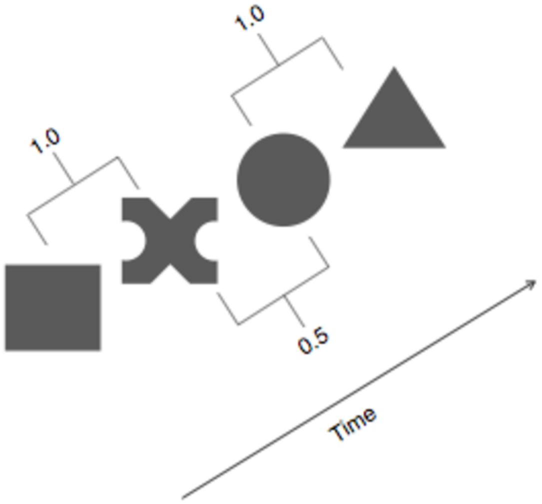
The test sequences consisted of the familiar stream and unfamiliar stream which was a semi-random generation of the 4 familiar shapes. The semi-random ordering was not constrained by any statistical relations within/between shapes however, two identical shapes never appeared consecutively (Figure 2.1). Importantly, the only difference between familiar and unfamiliar sequences was the statistical organization of the elements. The color of the shapes in each test sequence was a neutral grey color (RGB=74r, 74g, 74b; Color-Hex code #4a4a4a). Ideally, grey shapes should have been used in both familiarization and test sessions because such a neutral color minimizes the salience of the single shapes composing the sequence. In order to obtain maximal levels

---

[3] From now on, transitional probabilities will be reported as TPs.

of familiarization during a relatively brief exposure time, the orange color characterized only the familiarization sequence. E-Prime 1.1 was used to design the experiment and play the stimuli. We hypothesized that if chicks are able to distinguish the stimuli, and express a choice for one of them, this preference would be the result of the extraction of the familiar pattern.

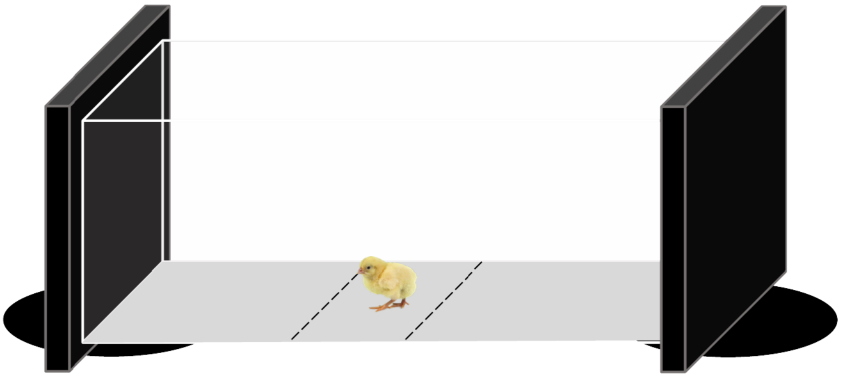
Figure 2.1. Illustration of the structure of the familiar-test sequence (compared to a semi-random generation of the same shapes). Numbers and parentheses indicate TPs within (1.0) and between (0.5) elements.



2.1.3 Apparatus

The apparatus was a white plywood runway (103x20x30 cm), divided into a central sector (25 cm) and two lateral sectors (each one of 44 cm). Two identical computer screens were placed at the opposite ends of the runway, simultaneously playing the two test sequences (Figure 2.2). Apart from the light arising from the monitor, the room was kept in complete darkness.

Figure 2.2. Illustration of the test apparatus.



2.1.4 Procedure

On the first day of life, 2 hours after hatching, chicks were taken from the dark hatchery in a closed cardboard box and placed individually in a familiarization box, where they were exposed to the familiar stimulus for 120 consecutive minutes (*familiarization phase*). The use of such a relatively short exposure time is not infrequent in the imprinting literature, especially when animated stimuli are used[4]. At the end of the exposure time, each chick was taken from its box and placed in another dark hatchery (completely empty) for 30 minutes. Previous evidence suggested that recognition memory following exposure to a stimulus is enhanced when chicks are maintained in a dark environment for some time after exposure (Johnson & Horn, 1988). At the end of this phase, each chick was individually placed in the central area of the test apparatus for the *test phase*. At the beginning of the test, the chick could face either one of the two long

---

[4] e.g., Lea, Slater & Ryan, 1996; Bolhuis, Johnson & Horn, 1989; Clara, Regolin, Zanforlin & Vallortigara, 2006; Mascalzoni, Regolin & Vallortigara, 2010. These studies used exposure time ranging from 90 to 240 minutes.

sides of the runway and the monitor playing the familiar sequence could be either the one on the right or on the left side of the apparatus. These two factors were counterbalanced between subjects. Chicks' movements along the runway were observed for 6 consecutive minutes and recorded by a camera placed above the apparatus. Entrance and permanence of the chick in one of the sectors were considered as a preference for the stimulus presented at that end. Chicks' behavior was scored on-line by an experimenter blind to the purposes of the research. When coding online, the experimenter watched the animal through a TV-screen connected to the video-camera (that is, she did not watch the animal by directly looking inside the apparatus). Fifty percent of the sample was re-coded offline by the experimenter blind to stimuli position, and the two codings were highly consistent (IntraClass correlation coefficient, single measures=0.87, $p>0.001$).

2.1.5 Results

As dependent variables I measured (1) the *first stimulus approached* (sequence played at the end of the lateral sector chosen by the chick for the first time during test phase), analyzed by a Chi-square test and (2) the *proportion of time spent near the familiar stimulus*, calculated by the formula:

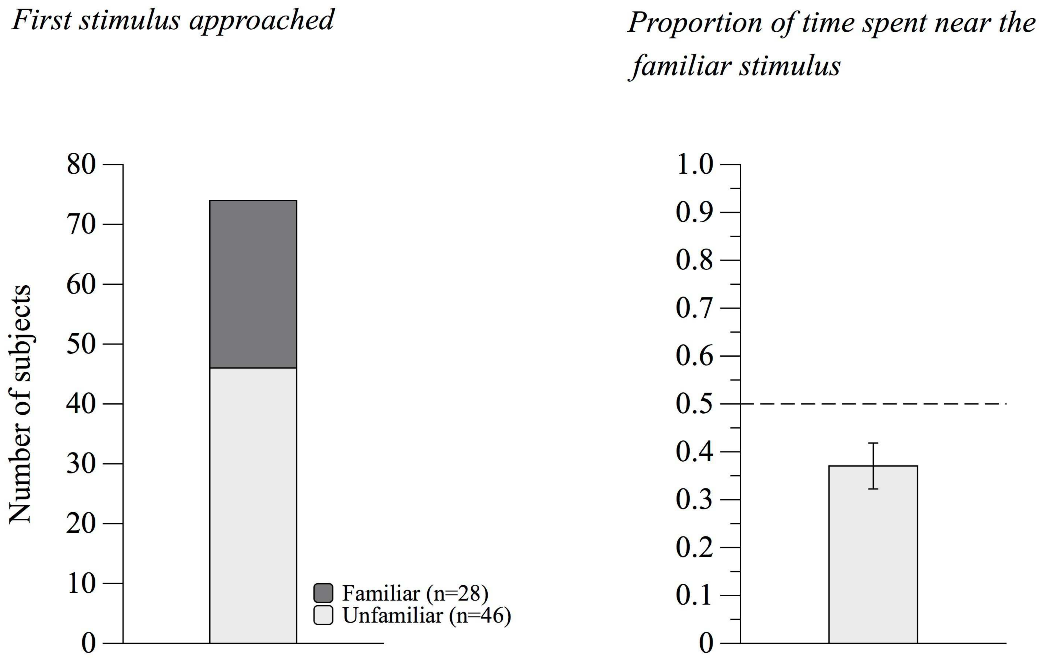
[Time spent by the familiar sequence /

(Time spent by the familiar sequence + Time spent by the unfamiliar sequence)]

This latter variable represents the proportion of time spent by the chicks in the lateral sector adjacent to the screen playing the familiar sequence. It consists of a proportion between 0 and 1, calculated from the time spent by the animals in each of the two lateral sectors; 0 represents the overall time spent near the screen playing the unfamiliar sequence whereas 1 represents the overall time spent near the screen playing the familiar sequence. The chance level was set at 0.5, representing the same amount of time spent near each screen. This measure was analyzed using a one-sample two-tailed t-test to determine whether the proportion of time spent near the familiar sequence differed from chance (0.5).

A significantly higher number of chicks approached the unfamiliar stimulus first (46 chicks out of 74, χ^2=4.37, p=0.036; Figure 2.3, left side). Consistent with this finding, a significantly higher proportion of time was spent by chicks near the screen playing the unfamiliar sequence ($t_{(73)}$=-2.69, mean=0.37, p=0.009, SD=0.41, Cohen's *d*=-0.317; see Figure 2.3, right side). A non-parametric test (one sample Wilcoxon signed rank) confirmed the latter finding (p=0.013).

Figure 2.3. *Left side.* Representation of the First stimulus approached. Along the Y axis, number of subjects are reported. The upper part of the column represents subjects that approached first the unfamiliar test sequence; the lower part represents subjects that approached first the familiar test sequence. *Right side.* Proportion of time spent by each chick near the screen presenting the familiar stimulus. The variable Error bars show standard errors. The dotted line indicates the chance level (0.5).


First stimulus approached

Proportion of time spent near the familiar stimulus

Chicks were clearly able to distinguish the familiar structured sequence from the unfamiliar, random presentation of the familiar items. Note that it is not surprising that chicks revealed the ability to recognize the familiar sequence by showing a novelty preference rather than by approaching the familiar sequence (see next paragraph).

Results of Experiment 1 suggested that visually-naive newborn chicks were sensitive to visual statistical patterns, being able to distinguish a familiar sequence from a random stimulus formed by the same elements. The next step consisted of investigating the robustness of this finding by comparing two structured sequences, and by increasing the number of elements composing the sequences.

2.1.6 Novelty preference

The tendency to explore the moderately novel stimulus with respect to the familiar one has been previously reported in this species especially when the procedure is based on a familiarization phase without reward (e.g., Bateson, 1974; Bateson, & Jaeckel, 1976; Bateson, 1979; Vallortigara & Andrew, 1991; Vallortigara & Andrew, 1994; Jones, Larkins & Hughes, 1996; Regolin et al., 2000). Evidence of novelty preference in chicks dates back to the 1970s, when Patrick Bateson developed a model predicting that a preference for the novel stimulus should appear whenever differences between familiar and unfamiliar objects are limited (Bateson, 1973). When the difference between familiar and unfamiliar stimuli is *small*, the model chooses the slightly novel, unfamiliar object. On the contrary, when the difference is *large*, the model prefers the familiar stimulus. The direction of the preference obtained in the present study fits this model: the small difference between familiar and unfamiliar sequences, provided by different order of pairs presentation while maintaining intact other perceptual aspects, may have led chicks to prefer the slightly novel stimulus.

On a more general note, the two test stimuli we used are formed by identical looming shapes, varying the order of their appearance. Therefore, as both sequences were familiar with respect to the elements and their modality of presentation, chicks preferred to explore the stimulus offering new information about the familiar elements. This is not very surprising if we consider the imprinting phenomenon by itself, operating in natural contexts where a social partner does not look the same when seen from the back or from the front.

2.2 EXPERIMENT 2

Results obtained in Experiment 1 suggested that newborn chicks are sensitive to the statistical structure of a visual stream of items since the animals differentiated the familiar structured sequence from a random stimulus composed of the same familiar elements. In Experiment 2, in order to investigate the robustness of this finding, test sequences were both characterized by a statistical structure. I reasoned that chicks' discrimination of test sequences in Experiment 1 could have been facilitated by the random nature of the unfamiliar one, since any random generation of discrete items provides high variability, especially when compared to a statistically-constrained appearance of the elements (familiar stimulus). Thus, Experiment 2 was designed to control for the variability of the two types of test streams, and to structured sequences have been created. In addition, the complexity of the stimuli was increased by adding a third shape-pair to both sequences.

2.2.1 Subjects

Forty-eight participants (all females)[5] were used for Experiment 2 since there was no significant difference between males' and females' performance in Experiment 1 (Chi-

---

[5] Males were not tested in Experiment 2. We strictly followed one of the 3 Rs principles (*Reduction*) applied to animal research (Passantino, 2008), leading us to minimize the number of subjects employed. We chose to use females because previous studies employing familiarization-discrimination paradigms with video-animated stimuli tested females only (e.g., Clarke & Jones, 2000; 2001). Moreover, females are sometimes found to respond better in tasks involving affiliative behavior and motivation to social reinstatement (e.g., Regolin, Marconato & Vallortigara, 2004).

square applied to 2x2 contingency table: $\chi^2=0.60$, $p>0.250$; Independent-samples t-test: $t=1.25$; $p=0.213$). The animals were hatched and housed in the same conditions described in Experiment 1.

