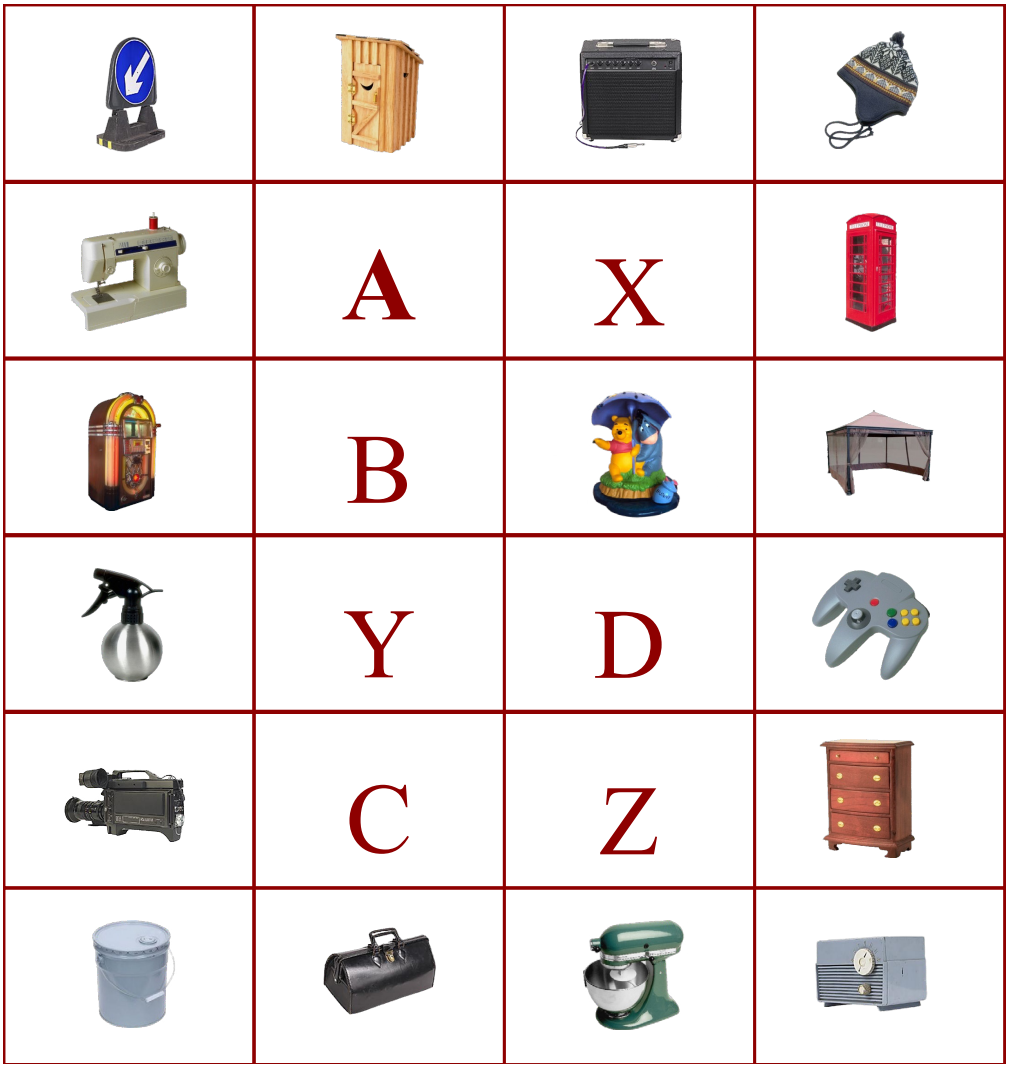
Investigating the Learning Mechanisms of Visual Statistical Learning



Ilayda Nazli

DONDERS SERIES

RADBOUD UNIVERSITY PRESS

Radboud Dissertation Series

Investigating the Learning Mechanisms of Visual Statistical Learning

Ilayda Nazli

The work described here was carried out at the Donders Institute for Brain, Cognition, and Behaviour, Radboud University. The work was funded by the Ministry of National Education of the Republic of Türkiye to Ilayda Nazli, as well as grants awarded to Prof. Floris de Lange by the European Union (ERC Consolidator Grant 101000942, "Surprise").

Author: Ilayda Nazli
Title: Investigating the Learning Mechanisms of Visual Statistical Learning

Radboud Dissertations Series
ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: Proefschrift AIO | Guus Gijben
Cover: Ilayda Nazli
Printing: DPN Rikken/Pumbo

ISBN: 9789465150321
DOI: 10.54195/9789465150321
Free download at: www.boekenbestellen.nl/radboud-university-press/dissertations

RADBOUD
UNIVERSITY
PRESS

Investigating the Learning Mechanisms of Visual Statistical Learning

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 21 januari 2025
om 10.30 uur precies

door

Ilayda Nazli
geboren op 23 april 1994
te Malatya, Türkiye

Promotor
Prof. dr. F.P. de Lange

Copromotor
Dr. A. Ferrari

Manuscriptcommissie
Prof. dr. R. Cools
Prof. dr. H. Boyaci (Bilkent Üniversitesi, Türkiye)
Prof. dr. M.V. Peelen