2.2.2 Stimuli

The stimuli were designed according to the same logic used in Experiment 1, with the following changes. First, we included an additional pair of shapes (*T-shape* and *butterfly*), for a total of six shapes (RGB=140r, 49g, 8b; Color-Hex code #8c3108). As in Experiment 1, statistical relations defined the order of appearance of each item with TPs within-pairs = 1.0, and TPs between-pairs = 0.5 since sequential repetitions of the same pair were not allowed[6], (Figure 2.4). To create novel shape pairs we combined together perceptually dissimilar thus very discriminable shapes. Familiar Pairs were 1) *square-X shape*, 2) *circle-triangle*, and 3) *T shape-butterfly*. Unfamiliar Pairs were formed by re-combining familiar shapes into new pairs, and were 1) *circle-butterfly*, 2) *triangle-X shape*, 3) *square-T shape*. Unfamiliar Pair 1 is therefore composed by the first element of Familiar Pair 2, and the last element of Familiar Pair 3; Unfamiliar Pair 2 is composed by shapes that ended Familiar Pairs 2 and 1; Unfamiliar Pair 3 is formed by shapes that started Familiar Pairs 1 and 3. As in Experiment 1, E-Prime 1.1 was used to design the experiment and play the stimuli.

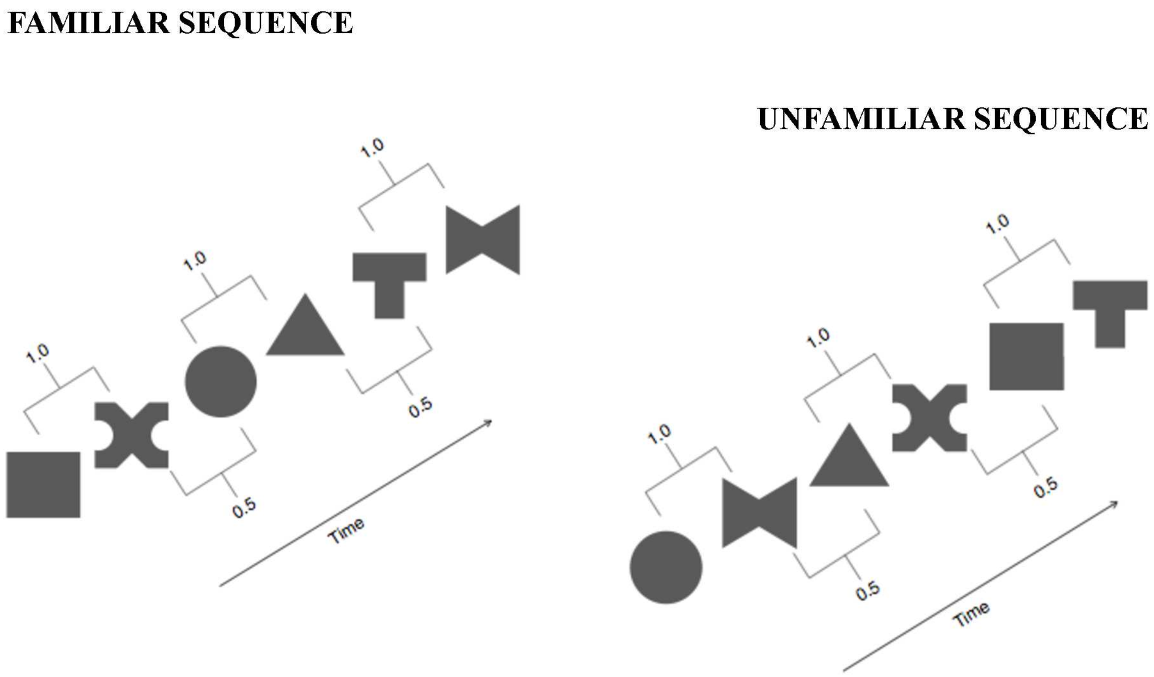
The second crucial manipulation concerned the structure of the unfamiliar sequence. Unlike Experiment 1, in which the unfamiliar test sequence consisted of a randomly-

---

[6] Sequential repetitions of the same pair were not allowed within a E-Prime cycle (that is a single presentation of the three pairs). However, repetitions were possible between the last pair of one cycle and the first pair of the following one (with a low probability of 0.33).

ordered presentation of the familiar shapes, the unfamiliar test sequence in Experiment 2 was composed of a novel structured combination of the familiar elements. In other words, the familiar statistical relations among shapes were reset by re-organizing the same elements into new pairs that the animals had never experienced before. For instance, in the familiar sequence the *circle* always predicts the *triangle* whereas in the unfamiliar one the *circle* always predicts the *butterfly*. This change made Experiment 2 more challenging than Experiment 1, in which chicks merely had to discriminate the familiar structure from a stimulus that was not characterized by statistical regularities. As the new unfamiliar sequence was also structured, chicks were required to learn the familiar pattern in order to differentiate it from the novel stimulus.

Figure 2.4. Schematic illustration of the structure of the familiar-test sequence (top, left) and unfamiliar test-sequence (bottom, right). Numbers and parentheses indicate transitional probabilities within (1.0) and between (0.5) elements.



FAMILIAR SEQUENCE

UNFAMILIAR SEQUENCE

The third difference from Experiment 1 was that test sequences were perceptually segmented by a 2-second visual break inserted in between each pair, formed by the alternation of a black screen and a grey screen (same color as the shapes: RGB = 90r, 90g, 90b; Color-Hex code #5a5a5a). The rapid transfer from a black screen to a grey one created a change in luminosity that worked as a flash, catching chicks' attention to the pause between pairs. The presentation of pairs as individual chunks increased the difficulty of the task compared to Experiment 1, in which test stimuli were presented as unbroken sequences. To successfully discriminate between the familiar and novel streams, chicks needed to recognize the statistical structure underlying the familiar stream which was presented in a perceptually different form. The combination of these three changes (an additional shape pair, a non-random unfamiliar pattern, and chunked test sequences) likely made the segmentation task in Experiment 2 more challenging than in Experiment 1.

2.2.3 Apparatus & Procedure

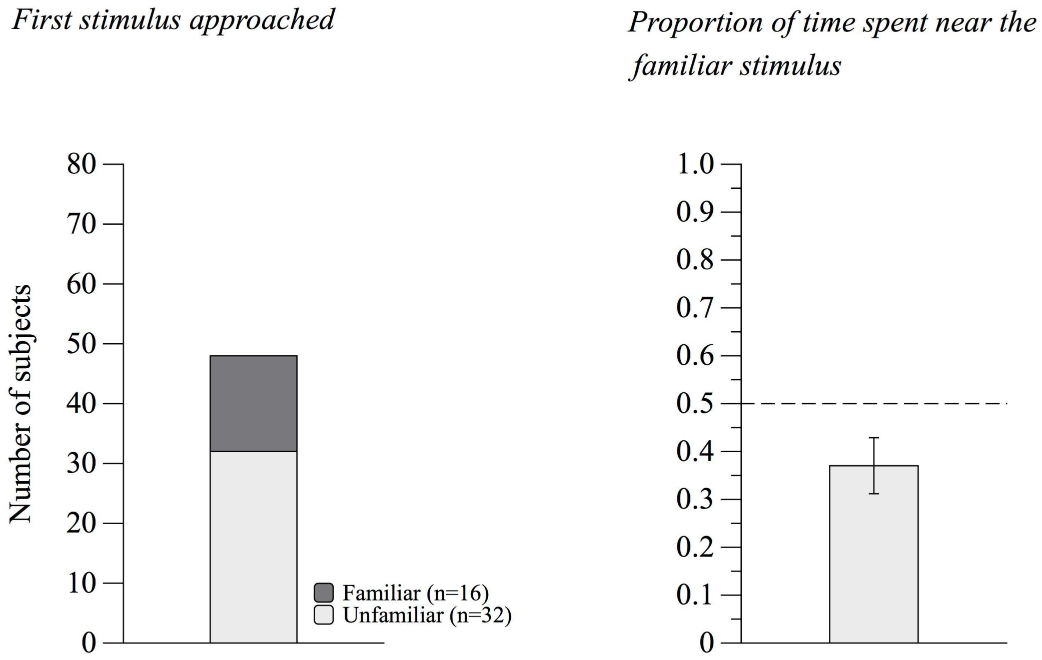
Identical to Experiment 1.

2.2.4 Results

Chicks' behavior was scored on-line by an experimenter blind to the purposes of the research and the 50% of the sample were also coded offline by a different coder, blind to both hypotheses of the research and stimuli position (IntraClass correlation coefficient, average measures=0.93, $p>0.001$). The dependent variables and statistical analyses were identical to Experiment 1. Results were similar to those obtained in Experiment 1. Chicks were still able to distinguish the familiar test sequence from the unfamiliar sequence, and this was expressed by chicks' choice of the novel stimulus. A significantly higher number of chicks first approached the unfamiliar stimulus (32 chicks out of 48, χ^2=5.33, p=0.021; Figure 2.5, left side) and, consistent with this finding, a significantly higher proportion of time was spent by chicks near the screen presenting the unfamiliar sequence ($t_{(47)}$=-2.21, mean=0.37, p=0.032, SD=0.4, Cohen's *d*=-0.325; Figure 2.5, right side). A non-parametric test (one sample Wilcoxon signed rank) confirmed the latter finding (p=0.042).

This evidence provides an important extension beyond the results of Experiment 1 because chicks recognized the familiar statistical pattern in the form of perceptually segmented pairs of shapes, discriminating it from an unfamiliar structured pattern consisting of the same shapes.

Figure 2.5. *Left side.* Representation of the First stimulus approached. Along the Y axis, number of subjects are reported. The upper part of the column represents subjects that approached first the unfamiliar test sequence; the lower part represents subjects that approached first the familiar test sequence. *Right side.* Proportion of time spent by each chick near the screen presenting the familiar stimulus. Error bars show standard errors. The dotted line indicates the chance level (0.5).



2.3 GENERAL DISCUSSION EXPERIMENTS 1 and 2

In two experiments, sensitivity to the probabilistic structure of a visual stimulus has been demonstrated in newborn chicks (Santolin, Rosa Salva, Vallortigara & Regolin, *under review*). The visual stimulus was presented as a continuous stream of shapes whose ordering was defined by transitional probabilities within/between pairs. After being exposed to a structured sequence and in the absence of any reward, chicks were capable of discriminating this pattern from a random generation of the same elements

(Experiment 1) and from a novel, structured combination of these elements (Experiment 2).

There are several striking aspects of this evidence. First, while studies on nonhuman species often use conditioning procedures based on extensive training sessions with food reward, in these experiments the learning process is unsupervised. Briefly, in unsupervised learning, the aim of the agent is to discover the hidden structure of a sequence of inputs without relying on an external teaching signal; that is, without receiving any feedback or reinforcement from the environment (Ghahramani, 2004). Unsupervised learning does not require any explicit output response following a sensory input, because the learning goal is a spontaneous extraction of useful regularities that allow to build an expressive, internal representation of the input patterns (Helmholtz, 1925; Hinton & Sejnowski, 1999). Since no feedback or reinforcement were provided to the chicks, the learning process took place through mere exposure to the stimulus, leading chicks to spontaneously extract the structure underlying the sequence. Second, these subjects belong to a species that does not possess a complex vocalization system and vocal learning abilities that could provide a selective advantage to, for instance, humans and songbirds. As described in Chapter 1, humans and some songbird species have shown the most sophisticated examples of statistical processing with respect to any other species tested so far. Third, a methodology based on visual exposure closely resembles procedures typically used with human neonates and infants, allowing to perform direct comparisons across the two species. Despite showing similarities between chicks' and infants' performance, these results highlight intriguing differences. Bulf and colleagues (2011) found that processing statistics from visual inputs may be still immature in 1- to 3- day-old neonates who showed no learning when more than 4 items composed the sequences; this is possibly due to constraints on the learning process regarding stimuli modality of presentation (e.g., spatial vs. sequential inputs; see Krogh et al., 2013, for further information) or to neonates' limited cognitive resources. Conversely, the current research demonstrated that visual statistical learning fully operates at the beginning of life, at least in the domestic chick. Day-old chicks' performance appears better than that of 1- to 3- day-old neonates since the animals successfully tracked statistical structure of both 4- and 6-element sequences. Notably, the 6-element sequence in Experiment 2 was not compared to a random presentation (as done with neonates and infants), but it was compared to a new combination of the familiar shapes, defined by a statistical structure itself. There are reasons to believe that this task was more demanding compared to that employed in Experiment 1 and to the tasks described in the neonate literature. Overall, in addition to being able to compute a stimulus comprising a larger number of items, chicks recognized the familiar structured sequence in contrast with another different structured sequence.

Chicks revealed segmentation abilities comparable to other nonhuman species such as, for instance, rats, which can distinguish familiar strings of syllables with high internal transitional probabilities from novel re-combination of the same syllables (words vs. non-words comparison described in Experiment 1 of Toro & Trobalòn, 2005).

2.4 EXPERIMENTS 3 and 4

In Experiments 1 and 2 it has been showed that newly-hatched chicks can differentiate visual stimuli displaying simple statistical regularities (e.g., a triangle always predicts a circle). In the present paragraph, I am going to describe two other experiments which were aimed at verifying whether chicks could do more complex computations than those performed in Experiments 1 and 2. The aim of these studies was to investigate chicks' ability to discriminate subtle differences between probabilistic structures of familiar and unfamiliar stimuli. I created a visual version of the *words vs. part-words* comparison described in Chapter 1 (e.g., Saffran et al., 1996; Toro & Trobalòn, 2005), comprising a string of items that always appear in that order (*words*) and a string of items that sometimes appear in that order (*part-words*), in the familiarization stream. Part-words in fact span word boundaries, being formed by the union of items that end a word, and items that start another subsequent word. Therefore, as the identity of the items composing a part-word strictly depends on what words follow one another during familiarization, the statistical coherence defining part-words is lower than that characterizing words. It seems quite clear that being able to discriminate this small difference requires a sophisticated level of statistical learning. I created two visual versions of the same contrast, named as pairs vs. part-pairs tested in Experiments 3 and 4.

The ultimate goal of this comparison would have been to explore what type of distributional properties defining the sensory information are learned by the animals. In Experiments 1 and 2 it was not possible to tell whether chicks really tracked transitional probabilities or whether they detected frequencies of co-occurrence within/between pairs. Transitional probabilities are a form of conditional probability so that the predictive relation between two arbitrary objects ("A predicts B") depends on the individual frequency of the first object (A). On the contrary, the frequency of co-occurrence represents how often two items occur together, with no predictive relation between them

(Miller & Selfridge, 1950). Unfortunately, since chicks did not succeed in the pairs vs. part-pairs comparison in both experiments, I was not able to pursue the latter aim of the research. Possible explanations of the non-significant results obtained in Experiments 3 and 4 will be proposed in Chapter 4.

2.4.1 EXPERIMENT 3

2.4.1.1 Subjects

A new group of 58 female domestic chicks was used. The animals were hatched and housed in the same conditions described in Experiment 1.

2.4.1.2 Apparatus & Procedure

Identical to Experiment 1.

2.4.1.3 Stimuli

The stimulus used in the familiarization phase was a continuous stream of 8 shapes[7] presented one at a time in the center of a computer screen. Similarly to Experiment 2, each shape was presented for 2 seconds and loomed from 2 to 10 cm in height, at a viewing distance of 50 cm (visual angle: 2.2°-11.2°). Shapes were created with Photoshop and have identical size of 316x316 px (resolution=71.6 ppi). As described in previous experiments, shapes were red (RGB= r140, g49, b8 e hex color code: 8c3108) in order to direct chicks' attention toward the familiar stimulus during the exposure

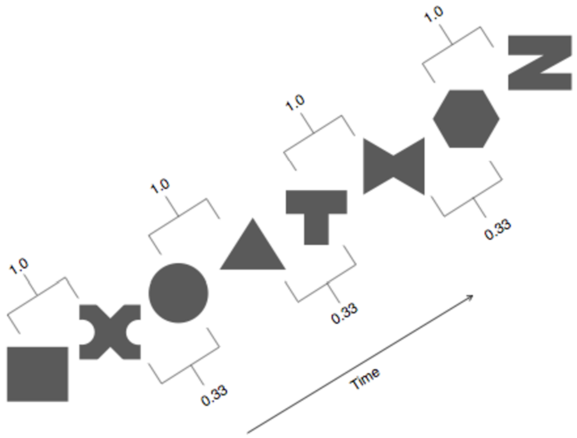
---

[7] Similarly to stimuli used in the words vs. part-words task (Saffran et al., 1996; Toro & Trobalòn, 2005), 8 shapes allow 4 pairs and 4 part-pairs.

phase. Six out of 8 elements were those used to create the sequences employed in Experiment 2 whereas the remaining 2 shapes were novel. All elements were organized into 4 shape pairs, defined by transitional probabilities within/between pairs. Pair 1 consisted of a *square* always followed by a *X-shape,* Pair 2 consisted of a *circle* always followed by a *triangle,* Pair 3 consisted of a *T-shape* always followed by a *butterfly*, and Pair 4 consisted of an *hexagon* always followed by a *Z-shape* (Figure 2.6). TPs within-pair=1.0 whereas TPs between-pairs=0.33[8]. E-Prime 1.1 was used to design the experiment and play the stimuli. Since there were no perceptual breaks in between elements, the only available cue to segment the familiar stream was the statistical structure of the sequence.

Figure 2.6. Schematic illustration of the structure of the familiar-test sequence (on the left) and unfamiliar test-sequence (on the right). Numbers and parentheses indicate transitional probabilities within (1.0) and between (0.33) elements.

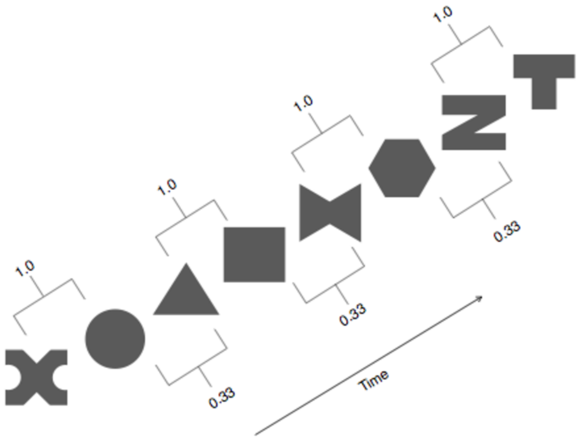
FAMILIAR SEQUENCE



---

[8] Similarly to Experiment 2, consecutive repetitions of the same pair were not allowed within a E-Prime cycle (that is, a single presentation of the three pairs). However, repetitions of the same pair were possible between the last pair of one cycle, and the first pair of the following cycle (with a low probability of 0.25).

UNFAMILIAR SEQUENCE



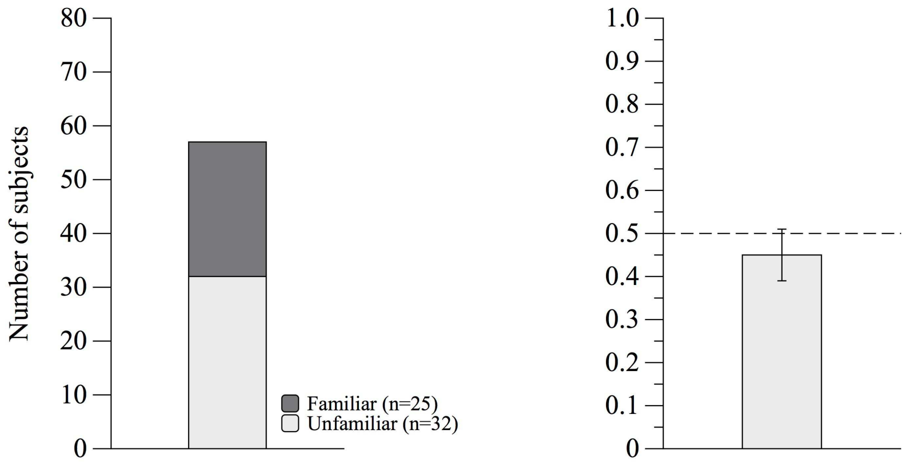
The test stimuli were the familiar sequence and an unfamiliar sequence which was composed of part-pairs. As mentioned above, part-pairs were by joining the last element of a familiar pair and the first element of another (subsequent) familiar pair. As shown in Figure 2.6 (right side), Part-pair 1 was formed by the *X-shape* and the *circle*, Part-pair 2 was formed by the *triangle* and the *square*, Part-pair 3 was formed by the *butterfly* and the *hexagon*, and Part-pair 4 was formed by the *Z-shape* and the *T-shape*. The key difference among pairs and part-pairs lied on the probabilistic structure of the two. For example, the *butterfly-hexagon* unfamiliar pair spanned the boundary between two familiar pairs: *T-shape-butterfly#hexagon-Z shape*. TPs within elements forming familiar pairs, thus within *T-shape* and *butterfly* as well as *hexagon* and *Z-shape*, were greater (=1.0) than TPs within elements forming a part-pair, thus within *butterfly* and *hexagon* (=0.33). This is due to the fact that each familiar pair could be followed by any of the remaining 3 pairs. The difficulty of the task is due to the fact that, by being composed of two familiar elements, part-pairs were experienced during familiarization but with a lower probability with respect to pairs. Therefore, in order to distinguish test sequences formed by pairs and part-pairs, chicks needed to detect a very small difference in conditional probability characterizing the two stimuli.

As reported in Experiment 2, test sequences were perceptually segmented by a 2-second visual break inserted in between each pair, and shapes composing the test streams were grey.

2.4.1.4 Results

The dependent variables and statistical analyses performed were identical to Experiment 2. Unfortunately, no statistically-significant preference emerged during test phase. The number of chicks which first approach was toward the familiar stimulus (n=25) did not differ from the number of chicks that approached the unfamiliar stimulus first (n=32; $\chi^2=0.86$, p=0.35; Figure 2.7, left side). A similar result was obtained from the analysis of the second dependent variable, the *Proportion of time spent near the familiar stimulus*. Chicks spent almost the same amount of time near the screens playing both test stimuli ($t_{(56)}=-0.69$, mean=0.45, p=0.49, SD=0.45; Figure 2.7, right side), suggesting that the animals were not able to differentiate them.

Figure 2.7. *Left side.* Representation of the First stimulus approached. Along the Y axis, number of subjects is reported. The upper part of the column represents subjects that approached first the unfamiliar test sequence; the lower part represents subjects that approached first the familiar test sequence. *Right side.* Proportion of time spent by each chick near the screen presenting the familiar stimulus. Error bars show standard errors. The dotted line indicates the chance level (0.5).



First stimulus approached

Proportion of time spent near the familiar stimulus

There are at least two possible explanations of chicks' failure in the present task. First, the animals did not distinguish test sequences because they could not track such a small difference characterizing the probabilistic structure underlying the two sequences. The animals may have thus perceived both stimuli as familiar, showing no preference. The second, alternative explanation of these findings relies on chicks' memory system. Since shapes appeared one at a time on the screen, it might be the case that chicks could not keep track of more than 6 elements presented as a temporal sequence. This could have negatively influenced chicks' capacity of detecting the structural difference between test sequences. However, before assuming that this species cannot track elaborate statistical patterns, I wanted to test this latter possibility. A new version of the current experiment has been created, whose stimuli were formed by a reduced number of elements (6 instead of 8). Any other aspect of the study remained unchanged with respect to Experiment 3.

2.4.2 EXPERIMENT 4

2.4.2.1 Subjects

A novel group of 24 female domestic chicks was used. The animals were hatched and housed in the same conditions described in Experiment 1.

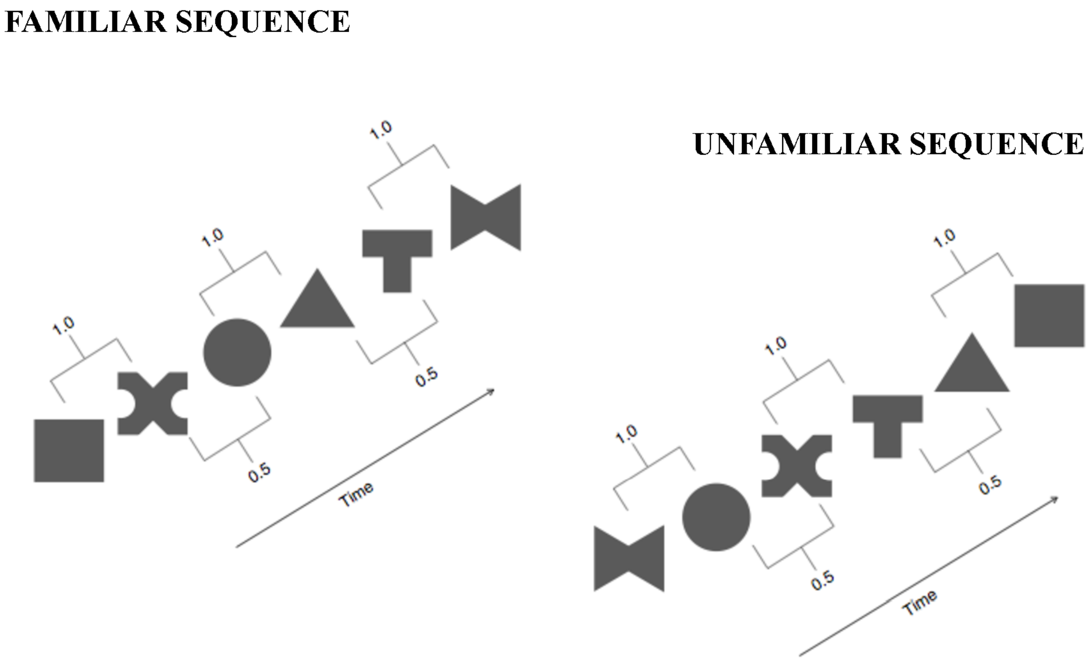
2.4.2.2 Apparatus & Procedure

Identical to Experiment 2.