Investigating the Learning Mechanisms of Visual Statistical Learning

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defended in public on

Tuesday, January 21, 2025
at 10.30 am

by

Ilayda Nazli
born on April 23, 1994
in Malatya, Türkiye

Supervisor
Prof. dr. F.P. de Lange

Co-supervisor
Dr. A. Ferrari

Doctoral Thesis Committee
Prof. dr. R. Cools
Prof. dr. H. Boyaci (Bilkent Üniversitesi, Türkiye)
Prof. dr. M.V. Peelen

Table of contents

Chapter 1	Introduction	9
Chapter 2	Forward and backward blocking in statistical learning	21
Chapter 3	What type of associations modulates statistical learning?	59
Chapter 4	Does the uniqueness of visual associations modulate visual activity?	77
Chapter 5	Discussion	92
	References	101
	Nederlandse samenvatting	108
	Acknowledgements	112
	About the author	114
	Research data management	116
	Donders Graduate School for Cognitive Neuroscience	118

Chapter 1

Introduction

Learning is a fundamental aspect of our life. It enables us to develop and enhance our internal representations of the world. A crucial element of learning is the ability to form associations between events that are systematically related across space or time (Gershman, 2017). Our environment is full of such regularities. Therefore, it is essential for us to form associations between repetitive structures to predict the upcoming input, to prepare adequate responses and to adapt to the environment flexibly. For example, books are arranged based on a regular classification system in libraries, enabling to find them easily (see Figure 1.1a), and the color of traffic lights follow a regular sequence to guide drivers and pedestrians to pass the road safely (see Figure 1.1b). Observers can automatically extract these regularities from the environment over multiple exposures, even without the intention or effort to learn and often without being aware of the learning process. This form of learning is known as statistical learning (Batterink et al., 2019; Frost et al., 2019; Saffran et al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). Statistical learning shapes the information processing of observers. In classical spatial and temporal statistical learning paradigms, participants are exposed to a stream of stimuli and later asked to discriminate structured and expected from random and unexpected shape stimuli sets (see Figure 1.1c-d). Statistical learning mostly results in facilitated behavioral responses such as faster and more accurate responses to structured and expected relative to unexpected stimuli (Fiser & Lengyel, 2019, 2022; Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et al., 2005). The neural consequences of statistical learning frequently involve suppressed neural responses for stimuli that are expected given the previous context (He et al., 2022; Richter et al., 2018; Richter & de Lange, 2019).