2.4.2.3 Stimuli

The stimulus used in the familiarization phase was the same familiar sequence used in Experiment 2 thus formed by 6 shapes (Figure 2.8). Test stimuli were a sequence identical to the familiar one in terms of its underlying structure, whereas the unfamiliar sequence was composed of part-pairs constructed as follow. Considering all possible combinations of the familiar pairs, 3 part-pairs have been selected. As in Experiment 3, part-pairs were formed by the union of the last element of a familiar pair and the first element of another (subsequent) familiar pair. Part-pair 1 was formed by the *butterfly* followed by *circle*, Part-pair 2 was formed by the *X-shape* followed by *T-shape*, Part-pair 3 was formed by the *triangle* followed by the *square.* By assembling elements spanning boundaries across familiar pairs, the unfamiliar sequence was formed by transitions between elements (e.g., the triangle predicted the square) that chicks actually experienced during familiarization, but with a lower probability. Therefore, also in this case, chicks needed to recognize a subtle difference between probabilistic structures defining familiar and unfamiliar test sequences. Any other aspect (shape presentation, colors, computer screens) was identical to Experiment 3.

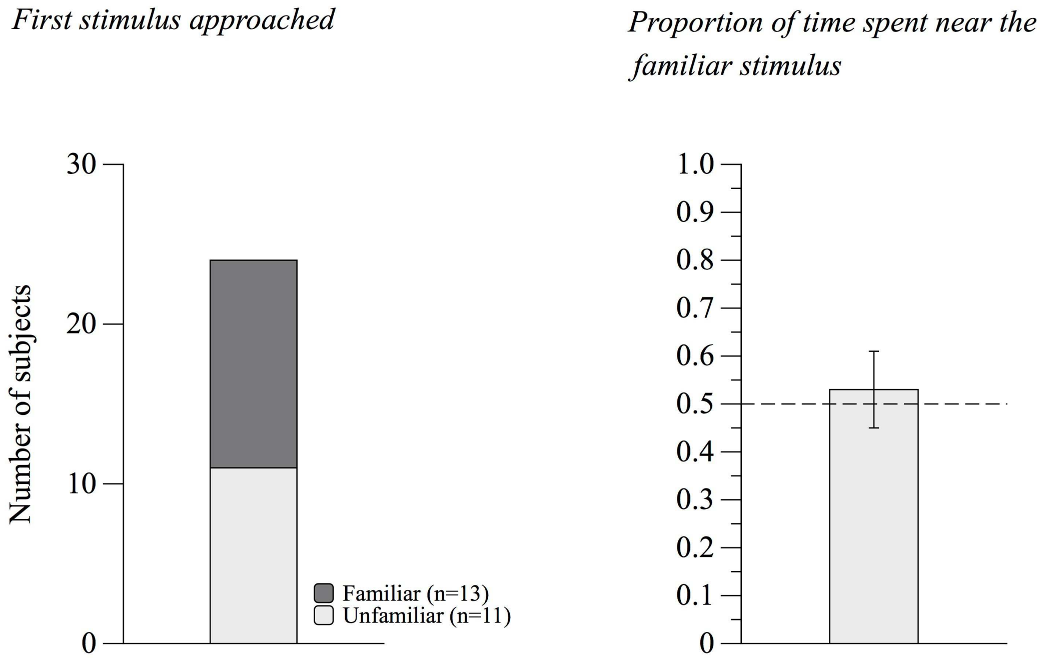
Figure 2.8. Schematic illustration of the structure of the familiar-test sequence (on the left) and unfamiliar test-sequence (on the right). Numbers and parentheses indicate transitional probabilities within (1.0) and between (0.5) elements.



2.4.2.4 Results

Dependent variables and statistical analyses performed were identical to Experiment 3. Again, chicks did not distinguish test sequences since no statistically-significant preference emerged during test phase. From the analysis of the *First stimulus approached*, the number of chicks which first approach was toward the unfamiliar stimulus (n=11) did not differ from the number of chicks that approached first the familiar one (n=13; χ^2=0.16, p=0.68; Figure 2.9, left side). A similar result was obtained from the analysis of the second dependent variable, the *Proportion of time spent near the familiar stimulus.* Chicks spent almost the same amount of time near the screens playing both stimuli ($t_{(23)}$ =0.36, mean=0.53, p=0.72, SD=0.42; Figure 2.9, right side). Please note that the sample size of this study is reduced with respect to previous experiments therefore, it might be premature to draw any conclusion about the fact that chicks cannot discriminate the two patterns. These data must be considered as preliminary, and further replications are needed. In spite of that, possible explanations of the overall non-significant results obtained in Experiments 3 and 4 will be provided in the next paragraph.

Figure 2.9. *Left side*. Representation of the First stimulus approached. Along the Y axis, number of subjects are reported. The upper part of the column represents subjects that approached first the unfamiliar test sequence; the lower part represents subjects that approached first the familiar test sequence. *Right side*. Proportion of time spent by each chick near the screen presenting the familiar stimulus. Error bars show standard errors. The dotted line indicates the chance level (0.5).



First stimulus approached

Proportion of time spent near the familiar stimulus

2.5 GENERAL DISCUSSION EXPERIMENTS 3 and 4

The non-significant results provided by Experiments 3 and 4 are not very surprising if interpreted in a general and ecological framework. My hypothesis is that, as the patterns to be learned become more complex, statistical learning is constrained by the perceptual nature of the stimuli. When regularities are relatively simple (Experiments 1 and 2), the learning process operates quickly and incidentally. However, given more complex patterns and computations such as those investigated in Experiments 3 and 4, statistical learning could be constrained by the perceptual modality of presentation.

Data from the human literature suggest that acoustic regularities are better learned when presented as a temporal sequence of sounds whereas visual patterns are better acquired when elements are simultaneously-visible (Saffran, 2001; 2002; Gebhart et al., 2009). Considering the nature of the sensory world, this is not surprising. In the auditory environment, sounds coming from a same source appear one at a time, and do not persist in time whereas visual scenes are likely to be formed by spatially-organized objects, visible all at once, with rare patterns of sequentiality (e.g., sign language, sequences of actions).

The findings described in this chapter may fit with this theory. The computations required to succeed in Experiments 1 and 2 can be plausibly considered simpler than those underlying Experiments 3 and 4. In the first two studies, chicks were asked to differentiate a familiar (structured) sequence from a random or a novel (structured) presentation of the same items; therefore, the familiar pattern was compared to something that the animals had never experienced during exposure. In contrast, the unfamiliar test sequences employed in Experiments 3 and 4 have been both experienced during familiarization, just with a lower probability as regards to the familiar pattern.

It might be the case that, when the difference in probability distribution amongst test stimuli is very subtle thus difficult to detect, perceptual aspects of the stimuli such as modality of presentation play a greater role in constraining the learning (Experiments 3 and 4), with respect to situations involving a more automatic process (Experiments 1 and 2). Ecologically speaking in fact, there might be no need for chicks to be able to master such a complex task involving temporally-presented visual stimuli. From a comparative point of view, rats, just like infants, successfully tracked such a small difference in transitional probabilities involving linguistic stimuli (the word vs. part-words comparison reported in Saffran et al., 1996 and Toro & Trobalòn, 2005). Why do rats show superior segmentation abilities of complex statistical patterns than those demonstrated by chicks? One explanation may refer to the coherence between the nature of the stimulus implementing the pattern (acoustic vs. visual) and the modality of presentation (sequential vs. spatial). In the words vs. part-words task, syllables were sequentially-presented to the rats, according to the stimulus-modality coherence. On the contrary, in the visual version of the same contrast chicks were presented with temporal sequences of objects. Coherence between type of stimulus and modality of presentation was thus respected for rats (sequences of syllables) but not for chicks (sequences of visual items).

Theoretical implications of these findings as well as future directions will be further discussed in Chapter 4.

CHAPTER 3 INVESTIGATING GENERALIZATION OF VISUAL REGULARITIES

Brief introduction

This chapter focuses on the generalization of visual structures to perceptually-novel stimuli. Generalization requires learners to go beyond the perceptual features of the input (i.e., the identity of the stimuli), and capture relationships between its components. A growing body of evidence has shown that generalization of simple patterns presented as strings of syllables or visual objects, is a quite wide-spread process that operates in multiple domains, modalities and species. However, as reported in Chapter 1, little is known about the earliest stages of this mechanism in the animal kingdom. Evidence about learning and generalization of regularities in newborn organisms is very limited since most part of the literature focused on nonhuman adult learners. This leaves open the issue of how this process operates at the very beginning of life: do newborn organisms with restricted experience with the sensory world show generalization capacities similar to those found in adult learners? Expanding this field would provide fundamental knowledge that would 1) clarify the ontogeny of this trait and, 2) allow comparisons between young human and nonhuman learners, shedding light on cross-species differences in early generalization abilities.

Aims & Hypotheses

The first aim of the present study is to investigate discrimination and generalization of visual regularities in a newborn animal model, the domestic chick (*Gallus gallus*).

Newborn chicks were reared in a controlled environment, with very limited experience with any visual stimulus prior to the experimental sessions. Afterward chicks were trained to recognize and generalize patterns displayed as triplets of simultaneously-visible shapes. I first considered the classic comparison AAB vs. ABA where A and B stand for arbitrary elements different to each other (Experiment 5).

The second aim of the study is to investigate whether generalization is constrained by perceptual aspects of the stimuli such as the presence/absence of adjacent repetitions. This research interest arose from the evidence reported in Chapter 1, suggesting that a repetition-detection mechanism positively influences learning and generalization of regularities in humans from the earliest stages of development. I was therefore interested in whether repetitions played such a predominant role in generalizing visual structures also in nonhuman species. To this purpose, I designed a second experiment aimed at testing comparisons where all patterns were characterized by adjacent reduplicated elements (AAB vs. ABB and AAB vs. BAA).

I hypothesized that newborn chicks should be able to 1) learn simple regularities such as AAB and ABA, and 2) recognize the same patterns implemented by novel items (Experiment 5). This hypothesis was based on the fact that the essential capacity for the survival of this species is learning about the environment within the first moments after hatching. It is therefore plausible that newborn chicks take advantage of regularities and contingencies characterizing relevant objects present in the environment (e.g., the familiar group of hen and siblings, edible items such as food, etc.). My prediction was that, if chicks use the presence/absence of reduplicated pairs of items as a cue to differentiate the stimuli, they would succeed in Experiment 5 but fail in Experiment 6.

However, I did not have clear hypotheses about whether or not configurations containing adjacent repetitions would be better learned than others (Experiment 6).

3.1 EXPERIMENT 5

In the present experiment, the comparison tested was AAB vs. ABA. A and B do not have specific identities but can be implemented by perceptually-different elements. AAB displays an asymmetrical structure, being formed by an adjacent repetition located at the beginning of the pattern. On the contrary, ABA possesses a non-adjacent repetition that confers a symmetrical structure to the pattern.

Two-day-old domestic chicks were trained to find a food reward hidden behind one of two plastic screens, placed in a triangular arena. The rewarded screen displayed a triplet of geometric shapes following, for instance, the AAB pattern; the neutral screen (no food behind it) displayed the same shapes organized according to a different pattern (i.e., ABA). Once completed the training session, chicks performed a generalization test with triplets composed of novel shapes, never seen by the animals, but implementing the training patterns.

3.1.1 Subjects & Rearing conditions

Subjects were 8 chicks (*Gallus gallus*), 4 males and 4 females, obtained from the commercial hatchery Agricola Berica few hours after hatching or as fertilized eggs to be hatched in the laboratory as described in Experiment 1 (Chapter 2). All the chicks were housed in pairs in standard metal cages (28x32x40 cm). The rearing room was kept under control for temperature (28–31 °C) and humidity (68%), and it was constantly illuminated by fluorescent lamps (36 W) located 45 cm above each cage. Food and water were available *ad libitum* in transparent glass jars (5 cm in diameter, 5 cm in high), placed in the center of the each cage. During the first day of life chicks were reared in pairs, while familiarizing to the housing environment. In order to get used to visual separations from conspecifics, the animals were divided and housed individually during Day 2 (individual training and testing is distressful to pair-reared chicks). They were also allowed to eat a few mealworms (*Tenebrio molitor* larvae) in their home cage as this kind of food will be used as reinforcement during the whole training session. Training lasted from Day 3 to Day 5, when testing took place. At the end of the test phase (no later than Day 6) all chicks were caged in pairs with food and water available *ad libitum*, and after a few hours they were donated to local farmers.