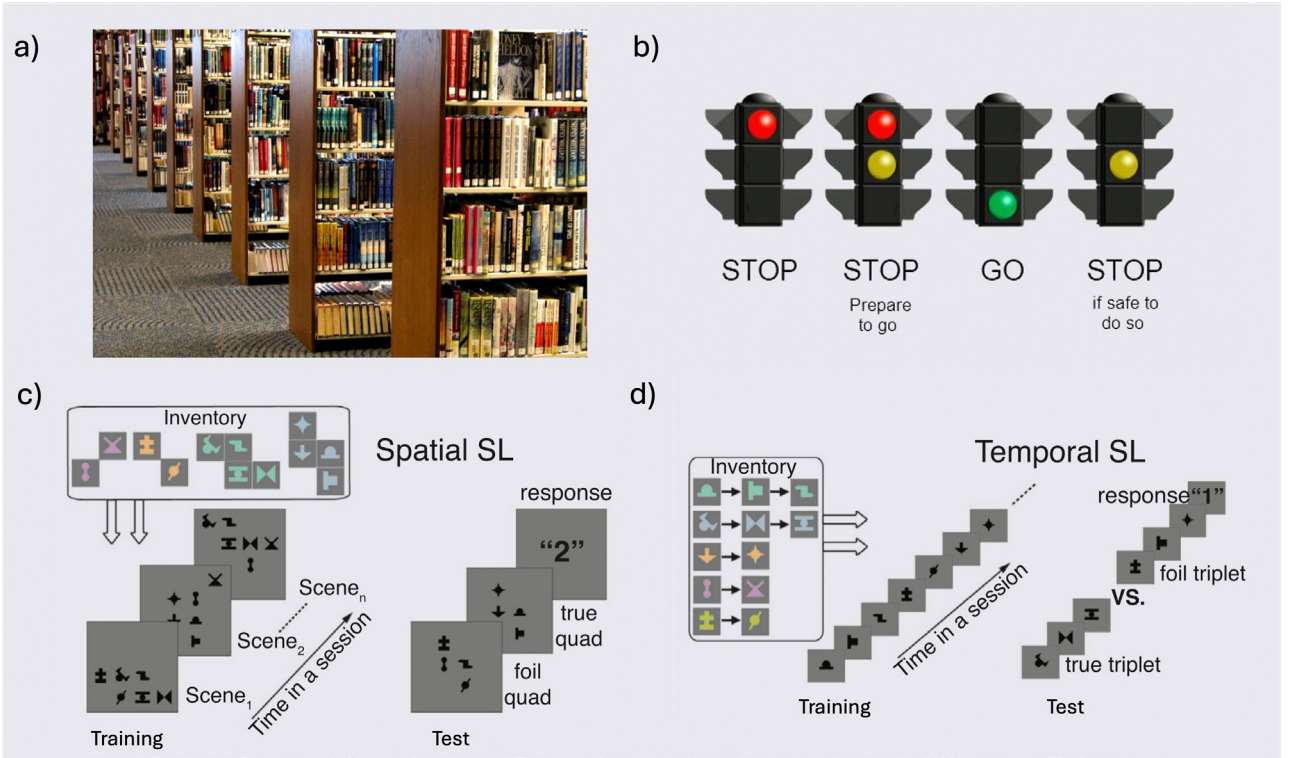


Figure 1.1. (**a**) Library Bookshelves: Libraries use a specific spatial organization system to arrange books on the shelves. (**b**) Traffic Lights: Traffic lights guide drivers and pedestrians on roads by following a specific temporal order of red, yellow, and green lights. (**c**) Spatial Statistical Learning Paradigm: In this paradigm, a stream of shape combinations is used for training. Both true and foil shape combinations are then used in a two-alternative forced-choice familiarity test (Fiser & Lengyel, 2019). (**d**) Temporal Statistical Learning Paradigm: In this paradigm, a stream of shape triplets is used as a training sequence. Subsequently, both true and foil shape triplets are presented consecutively as test stimuli in a two-alternative forced-choice familiarity test (Fiser & Lengyel, 2019).

What learning mechanism underlies statistical learning?

At the core of learning is the formation of associations between events. Once the separate events are connected, our brain can make inferences about an upcoming output given an input. The question is how our brain connects separate events and forms associations between them. This can be achieved by simply being exposed to relevant events. Simple contiguity-dependent Hebbian associative learning suggests that learning is the strengthening of associations following the co-occurrence of the relevant events systematically across time (Hebb, 1949). For example, simple Pavlovian learning or classical conditioning can be explained by Hebbian associative learning (Agliari et al., 2023; Pavlov, 1927). A dog automatically salivates in response to food. During conditioning, the sound of a bell is repeatedly paired with the food. Once learning is accomplished, the bell itself elicits salivation.

Although Hebbian associative learning can explain simple classical conditioning, it fails to explain more complex situations. Turning back to the example of Pavlov's dog, let's imagine that the food remains paired with the same bell but now also with a light. In this case, the light may fail to elicit salivation even if it is repeatedly paired with the food whereas the bell still can. This situation shows that learning does not only depends on the mere observation of input and output together across time. Instead, learning is moderated by the predictive power of an input over an output (Boddez et al., 2014; De Houwer et al., 2005; Luque et al., 2018; Schmidt & De Houwer, 2019). Recently it has been put forward that a primary function of the brain is to predict future states of the environment (Clark, 2013).The environment is continuously changing, which may lead to a mismatch between our prior expectations and observed reality. This discrepancy between the expected and the observed outcome is known as prediction error and it needs to be minimized to adapt to the changes in the environment and to interact with it (Clark, 2013; Friston, 2005). In this example above where the food is paired later with the light, the food is not unexpected given the bell, and thus the relationship between the food and the light may not be learned.

Many models of learning suggest that the occurrence of learning relies on prediction error. These models posit that the changes in associative strength between input and output are determined by the amount of discrepancy between the expected and the observed outcome, a.k.a. the prediction error, and the associative strength is only updated when the observed outcome is unexpected (Pearce & Hall, 1980; Rescorla & Wagner, 1972). Despite the necessity of prediction

error in learning, its function differs across different models (Roesch et al., 2012). In the Rescorla – Wagner model, the associative strength between events is directly modulated by prediction errors and it is primarily driven by the output. When the error is large, the change in the associative strength is also large, and when the error is zero, the associative strength is not updated. On the other hand, the Pearce – Hall model extends the prediction error account of learning to selective attention. The magnitude of the attention paid to the input is modulated by prediction errors, which in turn determines the associative strength. When the error is large, the attention devoted to the input is increased, which strengthens the association between events.

A key concept in these learning models is that prediction error modulates learning the relationship between reward-predictive cue and rewarding outcome, i.e. reinforcement learning. As opposed to statistical learning explained above, in reinforcement learning, observers typically learn the regularities intentionally and are aware of what has been learned. A deviation from the explicitly expected outcome, such as unexpected (omission of) reward or feedback, serves as an explicit reward prediction error in reinforcement learning (Gershman & Daw, 2017). Here, the central question is whether removing rewarding outcome from the association completely changes the dynamics. Despite the absence of explicit reward in statistical learning, the sensory prediction error between the observed (non-rewarding/punishing) and the expected outcome is strong enough for observers to update the association strength between events. Also, observers are intrinsically motivated for information search to build an accurate internal model of world, which guides learning (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018). This intrinsic motivation for information gain might be seen as an implicit reward prediction error involved in statistical learning. Therefore, considering both sensory prediction error and implicit-reward prediction error, we may ask more broadly: Is statistical learning driven by prediction error?

Several studies suggest that statistical learning may indeed similarly rely on prediction errors. It is well known that striatal dopaminergic neurons respond to reward prediction errors (Corlett et al., 2004; McClure et al., 2003; O'Doherty et al., 2004; Schultz et al., 1997). Interestingly, it is also found that dopaminergic activity in the ventral tegmental area of rats is important for the formation of an association between two non-rewarding stimuli (Keiflin et al., 2019; Sharpe et al., 2017). Similarly, in humans, statistical learning of stimulus-stimulus associations involves the striatum (den Ouden et al., 2009; Klein-Flügge et al., 2019). Despite the neural evidence, behavioral evidence is controversial. Cue competition is a crucial

category of phenomena in associative learning and generally taken as evidence that reinforcement learning is error-driven (Boddez et al., 2014). The most famous example of cue competition is Kamin or forward blocking (Kamin, 1969). In a typical blocking paradigm, first the cue A is paired with the outcome X (i.e., A→X). Later a new cue B is presented together with the cue A, and they are followed by the same outcome X (i.e., AB→X). As a result of blocking, the previously learned A→X association prevents the formation of an association between the second cue B and the outcome X. According to Rescorla – Wagner model, this is because the cue A already minimizes the prediction error during the exposure to the A→X. Few studies using variants of blocking did not find clear evidence for error-driven statistical learning. Beeslay and Shanks (2012) did not observe any blocking effect in a contextual cueing experiment. Notably, the learned associations in their study were based on the spatial relationship among distracters and targets in a visual search task. However, blocking typically involves a temporal prediction between a cue and a future outcome (Aggarwal et al., 2020; Aggarwal & Wickens, 2020; Blanco et al., 2014; De Houwer et al., 2005; De Houwer & Beckers, 2003; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Luque et al., 2018; Mitchell et al., 2006; Steinberg et al., 2013; Vandorpe et al., 2005). Similarly, Schmidt and de Houwer (2019) observed blocking in a series of color-word contingency learning studies only when participants were explicitly instructed to learn. Critically, they presented specific color-word associations more frequently, hence it is not clear whether learning in the blocked condition emerged because of cue predictability or, instead, because of mere increased familiarity with the more frequent associations. Importantly, predictability should determine associative learning according to error-driven accounts (Kamin, 1969, Rescorla & Wagner, 1972) rather than mere familiarity or co-occurrence (Hebb, 1949). Last, Moris et.al., (2014) found blocking effect in a repetition priming task. Critically, participants were explicitly informed about the presence of regularities among the stimuli and set out to learn them intentionally and explicitly. Such learning conditions substantially deviate from a typical statistical learning scenario, where observers automatically extract regularities without intention nor awareness (Batterink et.al., 2019; Frost et.al., 2019; Sherman et.al., 2020; Turk-Browne et.al., 2010). On the other hand, there is evidence for blocking in children (Griffiths et al., 2011; McCormack et al., 2009, 2013; Sobel et al., 2004) and in 8-month-old infants (Sobel & Kirkham, 2006, 2007) who clearly did not follow any explicit task instructions. This suggests that cue competition may be observable after statistical learning. I will explore the question of error-driven statistical learning in **Chapter 2** by borrowing the famous blocking paradigms of reinforcement learning to see whether and how they generalize to statistical learning.

What type of relationship is learned during statistical learning?

Statistical learning is often defined as the automatic extraction of statistical regularities from the environment (Batterink et al., 2019; Frost et al., 2019; Saffran et al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). This definition brings the main question to the forefront: What types of statistical regularities are extracted, and which metrics govern their extraction? Two metrics that have been extensively examined in the prior statistical learning research are joint probability and conditional probability. The joint probability is the total number of occurrences of stimulus pairs among other stimulus pairs. The conditional probability is the probability of an event occurring in a specific condition, reflecting how strongly a stimulus occurs given another stimulus. The prominent statistical learning study of Fiser and Aslin (2002) found that observers learned the relationship between events based on the conditional probability rather than the joint probability. Tracking conditional probability of an object given another object allows observers to make predictions about the future. Until now, conditional probability is considered as the main metric to determine statistical learning: observers better track strong relationships with high conditional probability between events.

The literature on intentional learning demonstrates that observers not only track the strength with which a stimulus follows another, but also track whether a stimulus uniquely predicts the other. Suppose that in most instances where A occurs, X follows A (i.e., A→X). This results in a high conditional probability of X given A, indicating a strong relationship between A and X. Consequently, we would expect observers to strongly learn the A→X association. However, if X also frequently appears without A, such as following a different stimulus B, then X is not uniquely predicted by A. Therefore, the A→X association may not be learned as strongly by observers. This is captured by a metric called ΔP (Allan & Jenkins, 1980). According to ΔP, learning is based not only on how often X follows A but also on how often X appears in the absence of A (i.e., $\Delta P=P(X|A)-P(X|\sim A)$). Therefore, in the example above where both A and B are predictive of X, ΔP of X given A is low, potentially leading to a weak A→X association. To the best of our knowledge, there is only one study testing if statistical learning is sensitive to ΔP rather than the conditional probability. Using a classical visual statistical learning task, Leshinskaya and Thompson-Schill (2021) found that participants failed to learn the relationship between events that had high conditional probabilities when the ΔP between them was low. This implies that statistical learning may be governed by unique predictive relationships rather than strong relationships, contrary to assumptions made in prior work.