3.1.2 Stimuli

Training stimuli were 10 triplet pairs, composed of shapes similar to those used in Fiser & Aslin (2002), and arranged according to AAB and ABA. One shape was repeated twice (A element) and the other one was presented only once (B element). Within a pair, the two triplets represented different structures: one triplet implemented AAB (e.g., circle-circle-cross), the other triplet implemented ABA (e.g., circle-cross-circle; Figure 3.1, top left). All stimuli were spatially-organized in a horizontal row and circumscribed in a black frame (5.15x2.18 cm), thus the three elements were all visible at the same time. Each stimulus was printed on a cardboard support and attached on the plastic screen (b=14 cm, h=18 cm). The cardboard was approximately aligned with the chick’s central visual field to guarantee an appropriate view of the entire stimulus.

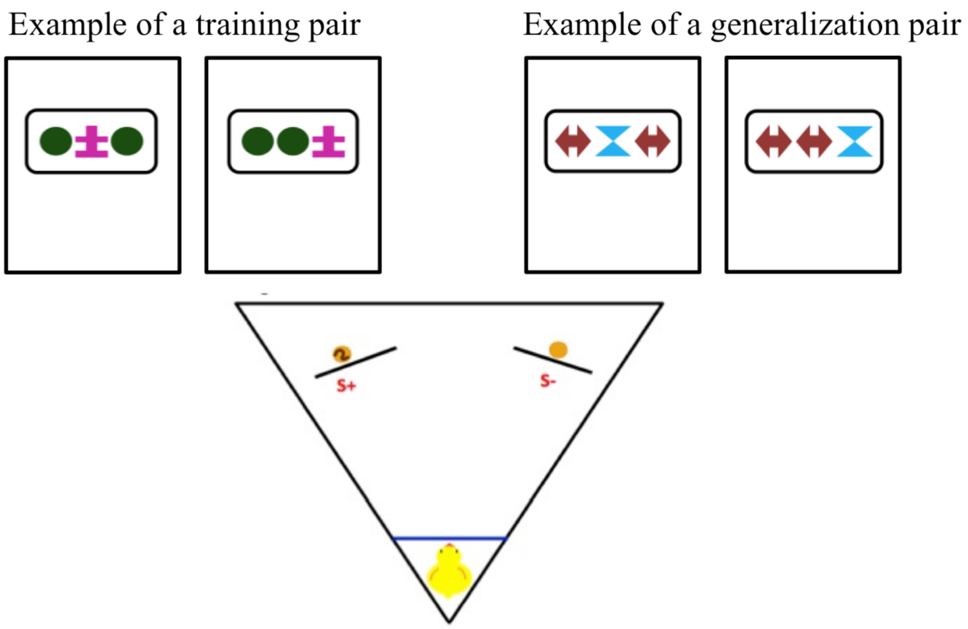
The animals were individually trained with only 5 triplet pairs; half of the sample was trained with pairs 1 to 5 (see Appendix, Training sequences 1) and the other half was trained with pairs 6 to 10 (see Appendix, Training sequences 2). Several exemplars of the training patterns were used to minimize the learning of some idiosyncratic regularity based on the perceptual aspects of the triplets (for a discussion, see ten Cate & Okanoya, 2012). For 4 chicks (2 males), AAB was the rewarded regularity (*S*+) whereas ABA represented the neutral pattern (*Sn*, not associated with food nor punished). For the other 4 animals (2 males), ABA represented *S*+ whereas AAB was *Sn,* so that the same stimuli were reinforced for one group of animals, and were neutral for the other group. During the generalization test all chicks were presented with 10 perceptually-novel triplet pairs, implementing the same stimuli AAB vs. ABA (Figure 3.1, top right). Since new shapes and colors formed the generalization triplets, the only difference occurring between generalization and training stimuli was the perceptual identity of the elements (see Appendix, Test sequences).

3.1.3 Apparatus

Training and test phases were carried out in a quiet experimental room, with temperature and humidity maintained at 25°C and 70% respectively. The apparatus showed in Figure 3.1 was used for both experimental sessions, and it consisted of a triangular arena whose walls (25x100 cm) and floor were uniformly lined with white plastic sheets. One of the three vertexes of the arena was the chick's starting position, delimited by a removable clear glass partition (20x10 cm), allowing it to see the inside of the arena. It is worth noting that the experimental setting (stimuli position and food reward) were positioned before placing the animal in its starting point.

During training, two identical plastic screens depicting the stimuli were positioned symmetrically in front of the starting point (35 cm away from it, and 20 cm apart from one another). Behind each screen a small circular dish (d=3 cm) was located however, only the dish behind the screen representing *S*+ was baited with food. By placing the dishes behind both screens, the animals could not tell which one was hiding the food until they detoured the correct screen, and looked behind it. The only way to discriminate the screens was by differentiating the pattern depicted in the frontal side of them. The experimental setting was identical for both training and test phases, except that during test any food reward was provided to the animals (the dishes behind the screens were both empty).

Figure 3.1. The upper part of the picture shows representations of the screens depicting two examples of training (on the left) and generalization (on the right) pairs. The lower part of the picture shows a schematic representation of the apparatus used for training and test (bird's eye view). Oblique lines indicate the position of the two screens within the apparatus. Food-reinforcement is represented by a schematic illustration of a mealworm inserted in the plastic dish (the orange dot). The blue line represents the clear-glass partition behind which the chick was placed at the beginning of each trial (starting point).



3.1.4 Procedure

To follow, a detailed description of training and test sessions is provided. Each animal was trained and tested individually.

Training phase. On the morning of the Day 3, following 2 hours of food deprivation (water was available *ad libitum*) chicks underwent to a 20-minutes pre-training session aimed at familiarizing the animals with the experimental environment, and learning the detour response. Through shaping, chicks were trained to go behind a screen to reach the food reinforcement. The screen was blank, without any stimulus depicted on it, and was placed centrally in the arena to reduce the possibility to develop a positional bias i.e., preference for the left or the right vertex of the apparatus. Training with the first triplet pair began as soon as the detour response was established. At the beginning of every training trial, the chick was placed behind the glass partition defining the starting point, and it could see both stimuli for 10 seconds. The partition was then removed, and the subject could express a choice for one of the screens. If the chick went behind the *S+* screen, it could eat from the baited feeder, and a correct response was scored. In contrast, if the chick chose the *Sn* screen, the response was scored as incorrect and the chick was not allowed to reach the food behind the other screen, being immediately placed back in the starting point until the next trial started. The detour response was considered valid whenever the chick's head was aligned with one of the horizontal margins of the screen or when that point was surpassed (as visible through the camera). Left-right position of the stimuli in the apparatus followed a semi-random order so that the rewarded triplet never appeared more than two consecutively times in the same position. The learning criterion was established at 17 correct choices out of 20 consecutive trials; once it was reached for the first pair, training of the second pair immediately started. As soon as the subject reached the criterion for the fifth pair, the last training session (shuffle phase) was carried out. The shuffle phase consisted of 20 trials in which all 5 training pairs were semi-randomly presented (no consecutive repetitions of a same pair were allowed). This session was aimed at familiarizing chicks with a rapid exposure to several exemplars of the patterns, similarly to what shall occur in the subsequent generalization test. The shuffle phase ended once the animals reached again the learning criterion (17 correct choices/20 consecutive trials). After this last training session, chicks spent a 2-hours resting period in their home cage before the generalization test started. The number of training trials conducted each day flexibly varied based on the animals' motivation to perform the task. Whenever the chick did not exhibit a detour response within 2 minutes after being released from the starting point, I assumed that the animal was no longer motivated to continue. The training was then suspended and the chick was placed back in its home cage without food in order to regain motivation (water was instead available *ad libitum*). On average chicks completed the training sessions between Day 3 and Day 5.

Generalization Test. This session included 20 probe trials featuring perceptually-new triplet-pairs implementing the AAB vs. ABA comparison. Over the 20 probe trials, each test pair was presented twice, and in a semi-random order so that the same pair never appeared twice in a row. The order of presentation of test pairs was counterbalanced between subjects. The 20 probe trials were alternated with 30 training trials, which were presented in a semi-random order, and reinforced as in the training phase in order to avoid response extinction. On the contrary, probe trials were not associated with food reward.

Chicks' responses in the probe trials were coded online by recording the first screen detoured. The entire sample was also scored off-line by a different coder blind to experimental hypotheses and conditions in order to confirm on-line coding accuracy. For subject n.8, only the first 10 test trials were scored off-line because of technical failure related to the video of the last 10 test trials). Results showed high consistency of the two codings (Intraclass correlation coefficient, average measure=0.98; p<0.001).

3.1.5 Data analysis

For the training phase, the mean number of trials needed to reach the criterion for each triplet pair and for shuffle phase was calculated. For the generalization test, the dependent measure analyzed was the number of trials during which chicks chose the triplet resembling the trained pattern (regarded as the correct choice) over 20 probe trials. The mean of correct choices has been compared to the chance level (10), using a one-sample two-tailed t-test. Individual performances have been also measured using a non-parametric (one tail) Binomial test; a cut off of 15 correct trials out of 20 has been set, establishing the significance of the test.

3.1.6 Results

Training phase

On average, chicks needed 204.25 (SD=49.11) trials to complete the entire training session. The animals needed a progressively lower number of trials to reach the learning criterion as the training went on (1st trained pair = ~66 trials - 5th trained pair = ~27 trials).

Generalization Test

Chicks chose the correct pattern significantly above chance (t_7=11.121, p=0.000, mean=16.63, SD=1.68, effect-size *r*=0.97), (Figure 3.2, left column), suggesting that they recognized the familiar regularity even when implemented by previously-unseen items. Individual results showed that 7 out of 8 animals performed significantly above chance scoring 15 or more correct responses out of 20 test trials (p=0.02; Table 3.1), confirming the result obtained at the group level.

These findings revealed chicks' capacity to recognize the trained regularity (both AAB and ABA) and, to generalize to novel exemplars of the same pattern. The variety of stimuli used during training and test decreased the chance that chicks' performance reflected a response to low-level features of the stimuli (luminance, contrast, color, shape, etc).

Having obtained evidence of chicks' ability to discriminate and generalize AAB and ABA, I wanted to investigate which mechanisms chicks may have used to do so. Chicks could have indeed discriminated the patterns based on the asymmetrical structure of AAB with respect to ABA. As mentioned earlier, this comparison comprises an asymmetrical configuration (AAB) characterized by an adjacent reduplicated element, and a symmetrical structure (ABA) characterized by a non-adjacent reduplicated element, interleaved by a different one. Therefore, chicks may have easily learned to differentiate these patterns based on this difference. In order to control for this plausible strategy new contrasts were created, in which all patterns included an adjacent repetitions that conferred an asymmetrical structure to the stimuli.

3.2 EXPERIMENT 6

The current experiment was aimed at investigating the role of adjacent repetitions in discrimination and generalization of visual regularities. All regularities compared in this experiment comprised an adjacent reduplicated item that gave the triplets an asymmetrical structure. In Condition 1, AAB was compared to ABB; the first pattern was characterized by a repetition placed at the beginning whereas the second pattern was characterized by a repetition placed at the end. In Condition 2, AAB was instead compared to BAA, again characterized by a repetition located at the final edge of the pattern. Testing these configurations allowed to explore the role played by the presence/absence of adjacent repetitions and perceptual symmetry/asymmetry in the generalization process.

3.2.1 Subjects & Rearing conditions

A novel group of 8 domestic chicks was used: 4 (2 males) were randomly assigned to Condition 1 and the other 4 to Condition 2. The rearing environment was identical to Experiment 5.

3.2.2 Stimuli

New triplets of shapes instantiating AAB, ABB and BAA were created. As mentioned above, the crucial manipulation of this study is the fact that all patterns were characterized by asymmetrical structures, conferred by the presence of an adjacent repetition.

Additionally, new shapes and colors were employed to construct training and test triplets improving the stimuli design (for examples, see Figure 3.3 and 3.4; for the whole sets of stimuli see Appendix). Each shape could be inscribed in an invisible square of 2x2 cm, and the colors employed were just 4 complementary shades: red and orange (warm hues), and green and blue (cold hues), (e.g., Jones, Osorio & Baddeley, 2001). Previous findings showed that chicks can differentiate these colors and categorize them based on perceptual similarities, similar to what humans do (Ham & Osorio, 2007). A pilot study with 4 chicks also revealed that the animals were capable of learning and generalizing ABA vs. AAB implemented by these new forms and colors (similarly to what obtained in Experiment 5, chicks' average performance at test was of 15.5/20 correct trials). Triplet-pairs were printed on photographic paper and controlled for the absence of UV-radiations with a spectrophotometer. In order to avoid generalization based on textural properties of the stimuli, color contingencies between training and generalization stimuli was reversed. One of the warm hues (red or orange) was assigned to the A items implementing the training set, and one of the cold hues (blue or green) was assigned to the A items implementing test stimuli. As in Experiment 5, shapes were simultaneously-visible, spatially organized in a horizontal row, and circumscribed in a black frame.