The literature on uniqueness suggest that there are two types of unique relationships: *ΔP* (Allan & Jenkins, 1980) and Dual Factor Heuristic (*DFH*, Hattori & Oaksford, 2007). They describe the relationship between two events based on a 2×2 matrix as shown in Figure 1.2. This table summarizes the association between A and X: A is followed by X. A and -A respectively represent the occurrence and non-occurrence of leading stimulus A; X and -X respectively represent the occurrence and non-occurrence of trailing stimulus X. The letters in the cells (i.e., a, b, c, d) represent the relative frequencies of the presence and absence of A and X: a cell shows the number of 'A is followed by X (A→X)' observations, b cell shows the number of 'A is followed by a different trailing stimulus (A→Y)' observations, c cell shows the number of 'X follows a different leading stimulus (B→X)' observations and d cell shows the occurrence of neither A nor X (B→Y). Observers using the ΔP strategy to form associations focus equally on both the occurrence and non-occurrence of events and systematically and rationally process all four cells (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005). However, it has been stated that observers focus on these four cells differentially (Béghin et al., 2021; Matute et al., 2015) , primarily concentrating on the occurrence of events while disregarding the d cell (Béghin et al., 2021; I. Hattori et al., 2017). To capture this, Hattori and Oaksford (2007) proposed a different index called the Dual Factor Heuristic ($DFH = \sqrt{P(X|A) \times P(A|X)}$. Unlike the rational and analytic ΔP, observers using the DFH tend to focus on the occurrence of events, disregard the d cell, and process the relative frequencies of the presence and absence of A and X rapidly and with low effort (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005).

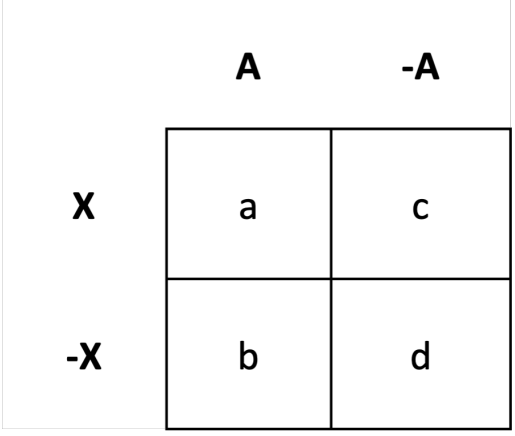


Figure 1.2. A matrix representing the relationship between event A and event X. A and -A respectively represent the occurrence and non-occurrence of leading stimulus A; X and -X respectively represent the occurrence and non-occurrence of trailing stimulus X.

The formulas of ΔP and DFH mostly share the same components. Therefore, strengthening or weakening one of them makes the other parameter weaker or stronger. The study of Leshinskaya and Thompson-Schill (2021) suggests that not strong but unique predictive relations govern statistical learning by means of . Unfortunately, we cannot compute DFH using the limited information related to the relationship between events provided in their paper. On the other hand, we can compute DFH values using the design matrix of Ramachandran et.al. (2016). In their study, two macaque monkeys engaged in passive viewing of pairs of images in which a leading image was followed by a trailing image based on a certain conditional probability (see Figure 1.3a-b). In the 1:1 conditional probability condition, the leading image was perfectly predictive of one single trailing image (i.e., $P(Trailing|Leading)$=1 and $P(Leading|Trailing)$=1). In the 2:1 conditional probability condition, the leading image was perfectly predictive of the trailing image (i.e., $P(Trailing|Leading)$=1), yet the trailing image was also predicted by a different leading image (i.e., $P(Leading|Trailing)$=0.5). And, in the 1:2 conditional probability condition, the leading image was equally predictive of two different trailing images (i.e., $(Trailing|Leading)$=0.5 and $P(Leading|Trailing)$=1). They found strong expectation suppression in monkey inferotemporal cortex for 1:1 condition compared to 1:2 and 2:1 conditions, but these two latter were not differrent from each other. Richter et.al. (2018) replicated the study of Ramachandran et.al. (2016) with human participants using the similar experimental paradigm and the same design matrix. In their study, participants were exposed to pairs of object images in a statistical learning paradigm, in which the first object predicted the identity of the second object (see Figure 1.3c). Similarly, they found that, there was no behavioral or neural difference between 1:2 and 2:1 conditions. The findings of these studies cannot be explained by conditional probability because in the 2:1 condition, the leading image was strong predictive of the trailing image with a conditional probability of 1. On the other hand, these results can be captured by uniqueness. In the 1:2 condition, ΔP is 0.5 and DFH is 0.7 whereas in the 2:1 condition, ΔP is 0.9 and DFH is 0.7. Although the value of ΔP is different in two conditions, the value of DFH is the same, implying that statistical learning may be governed by DFH. Thus, these studies suggest that statistical learning is more sensitive to uniqueness rather than conditional probability: however, it is not clear which forms of uniqueness determines statistical learning. This question will be explored in **chapter 3** and **chapter 4**.