Figure 3.3. On the left, example of triplet pair used in training phase and, on the right, example of triplet pair used in generalization phase. Both pairs represent the AAB vs. ABB contrast (Condition 1).

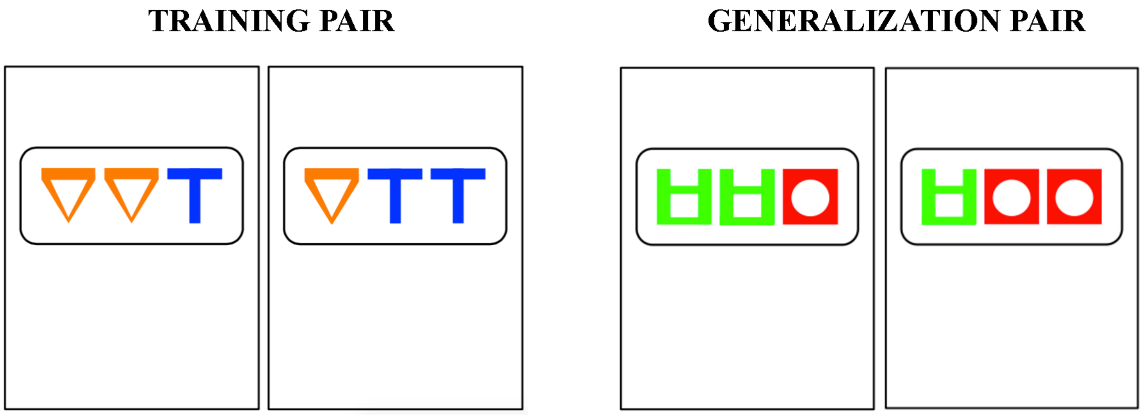
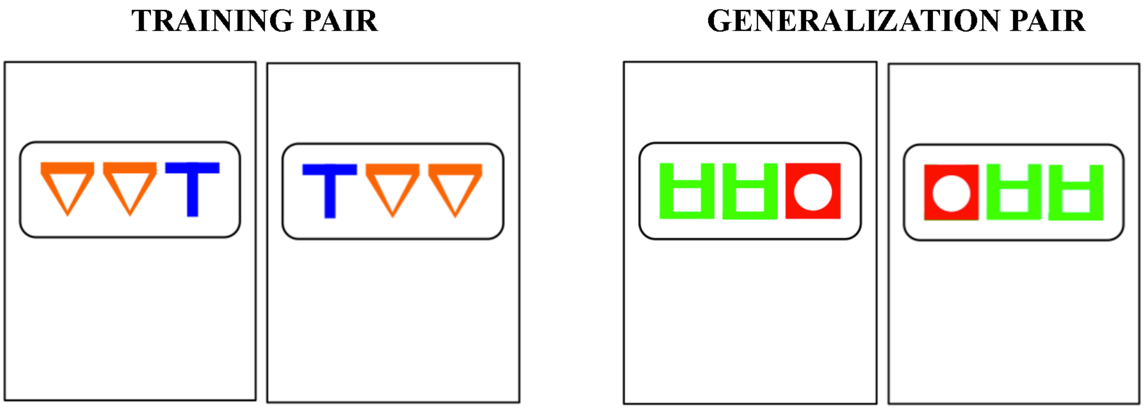


Figure 3.4. On the left, example of triplet pair used in training phase and, on the right, example of triplet pair used in generalization phase. Both pairs represent the AAB vs. BAA contrast (Condition 2).



3.2.3 Apparatus & Procedure

Apparatus and procedure used in training and test sessions were identical to Experiment 5. In order to confirm on-line coding accuracy, the entire sample was scored off-line by a second coder, blind to experimental hypotheses and conditions. Results showed high consistency between coders (Intraclass correlation coefficient, average measure=.98; p<0.001).

3.2.4 Data analysis

Dependent measure and statistical analyses were identical to Experiment 5. In addition, a Mann-Whitney U test was performed to compare the two experimental groups

(Condition 1 and Condition 2)[9], and an Independent-samples t-test was operated to draw a final comparison between experiments.

3.2.5 Results

Training Phase

To complete the training session, chicks assigned to Condition 1 needed on average 275.25 (SD=90.26) trials whereas chicks assigned to Condition 2 needed 196.75 (SD=48.57) trials.

Generalization Test

A Mann–Whitney U performed on the number of correct responses over 20 probe trials revealed a non-significant difference between conditions (U=9.500; p=0.686), thus data of the two groups were pooled together and considered as a single group for subsequent analyses. The mean of correct choices operated during test phase showed that chicks chose the triplet resembling the reinforced pattern significantly above chance (16.13 of correct test trials over 20; t_7=10.032, p=0.000, SD=1.72, effect-size *r*=0.97), (Figure 3.2 right column). This result was also confirmed by the analysis of individual performances; in both conditions, 3 out of 4 subjects performed significantly above chance, scoring 15 or more correct responses out of 20 test trials (p=0.02; Table 3.1).

---

[9] This non-parametric test was applied given the small sample size of the two groups.

Figure 3.2. Number of correct responses over 20 test trials of Experiment 5 (on the left) and 6 (on the right). Error bars show standard errors. Asterisks indicate significant (p<0.001) departures from chance level (that indicates 10 correct responses), marked by the dotted line.

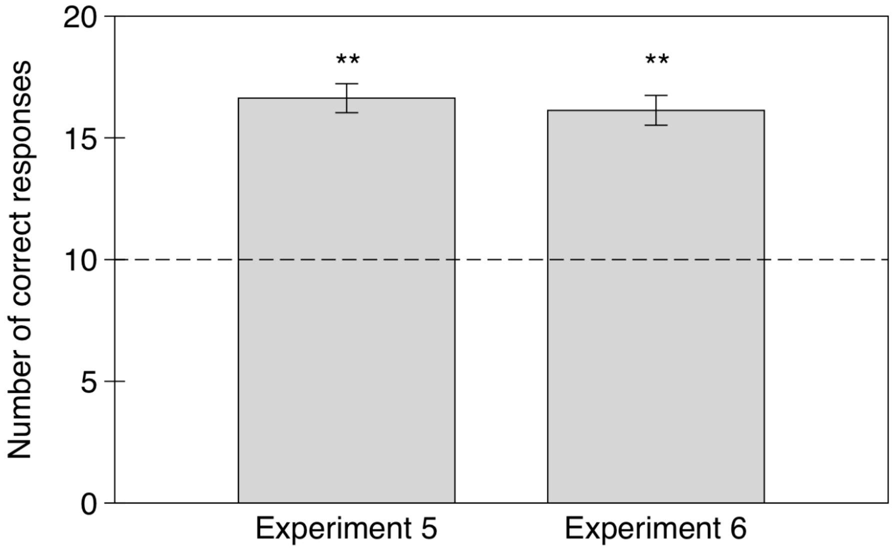


Table 3.1. Individual performances at the generalization test for Experiments 5 and 6.

EXPERIMENT 5 AAB vs. ABA	Correct choices/20	**EXPERIMENT 6**	Correct choices/20
Subject 1	18	Subject 1 (AAB vs. ABB)	17
Subject 2	14	Subject 2 " "	16
Subject 3	18	Subject 3 " "	14
Subject 4	17	Subject 4 " "	19
Subject 5	19	Subject 5 (AAB vs. BAA)	14
Subject 6	16	Subject 6 " "	15
Subject 7	15	Subject 7 " "	17
Subject 8	16	Subject 8 " "	17

A comparison between experiments revealed no differences related to the dependent measure (Experiment 5: mean 16.63, SD=1.68; Experiment 6: mean=16.13, SD=1.72; $t_{(7)}=0.586$, $p=0.567$), supporting the idea that chicks can extract visual patterns, independently of the presence/absence of adjacent repetitions. Further discussion of this evidence will be provided in the next paragraph.

Results of Experiment 6 have some implications for chicks' learning abilities. During training of Condition 1 (AAB vs. ABB), the rewarded regularity could have been learned by detecting the different identity of the repeated element within pairs (AA vs. BB). For example, for a given training pair, chicks could have just learned that the reinforced pattern was the stimulus where the red cross was repeated as opposed to that where the repetition involved the green hourglass. However, as test stimuli were formed by new colors and forms, this strategy would not have been sufficient to support generalization. On the other hand, Condition 2 (AAB vs. BAA) allowed to exclude that at training chicks differentiated the patterns solely based on the identity of the repetition. Following the above-mentioned example, subjects could not have learned that the reward was associated to the reduplicated red cross since that shape was repeated in both triplets composing a pair. In fact, results revealed that chicks distinguished AAB when compared to BAA (and vice versa), indicating that they can discriminate structures characterized by *identical* adjacent repetitions (AA vs. AA) placed in different positions within the triplets. It is worth noting that this evidence confirms the results obtained in Exp. 1 that showed, for the first time, that chicks can recognize the training pattern even when instantiated by novel elements. Results of Experiment 6 give further support to this initial evidence, demonstrating that chicks can do that even in the most challenging contrast that comprises identical repetitions in both triplets forming a pair.

3.3 GENERAL DISCUSSION EXPERIMENTS 5 and 6

In two experiments it has been showed that newborn chicks can detect regularities implemented by triplets of simultaneously-visible elements (Santolin, Rosa Salva, Regolin & Vallortigara, *submitted*). Generalization is revealed by chicks' capacity to respond differentially to pattern-violating and pattern-following test triplets formed by novel elements. This evidence suggests that chicks are able to abstract away from the surface information of the stimuli (forms and colors of the training stimuli), picking up underlying structures.

Interestingly, based on the results provided by Experiment 6, chicks seem to equally generalize patterns that are not discriminable on the basis of perceptual symmetry/ asymmetry of their structures, and presence/absence of adjacent reduplications. This indicates that generalization does not prioritize regularities with repetitions or with a certain type of structure in this species. Further discussion of this issue will be provided in Chapter 4.

The current experiments provide the first evidence of generalization of regularities presented as triplets of visual elements in a newborn animal model. Even though the learning was obtained through a conditioning procedure (whereas generalization was unreinforced), it is plausible that chicks are predisposed to track relations between objects characterizing the visual environment within the first days of life. Controlled-reared studies indeed imply that abilities exhibited by young animals are likely to be the result of early predispositions (Vallortigara, 2012). Data from this research can not be directly compared to other species as regards to stimuli and methodologies employed as well as the age of the animals. Nevertheless chicks demonstrated similar generalization capacities to those revealed by rats (Murphy et al., 2008; de La Mora & Toro, 2013), rhesus monkeys (Hauser & Glynn, 2009) and, most of all, human infants (Saffran et al., 2007; Johnson et al., 2009).

Theoretical implications of these findings as well as future directions will be further discussed in Chapter 4.

CHAPTER 4 CONCLUSIONS

The present research started because there was no general agreement about whether and how statistical learning and generalization of visual regularities fully operate at birth. A growing body of research have demonstrated the existence of these processes at different stages of human development and in different species, revealing that statistical learning and generalization are quite wide-spread mechanisms in the animal kingdom (see ten Cate & Okanoya, 2012; Krogh et al., 2013; Frost et al., 2015; Dehaene, Meyniel, Wacongne, Wang & Pallier, 2015 for review of the literature). However, very little was known about the very beginning of these mechanisms.

The research illustrated in this thesis provides novel data showing that statistical learning and generalization of visual structures fully function at birth. The newborn domestic chick (*Gallus gallus*) has been validated, for the first time, as behavioral model for this topic of investigation. The results obtained in the present work are consistent with the *generality* assumption underlying these cognitive traits. The domestic chick indeed belongs to a species phylogenetically more distant to humans than the nonhuman primates tested so far. Primates in fact may share linguistic precursors with the human being. In addition, evidence of statistical learning and generalization obtained in this model seem consistent with those found in humans and songbirds, in spite of the fact that chicks' skills are unrelated to language and communication. In the next section I will briefly summarize the principal findings obtained in this thesis, and propose possible future lines of investigations.

Chapter 2 provides the first evidence of sensitivity to the probabilistic structure of a visual stimulus in a newborn animal. The visual input was a continuous stream of shapes whose ordering was defined by transitional probabilities (e.g., a triangle always predicted a circle). After solely being exposed to this sequence, chicks were capable of discriminating its structure from a random generation (Experiment 1) or from a novel, structured combination (Experiment 2) of the same elements. Chicks thus showed spontaneous processing of a statistically-defined visual input comparable to that shown by other species. Furthermore, while studies on nonhuman species regularly use food-rewarded conditioning procedures, chicks were just exposed to the visual stimulus before being tested. Since no reinforcement nor feedback were given to the animals, the type of learning underlying this mechanism was unsupervised. Moreover, a procedure based on visual exposure allowed a comparison between newborn chicks and neonates, and revealed interesting differences. Day-old chicks performed better than of 1- to 3- day-old neonates (Bulf et al. 2011) as the chicks successfully detected the statistical structure underlying sequences comprising up to 6 temporally-presented elements. On the contrary, neonates showed no learning when more than 4 items formed the sequences. Besides being able to compute a stimulus formed by a larger number of elements, chicks recognized the familiar structured sequence compared with another different structured sequence. To the best of my knowledge this contrast has never been described before in the literature. Overall, contrary to what showed by humans, processing simple regularities from visual inputs is a fully functioning process at birth, at least in the domestic chick.