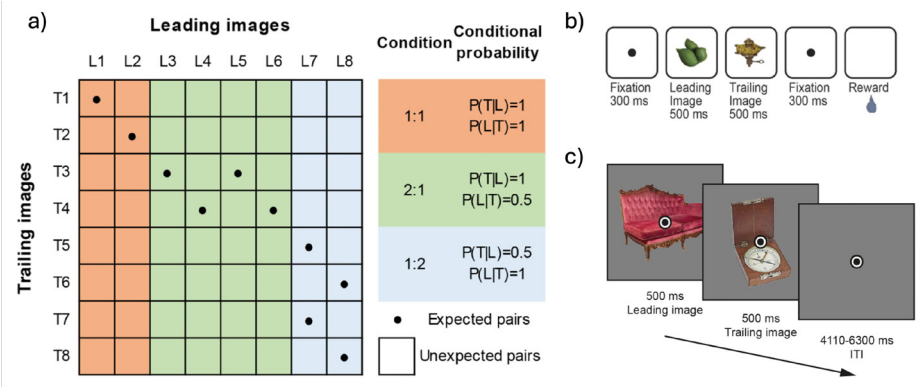


Figure 1.3. (**a**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase in the Ramachandran et.al. (2016) and Richter et.al. (2018). Ls represent leading stimuli, and Ts represent trailing stimuli. (**b**) The timing of single trial during data training and data collection in the study of Ramachandran et.al. (2016). (**c**) The timing of single trial during data training and data collection in the study of Richter et.al. (2018).

Overview of this thesis

In sum, at the core of this thesis is the question of how and what type of statistical regularities are extracted. **Chapter 2** aims to assess whether statistical learning is based on error-driven learning. For this, we borrowed the famous forward and backward blocking paradigms of reinforcement learning. **Chapter 3 and Chapter 4** are devoted to exploring if statistical learning is more sensitive to uniqueness rather than conditional probability and which forms of uniqueness (i.e., ΔP or DFH governs statistical learning using online studies and fMRI. **Chapter 5** summarizes and integrates the results presented in chapters 2-4. Importantly, I will highlight the core conclusions we can draw from the work presented in this thesis and the literature.

1

Chapter 2

Forward and backward blocking in statistical learning

Abstract

Prediction errors have a prominent role in many forms of learning. For example, in reinforcement learning, agents learn by updating the association between states and outcomes as a function of the prediction error elicited by the event. One paradigm often used to study error-driven learning is blocking. In forward blocking, participants are first presented with stimulus A, followed by outcome X (A→X). In the second phase, A and B are presented together, followed by X (AB→X). Here, A→X blocks the formation of B→X, given that X is already fully predicted by A. In backward blocking, the order of phases is reversed. Here, the association between B and X that is formed during the first learning phase of AB→X is weakened when participants learn exclusively A→X in the second phase. The present study asked the question whether forward and backward blocking occur during visual statistical learning, i.e., the incidental learning of the statistical structure of the environment. In a series of studies, using both forward and backward blocking, we observed statistical learning of temporal associations among pairs of images. While we found no forward blocking, we observed backward blocking, thereby suggesting a retrospective revaluation process in statistical learning and supporting a functional similarity between statistical learning and reinforcement learning.

This chapter has been published as:

Nazlı, İ., Ferrari, A., Huber-Huber, C., & De Lange, F. P. (2024). Forward and backward blocking in statistical learning. *PloS one*, *19*(8), e0306797.

Introduction

Learning is an essential feat of animal cognition. It allows us to build and refine our internal models of the world, so that we predict and flexibly adapt to our dynamic environment. A key feature of learning is the ability to form associations between events that take place in a systematic relationship across space or time (Gershman, 2017). For example, in a typical classical conditioning experiment (Pavlov, 1927), a dog automatically salivates (i.e., unconditioned response) in response to food (i.e., outcome or unconditioned stimulus). During conditioning, the sound of a bell (i.e., cue or conditioned stimulus) is repeatedly paired with the food. Once conditioning is accomplished, the bell itself elicits salivation (i.e., conditioned response).

Cue competition is a crucial category of phenomena in associative learning. It refers to the observation that learning which cues predict an outcome not only depends on the presence of the cues before the outcome. Rather, cues compete with each other to gain predictive power over the outcome, and this moderates the learning process (Boddez et al., 2014; De Houwer et al., 2005; Luque et al., 2018; Schmidt & De Houwer, 2019).

One key example of cue competition is Kamin blocking, also known as forward blocking (Kamin, 1969). In a typical forward blocking paradigm (see Table 2.1), observers first learn the association between cue A and outcome X (A→X), and later they are trained with the association between cues A + B and outcome X (AB→X). As a result of forward blocking, observers learn the association between cue B and outcome X less strongly, because X is already completely predicted by cue A. In other words, the previously learned A-X association blocks learning the association between cue B and outcome X. Forward blocking cannot be explained by simple contiguity-dependent Hebbian associative learning (Hebb, 1949). Thereby, it suggests that the simple temporal co-occurrence of different stimuli is not sufficient for learning to occur. Instead, the model developed by Rescorla and Wagner (1972) provides an explanation for blocking (though see Spicer et al., 2021 for a modification of the traditional model). According to the Rescorla-Wagner model, changes in associative strength are determined by the amount of discrepancy between the expected and the observed outcome, i.e. the prediction error. In the forward blocking procedure, the previously learned A→X association prevents the formation of an associative link between the second cue B and the outcome X, because the cue A already minimizes the prediction error during the exposure to the A→X pairs in the first training phase.

Table 2.1. General experimental design. Letters denote conditions (i.e., A for Antedating, B for Blocked, C-D for Control).

	Training phase 1	***Training phase 2***	***Test phase***
Forward	A → X	AB → X	A → X
blocking		CD → Y	B → X D → Y
Backward blocking	AB → X CD → Y	A → X	A → X B → X D → Y

A similar, but distinct form of cue competition is backward blocking, which is an example of retrospective revaluation: a change in the associative strength occurs because the association between the companion cue (i.e., the cue that is previously associated with the target cue and outcome) and the outcome is revaluated. In the backward blocking paradigm (Shanks, 1985), observers are first trained with AB→X association, and subsequently with A→X association. In spite of the reversed order of training phases compared to forward blocking, backward blocking leads to a similar outcome as forward blocking: a lack of association between blocked cue B and outcome X. Here, in the first training phase, both A-X and B-X associations are formed equally (i.e., depending on the saliency of cues). However, in the second training phase, as observers are trained with A→X association, the associative strength between cue A and outcome X becomes stronger, which in turn weakens the association between cue B and outcome X. While this form of retrospective revaluation cannot be explained by the traditional Rescorla – Wagner model, as this model assumes that the relevant cue must be present in order to change the associative strength (Kruschke, 2008; Miller & Witnauer, 2016; Rescorla & Wagner, 1972), backward blocking can be successfully modeled by a slightly revised version of the traditional model. For example, backward blocking can be explained by a Rescorla-Wagner learning model that assigns non-zero salience to non-presented blocked stimuli whose memories or representations are retrieved by competing stimuli that had previously been paired with those blocked stimuli (Van Hamme & Wasserman, 1994) or by a Bayesian generalization of the Rescorla – Wagner model, the Kalman filter (Gershman, 2015; Kalman, 1960; Kruschke, 2008), where the weights of all possible cues are updated simultaneously, and the sum of all possible weights equals to 1.

In typical blocking experiments, associations are learned either when the outcome is a reward (Aggarwal et al., 2020; Aggarwal & Wickens, 2020; Sharpe et al., 2017; Steinberg et al., 2013) or when performance-related feedback is provided (Blanco et al., 2014; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Luque et al., 2018;

Mitchell et al., 2005, 2006). This provides support that reinforcement learning (i.e., learning associations between events via trial and error) relies on an error-driven learning algorithm (Gershman & Daw, 2017). Another powerful form of learning is known as statistical learning, often defined as the incidental extraction of regularities from the environment without intention (Batterink et al., 2019; Frost et al., 2019; Saffran et al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). In the context of statistical learning, we have limited information about how the learning process itself occurs. Several studies suggest that statistical learning may indeed similarly rely on prediction errors. In rats, dopaminergic activity in the ventral tegmental area is important for the formation of an association between two non-rewarding stimuli (Keiflin et al., 2019; Sharpe et al., 2017). In humans, statistical learning involves the ventral striatum (Klein-Flügge et al., 2019), which has been hypothesized to signal prediction errors (Klein-Flügge et al., 2019; McClure et al., 2003; O'Doherty et al., 2004). However, other researchers, using variants of forward blocking, did not find clear-cut evidence for error-driven statistical learning. Beesley and Shanks (2012) did not observe any forward blocking in contextual cueing experiments, where participants incidentally learnt the spatial relationship among distractors and targets in a visual search task. This procedure however deviates from classic forward blocking paradigms, which rely on a *temporal* prediction between a cue and a future outcome (Aggarwal et al., 2020; Aggarwal & Wickens, 2020; Blanco et al., 2014; De Houwer et al., 2005; De Houwer & Beckers, 2003; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Luque et al., 2018; Mitchell et al., 2006; Steinberg et al., 2013; Vandorpe et al., 2005). Two subsequent experiments (Morís et al., 2014; Schmidt & De Houwer, 2019) observed forward blocking of temporal associations only for material that was intentionally learnt, but not for incidentally learnt stimulus associations. Such learning conditions substantially deviate from a typical statistical learning scenario, where observers extract regularities without intention (Batterink et al., 2019; Frost et al., 2019; Sherman et al., 2020; Turk-Browne et al., 2010). While few studies investigated forward blocking in incidental learning, less is known about backward blocking in incidental learning. Importantly, there is evidence of retrospective revaluation (of which backward blocking is an instance) not only in adults and children (Griffiths et al., 2011; McCormack et al., 2009, 2013; Sobel et al., 2004) but also in 8-month-old infants (Sobel & Kirkham, 2006, 2007), who clearly did not follow any explicit task instructions. This suggests that backward blocking may be present even in incidental learning, where observers attune themselves to statistical regularities by simple passive exposure.

We set out to examine forward and backward blocking during statistical learning in a series of experiments. In some statistical learning experiments, participants

are exposed to a continuous stream of stimuli containing statistical regularities (Batterink et al., 2019; Batterink & Paller, 2017; Henin et al., 2021; Saffran et al., 1996; Turk-Browne et al., 2005, 2009). Other studies have instead presented two successive stimuli on each trial, with conditional probabilities controlling their pairing (Richter et al., 2018; Richter & de Lange, 2019). In terms of neural processing, both continuous streams (Kaposvari et al., 2018) and pairs (Meyer & Olson, 2011) show identical modulations of sensory responses after statistical learning, suggesting that both paradigms elicit similar learning processes. We opted for pairs of stimuli in order to connect our study to the classic forward and backward blocking paradigms (Kamin, 1969; Shanks, 1985). On every trial, we presented participants with two consecutive visual object stimuli and asked them to categorize the trailing object as either electronic or non-electronic. Unbeknownst to participants, we manipulated the conditional probabilities between the leading and trailing stimuli, such that each trailing image could be predicted on the basis of its preceding, leading image. After learning, we evaluated statistical learning by presenting participants with expected and unexpected image pairs and measuring their reaction time for categorization judgments of the trailing image. Successful learning was indexed by faster reaction times to expected relative to unexpected trailing stimuli (Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et al., 2005).