One question raised by these findings is the possibility that chicks' ability to track visual patterns is limited specifically to the context of filial imprinting. Imprinting can be considered as a process that relies on the extraction of regularities. More precisely, it is a type of learning-by-exposure that works in an unsupervised way since the learner does not require an external teacher who signals the specific features defining salient objects in the environment (Goldstone, 2003). Indeed, the principal survival function of this phenomenon for newborn chicks is identifying and categorizing a relevant object in the environment by learning its salient features (e.g., conspecifics' facial features, plumage colors, etc). Additionally, this mechanism has to be powerful enough to detect invariant structures of the familiar object no matter how it appears (e.g., occluded by a natural element or under different angles of observation), in order to enable chicks to develop an internal representation of that familiar object (Bateson, 1990). For these reasons, it might be the case that chicks' learning by solely being exposed to the stimulus was enhanced by the innate tendency of this species to learn the features of the first salient object seen immediately after hatching. To further address the plausible close relation between filial imprinting and statistical learning, it would be interesting to test chicks using a paradigm that does not involve filial imprinting for instance, a conditioning procedure similar to that employed in Chapter 3. Furthermore, it would be interesting to test other precocial species characterized by this phenomenon such as ducklings, goslings and quail chicks to verify whether they succeed in statistical learning tasks similar to those presented here. Even more fascinating it would be testing altricial species that exhibit reduced filial imprinting to see if they differ in their ability to track regularities from the environment within the same visual statistical learning paradigm.

The second series of experiments reported in Chapter 2 (Experiments 3 and 4) were aimed at getting a neat understanding of chicks' statistical learning capacities. I was in fact interested in verifying whether chicks could do more complex computations than those explored in Experiments 1 and 2, such as detecting subtle differences in probability distribution characterizing familiar and unfamiliar sequences. Essentially, test stimuli were composed by the same shapes but differed from each other based on high or low transitional probabilities within shape-pairs. The familiar sequence was formed by pairs with high internal transitional probabilities meaning that the shapes forming those pairs had higher probability to appear in that order with respect to the pairs composing the unfamiliar sequence (characterized by low internal transitional probabilities). The difficulty of this task is given by the fact that the part-pairs forming the unfamiliar stimulus were created by assembling together the last shape of a familiar pair and the first shape of another familiar pair. Therefore, the order of shape appearance characterizing the unfamiliar stimulus has been experienced during familiarization but with a lower probability. Detecting the structural difference between familiar and unfamiliar sequences was thus more challenging than discriminating test stimuli presented in Experiments 1 and 2, whose unfamiliar sequences had never been experienced before. Unfortunately, the animals could not differentiate familiar and unfamiliar streams showing no discrimination at test when sequences comprised 8 (Experiment 3) or 6 (Experiment 4) items. My last goal would have been to explore which type of distributional properties are learned by this species. In Experiments 1 and 2 chicks could have succeeded by simply tracking frequencies of co-occurrence rather than picking up transitional probabilities. As mentioned earlier in this thesis, computing frequencies of co-occurrence requires detecting what co-occurs with what, and how often does this happen, whereas computing transitional probabilities requires detecting what predicts what, based on different levels of probability. Since the animals were unable to distinguish strings of items linked by high transitional probabilities from strings of items linked by low transitional probabilities (Experiments 3 and 4), the investigations were terminated, leaving open the issue of whether chicks are effectively sensitive to transitional probabilities or not.

One possible explanation of the fact that chicks could not tease apart familiar and unfamiliar stimuli in Experiments 3 and 4 points to the ecological value of the experimental setting. The evidence obtained in Chapter 2 appears to be consistent with a theory based on the primary essence of the sensory input. Research conducted on human subjects have demonstrated that acoustic regularities are better acquired when presented as a sequence of sounds and visual patterns are better learned when elements are visible all at once (Saffran, 2001; 2002; Gebhart et al., 2009). This has been explained considering the nature of the sensory information which is predominantly composed of sounds (of a same source) appearing one by one, and simultaneously-visible visual objects persisting over time. Results reported in Chapter 2 may fit with this theory. It might be the case that, when the computations required are quite simple (e.g., discriminating a familiar sequence from previously-unseen sequences involving the same objects; Experiments 1 and 2) there is no perceptual constraint that influences the process. On the contrary, when the task is more challenging (e.g., differentiating test sequences that have been both experienced during familiarization but with a different probability; Experiments 3 and 4), the process might be affected by the perceptual nature of the stimuli and, more precisely, by the coherence between the type of stimulus (sounds vs. visual objects) and its modality of presentation (sequential vs. spatial).

In order to test this possibility, future studies will be devoted to implement the same task according to a spatiotemporal modality of presentation, during which items appear one at a time in different positions within the screen, resulting in a full-visible triplet. In this way, the chicks have the chance to see the entire sequence and may be facilitated in learning its underlying structure.

Chapter 3 provides the first evidence of generalization of visual regularities in a newborn animal model. Patterns were ABA, AAB, ABB and BAA presented as static triplets of simultaneously-visible elements (e.g., circle-cross-circle). The core aspect of these studies is the generalization test, during which chicks are required to identify the trained pattern when implemented by completely novel elements (e.g., arrow-hourglass-arrow). The animals were able to generalize beyond perceptual features, and extract the structures underlying test triplets.

Moreover, chicks seem not to privilege regularities characterized by certain properties. All comparisons tested in Experiment 6 possessed asymmetrical structures conferred by the presence of an adjacent reduplicated item. One contrast involved two patterns characterized by a repetition located at the beginning (AAB) and at the final edge (ABB) of the triplets as opposed to the other contrast, in which the *same* item was repeated at the beginning and at the end of the triplets (AAB vs. BAA). The animals generalized in both conditions, proving able to detect the reinforced patterns independently of the presence/ absence of adjacent repetitions. In a comparative perspective, this evidence is particularly interesting since human data revealed that adjacent repetitions do enhance generalization.

Neonates in fact, prefer repetition-based regularities, being able to distinguish strings of syllables displayed as ABB but not ABA compared with random (ABC) stimuli (Gervain et al. 2008). This finding has been interpreted as the result of a repetition-detection mechanism that might facilitate subsequent speech perception abilities. Gervain and collaborators (2008) pointed out that neonates are consistently bombarded by linguistic information, and that this input, in particular the infant-directed speech, is marked by contiguous repetitions of words and sentences (Ferguson, 1983). Furthermore, infants' first attempts to speech production are characterized by repeated elements (babbling and first words). According to this view, the neonate brain appears to be tuned to the structure of the sensory information, to guarantee adequate processing and representation.

The findings illustrated in this research may be interpreted within a similar framework. Chicks succeeded in all generalization tests, being able to differentiate patterns with or without repetitions (ABA, AAB, ABB, BAA). This evidence points to the nature of the sensory world: ecologically speaking, there might be limited need to master reduplicated elements in a visual scene for chicks as well as for humans because it seems unlikely that the visual environment comprises such a great number of identical objects adjacently repeated. The chick brain could be predisposed to detect the structure of the visual input not necessarily consisting of repeated objects, leading the animals to equally process patterns with non-adjacent and adjacent repetitions.

Future investigations will be devoted to explore whether this process is constrained by other perceptual aspects of the stimuli. As mentioned above, modality-specificity represents one of the aspects that broadly influences learning and generalization of regularities. In line with the coherence between stimulus and modality of presentation, it is plausible that the infants tested in previous studies had failed to generalize some visual patterns when implemented by temporal sequences (Johnson et al. 2009), and succeeded when similar patterns were displayed as string of simultaneously-visible objects (Saffran et al. 2007). By presenting items one at a time, it may be possible to observe interesting differences in the generalization capacity exhibited by this model. Is generalization prevented when visual objects are presented as a temporal sequence? Differences in sensitivity to repetitions placed in diverse positions within the sequence may be found in the chick, as shown by infants in the visual domain (see Johnson et al. 2009 for further information about early vs. late repetitions), and by human adults in the linguistic domain (Endress et al. 2005). The use of sequentially-presented visual patterns would provide new insights about the role of perceptual constraints on the learning process, investigating whether some regularities would be easier to learn than others in this species. This would have important implications about cross-species similarities or constraints underlying learning mechanisms.

The relationship between statistical learning and generalization puzzles researchers from all over the world and disparate fields such as cognitive neuroscience, developmental psychology, animal cognition, artificial intelligence, etc. The broad, common purpose is to figure out whether statistical learning and generalization (also known as rule learning) are two separate processes as opposed to a unitary mechanism (see Aslin & Newport, 2012 for review about this issue). The hypothesis involving two distinct mechanisms consider the fact that, while statistical learning depends on the identity of input elements, rule learning exploits perceptually-novel stimuli (see Marcus, 2000; Endress & Bonatti, 2007). On the contrary, some studies have revealed that statistical learning could lead to generalization to novel stimuli under certain circumstances (for example, on the basis of context information; Gerken, 2006; Reeder, Newport, Aslin, 2009; 2010). This suggests that statistical and rule learning might be framed under the same process that produces different learning outputs, some of them are restricted to familiar inputs, some others involve novel stimuli (see also Orban, Fiser, Aslin & Lengyel, 2008). At present, it is quite challenging to conceive a single study clarifying this complex issue. New insights may come from studies on domestic chicks, a model that has revealed statistical learning as well as generalization capacities with similar visual inputs, and that allows the use of paradigms possibly suitable to investigate both processes.

Results of the present research suggest that newborn chicks are predisposed to track visual regularities in their postnatal environment. The use of naive organisms with no prior experience with the visual world enable to explore the origins of several cognitive functions (see Vallortigara, 2012 for review). Despite the very limited previous experience, after exposure to a structured input or a 3-day training session, significant learning and generalization effects have been obtained, pointing to the presence of biological predispositions serving the development of these cognitive abilities. In sum, the findings illustrated here broaden the ecological value of statistical learning and generalization as mechanisms that allow efficient processing of the environment, in order to rapidly detect their components. The availability of a viable model to pursue further behavioral studies on this topic paves the way to novel, and even more insightful investigations on the genetics and neural basis of these phenomena.

REFERENCES

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14*(8), 1067-1074.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170-176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Bateson, P. (1973). Preferences for familiarity and novelty: a model for the simultaneous development of both. *Journal of Theoretical Biology*, *41*(2), 249-259.

Bateson, P. (1974). Length of training, opportunities for comparison, and imprinting in chicks. *Journal of Comparative and Physiological Psychology*, *86*(4), 586.

Bateson, P., & Jaeckel, J. B. (1976). Chicks' preferences for familiar and novel conspicuous objects after different periods of exposure. *Animal Behaviour*, *24*(2), 386-390.

Bateson, P. (1979). Brief exposure to a novel stimulus during imprinting in chicks and its influence on subsequent preferences. *Animal Learning & Behavior*, *7*(2), 259-262.

Bateson, P. (1990). Is imprinting such a special case? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 329*(1253), 125-131.

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: an appreciation and an update. *Trends in Ecology & Evolution*, *28*(12), 712-718.

Beckers, G. J., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport*, *23*(3), 139-145.

Bolhuis, J. J., Johnson, M. H., & Horn, G. (1989). Interacting mechanisms during the formation of filial preferences: The development of a predisposition does not prevent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(4), 376.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Butler, A. B., & Hodos, W. (2005). Comparative vertebrate neuroanatomy. *Hoboken*.

Chen, J., van Rossum, D., & ten Cate, C. (2015). Artificial grammar learning in zebra finches and human adults: XYX versus XXY. *Animal Cognition*, *18*(1), 151-164.

Clara, E., Regolin, L., Zanforlin, M., & Vallortigara, G. (2006). Domestic chicks perceive stereokinetic illusions. *Perception-London*, *35*(7), 983.

Clarke, C. H., & Jones, R. B. (2000). Effects of prior video stimulation on open-field behaviour in domestic chicks. *Applied Animal Behaviour Science*, *66*(1), 107-117.

Clarke, C. H., & Jones, R. B. (2001). Domestic chicks' runway responses to video images of conspecifics. *Applied Animal Behaviour Science*, *70*(4), 285-295.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, *67*(4), 1323-1332.

Comins, J. A., & Gentner, T. Q. (2013). Perceptual categories enable pattern generalization in songbirds. *Cognition*, *128*(2), 113-118.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539-546.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24-39

Daniela, M., & Toro, J. M. (2013). Rule learning over consonants and vowels in a non-human animal. *Cognition*, *126*(2), 307-312.

Daisley, J. N., Vallortigara, G., & Regolin, L. (2010). Logic in an asymmetrical (social) brain: transitive inference in the young domestic chick. *Social Neuroscience*, *5*(3), 309-319.

Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, *111*(3), 378-382.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2-19.

Deng, C., & Rogers, L. J. (1998). Bilaterally projecting neurons in the two visual pathways of chicks. *Brain Research*, *794*(2), 281-290.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, *134*(3), 406-419.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247-299.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, *105*(3), 577-614.

Ferguson, C. A. (1983). Reduplication in child phonology. *Journal of Child Language*, *10*(01), 239-243.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499-504.

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458-467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822-15826.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134(4), 521-537

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125.

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70-74.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486-490.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204-1207.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67-B74.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222-14227.

Ghahramani, Z. (2004). Unsupervised learning. *Advanced Lectures on Machine Learning Lecture Notes in Computer Science,* 72-112.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of ‘sameness’ and ‘difference’ in an insect. *Nature*, *410*(6831), 930-933.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. *Perceptual organization in vision: Behavioral and neural perspectives*, 233-278.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183-206.

Goujon, A., & Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (Papio papio). *Behavioural Brain Research*, *247*, 101-109.

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (Papio papio). *Science*, *336*(6078), 245-248.

Graf-Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254-260.

Güntürkün, O., Miceli, D., & Watanabe, M. (1993). Anatomy of the avian thalamofugal pathway. *1993) Vision, Brain and Behavior in Birds Cambridge: MIT*, 115-135.

Ham, A. D., & Osorio, D. (2007). Colour preferences and colour vision in poultry chicks. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1621), 1941-1948.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64.

Hauser, M. D., & Glynn, D. (2009). Can free-ranging rhesus monkeys (Macaca mulatta) extract artificially created rules comprised of natural vocalizations?. *Journal of Comparative Psychology*, *123*(2), 161-167.

Helmholtz, H. v. (1925). Treatise on psychological optics (JPC Southall, Trans.).(Vol. 3). *Menasha, WI: Banta.(Original work published 1910)*.

Hinton, G. E., & Sejnowski, T. J. (1999). *Unsupervised learning: foundations of neural computation*: MIT press.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., ... & Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151-159.

Johnson, M.H., Bohluis, J.J. & Horn, G. (1985). Interaction between acquired preferences and developing predispositions during imprinting. *Animal Behaviour*, 33*(3)*, 1000-1006.

Johnson, M., & Horn, G. (1988). Development of filial preferences in dark-reared chicks. *Animal Behaviour, 36*(3), 675-683.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8-and 11-month-olds. *Infancy*, *14*(1), 2-18.

Jones, C. D., Osorio, D., & Baddeley, R. J. (2001). Colour categorization by domestic chicks. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*(1481), 2077-2084.

Jones, R. B., Larkins, C., & Hughes, B. O. (1996). Approach/avoidance responses of domestic chicks to familiar and unfamiliar video images of biologically neutral stimuli. *Applied Animal Behaviour Science*, *48*(1), 81-98.

Karten, H. J., & Shimizu, T. (1989). The origins of neocortex: connections and lamination as distinct events in evolution. *Journal of Cognitive Neuroscience 1*(4), 291-301.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559-1571.

Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: flexible yet constrained. *Frontiers in psychology, 3*.

Lazareva, O. F., Freiburger, K. L., & Wasserman, E. A. (2004). Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychonomic Bulletin & Review*, *11*(6), 1111-1117.

Lea, S. E., Slater, A. M., & Ryan, C. M. (1996). Perception of object unity in chicks: A comparison with the human infant. *Infant Behavior and Development*, *19*(4), 501-504.

Lorenz K.Z. (1937). The companion in the bird's world. *The Auk, 54*(3), 245-273.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

Marcus, G. F. (2000). Pabiku and Ga Ti Ga Two Mechanisms Infants Use to Learn About the World. *Current Directions in Psychological Science*, *9*(5), 145-147.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*(5), 387-391.

Mayer, U., Watanabe, S., & Bischof, H. J. (2013). Spatial memory and the avian hippocampus: Research in zebra finches. *Journal of Physiology-Paris*, *107*(1), 2-12.

Mascalzoni, E., Regolin, L., & Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proceedings of the National Academy of Sciences, 107*(9), 4483-4485.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*(2), 91-121.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences, 108*(48), 19401-19406.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology*, *63*(2) 176-185.

Misyak, J. B., & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. In *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1307-1312).

Murphy, R. A., Mondragón, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, *319*(5871), 1849-1851.

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

Passantino, A. (2008). Application of the 3Rs principles for animals used for experiments at the beginning of the 21st century. *Annual Review of Biomedical Sciences*, *10*, T27-T32.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244-247.

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2009). The role of distributional information in linguistic category formation. In *Proceedings of the 31st annual meeting of the Cognitive Science Society* (pp. 2564-2569).

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2010). Novel words in novel contexts: The role of distributional information in form-class category learning. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2063-2068).

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, *57*(7), 971-976.

Regolin, L., Tommasi, L., & Vallortigara, G. (2000). Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Animal Cognition*, *3*(1), 53-60.

Regolin, L., Marconato, F., & Vallortigara, G. (2004). Hemispheric differences in the recognition of partly occluded objects by newly hatched domestic chicks (Gallus gallus). *Animal Cognition*, *7*(3), 162-170

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: Current research and theory*.

Rosa Salva, O., Daisley, J. N., Regolin, L., & Vallortigara, G. (2010). Time-dependent lateralization of social learning in the domestic chick (Gallus gallus domesticus): effects of retention delays in the observed lateralization pattern. *Behavioural Brain Research*, *212*(2), 152-158.

Rosa Salva, O., Regolin, L., & Vallortigara, G. (2010). Faces are special for newly hatched chicks: evidence for inborn domain−specific mechanisms underlying spontaneous preferences for face−like stimuli. *Developmental Science*, *13*(4), 565-577.

Rosa Salva, O., Regolin, L., Mascalzoni, E., & Vallortigara, G. (2012). Cerebral and behavioural asymmetries in animal social recognition. *Comparative Cognition & Behavior Reviews CCBR, 7*(10), 110-138.

Rosa Salva, O., Rugani, R., Cavazzana, A., Regolin, L., & Vallortigara, G. (2013). Perception of the Ebbinghaus illusion in four-day-old domestic chicks (Gallus gallus). *Animal Cognition*, *16*(6), 895-906.

Rugani, R., Regolin, L., & Vallortigara, G. (2007). Rudimental numerical competence in 5-day-old domestic chicks (Gallus gallus): identification of ordinal position. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(1), 21.

Rugani, R., Kelly, D. M., Szelest, I., Regolin, L., & Vallortigara, G. (2010). Is it only humans that count from left to right?. *Biology Letters*, *6*(3), 290-292.

Rugani, R., Vallortigara, G., Vallini, B., & Regolin, L. (2011). Asymmetrical number-space mapping in the avian brain. *Neurobiology of Learning and Memory*, *95*(3), 231-238.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language,* 44(4), 493-515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172-196.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669-680.

Saffran, J. R., & Thiessen, E. D. (2007). Domain−General Learning Capacities. *Blackwell handbook of language development*, 68-86.

Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition, 107*(2), 479-500.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105-134.

Santolin, C., Rosa Salva, O., Vallortigara, G. & Regolin, L. Unsupervised visual statistical learning in the newborn chick (*Gallus gallus*). *Manuscript under review*

Santolin, C., Rosa Salva, O., Regolin, L. & Vallortigara, G. Generalization of visual regularities in newly-hatched chicks (*Gallus gallus*). *Manuscript* s*ubmitted*

Schmid, K. L., & Wildsoet, C. F. (1998). Assessment of visual acuity and contrast sensitivity in the chick using an optokinetic nystagmus paradigm. *Vision Research*, *38*(17), 2629-2634.

Shimizu, T., & Karten, J. H. (1993). The avian visual system and the evolution of the neocortex (H. P. Zeigler & H. J. Bischof, Eds.). In *Vision, Brain and Behaviour in Birds* (pp. 103-114). MA: MIT Press.

Soto, F. A., & Wasserman, E. A. (2014). Mechanisms of object recognition: what we have learned from pigeons. *Frontiers in Neural Circuits*, *8*.

Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and prosodic cues for song segmentation learning by Bengalese finches (Lonchura striata var. domestica). *Ethology, 116*(6), 481-489.

ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1984-1994.

Thiessen, E. D. (2012). Effects of inter-and intra-modal redundancy on infants' rule learning. *Language Learning and Development*, *8*(3), 197-214.

Tinbergen, N. (1963) On aims and methods of ethology. Zeitschrift fü r Tierpsychologie, 20, 410–433.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*(5), 867-875.

Vallortigara, G., & Andrew, R. (1991). Lateralization of response by chicks to change in a model partner. *Animal Behaviour, 41*(2), 187-194.

Vallortigara, G., & Andrew, R. J. (1994). Differential involvement of right and left hemisphere in individual recognition in the domestic chick. *Behavioural Processes*, *33*(1), 41-57.

Vallortigara, G. (2012). Core knowledge of object, number, and geometry: a comparative and neural approach. *Cognitive Neuropsychology, 29*(1-2), 213-236.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270-276.

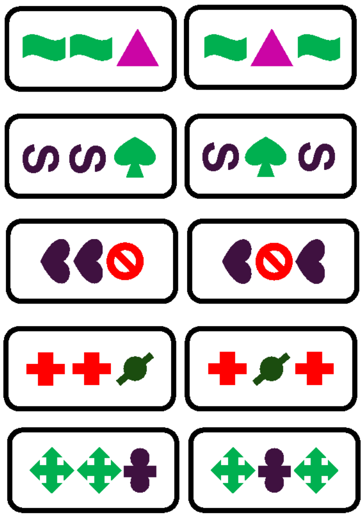
Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews CCBR*, *3*, 13-45.

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K., et al. (2013). Auditory artificial grammar learning in macaque and marmoset monkeys. *The Journal of Neuroscience, 33*(48), 18825-18835.

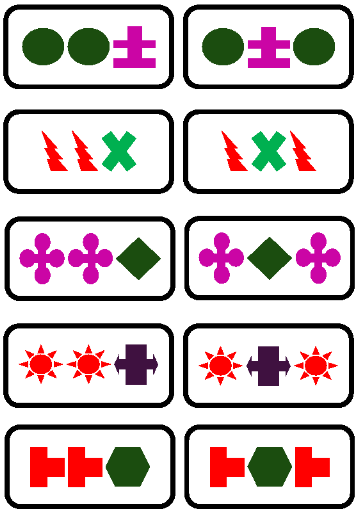
APPENDIX

Experiment 5 - AAB vs. ABA

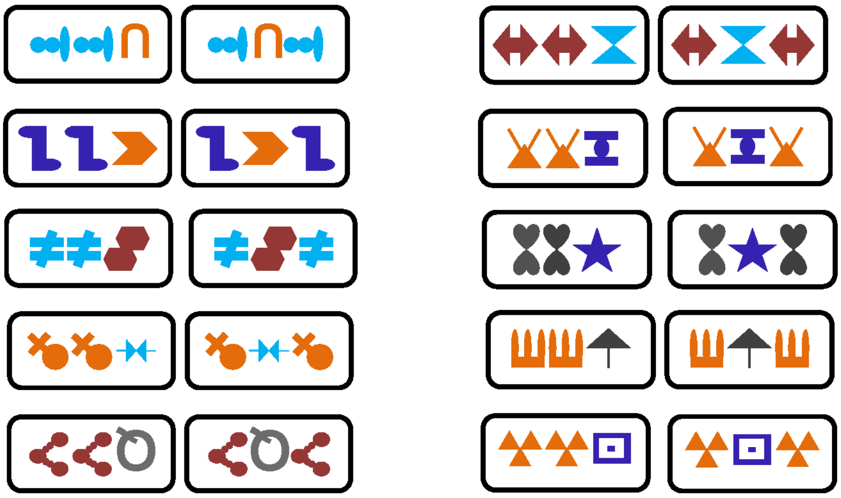
Training sequences 1



Training sequences 2

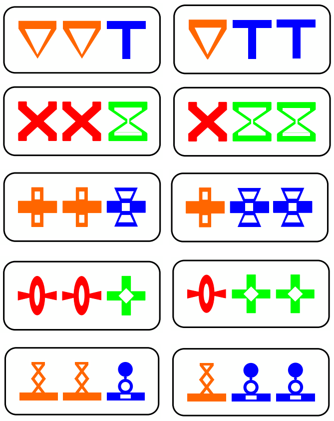


Test sequences



Experiment 6 - Condition 1 (AAB vs. ABB)

Training sequences



Test sequences