Experiment 1

Method

Preregistration and data availability

All experiments were preregistered on the Open Science Framework (https://osf.io/r243e for Experiment 1; https://osf.io/7kmtv for Experiment 2). All data and code used for the analyses are freely available on the Donders Repository (https://doi.org/10.34973/pwza-qh43). Deviations from the preregistration are mentioned as such and justified in the corresponding sections below.

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 92 participants performed the experiment. 42 of them were excluded based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' below) before they started the second training phase (i.e. before the relevant data for the analysis was collected). Importantly, our selection criteria

applied to a very simple task where the general population is expected to score at ceiling (Ferrari et al., 2022; Richter et al., 2018; Richter & de Lange, 2019). Thus, our criteria (i.e. accuracy below 80%) allowed us to exclude outliers who clearly underperformed either because they did not read the instructions carefully, or did not understand the requirements of the task, or did not pay enough attention to stimuli; accordingly, it is common in online experiments that approximately half of the participants shows careless and inattentive behavior (Al-Salom & Miller, 2019; Brühlmann et al., 2020). Consequently, we carried out our analyses on a subset of the population who showed high motivation and adequate attention to the stimuli, as required to support statistical learning (Richter & de Lange, 2019). 50 participants (18 females; mean age 25.80, range 18-40 years) were included in the final data analysis. In Supplementary Forward Blocking Experiment 1 with 100 participants (see Supplementary information 2), we found successful learning of stimulus transition probabilities ($b = 11.23$, $CI = [6.80, 15.59]$, Cohen's $d_z = 0.54$). From this observation, we concluded that 50 participants were an adequate sample size for Experiment 1.

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

In each experimental trial, participants were exposed to two images presented on the left or right side of the central fixation point in quick succession: a leading stimulus was followed by a trailing stimulus. For each participant, there were 4 leading objects and 4 trailing stimuli objects. Everyday objects were randomly chosen from a pool of 64 stimuli derived from Brady et al. (2008) per participant, thereby eliminating potential effects induced by individual image features at the group level. In each stimulus set, 50% of objects were electronic (consisting of electronic components and/or requiring electricity to function) and 50% were non-electronic. The expectation manipulation consisted of a repeated pairing of objects in which the leading object predicted the identity of the trailing object, thus over time making the trailing object expected given the leading object. Importantly, each trailing object was only (un)expected depending on which leading object it was preceded by. Thus, each trailing object served both as an expected and unexpected object depending on the leading object at test phase. In addition, trial

order was pseudo-randomized, with the pairs distributed equally over time. In sum, any difference between expected and unexpected occurrences cannot be explained in terms of familiarity, adaptation, or trial history. In addition, object position (left / right) was counterbalanced with respect to Expectation (expected / unexpected) and Condition (antedating / blocked / control). In other words, leading and trailing objects appeared equally often on the left or right side of the central fixation point across trials. As a result, the expectation manipulation did not depend on spatial position. Also, both hemi-fields were equally task-relevant, which fostered participants' attention to both sides. Throughout the experiment, participants needed to categorize the trailing object as electronic or non-electronic as fast as possible. This task was aimed at assessing any implicit reaction time (RT) benefits due to incidental learning of the temporal statistical regularities: upon learning, leading object could be used to predict the correct categorization response before the trailing object appeared. In addition to the main object categorization task, there was an oddball detection task involving the leading stimuli in the training phases (16% of all trials per participant): participants were required to press a specific button as soon as they saw an animate leading stimulus. The aim of the animate detection task was to ensure that participants also paid attention to the leading stimuli, such that the association would be better learnt. For each participant, 4 animate leading stimuli (i.e., 2 for antedating leading stimulus and 2 for blocked leading stimulus) were randomly chosen from a pool of 8 stimuli (Brady et al., 2008). Finally, there were attention check trials where participants were simply asked to press a specific key based on a message on screen (e.g., "Press left-arrow key"). The aim of these trials (7% of all trials per participant) was to monitor participants' vigilance (see 'Exclusion and inclusion criteria'). A fixation bull's-eye was presented in the center of the screen throughout the experiment.

The blocking paradigm comprised two consecutive training phases, followed by one test phase (see Figure 2.1a). During the two training phases, leading objects were perfectly predictive of their respective trailing objects (i.e. P(trailing | leading = 1) ; see Figure 2.1b). Participants were not informed about this deterministic association, nor were they instructed to learn this association at the beginning of the experiment. Therefore, the pair associations were likely learned incidentally. Note that the participants may, however, still develop explicit knowledge of the associations over the course of the experiment, which we tested in a final recognition task. In training phase 1, the leading object (A) was always followed by the same trailing object (X). In training phase 2, a novel leading object (blocked [B] leading object) was presented along with the leading object presented in training phase 1 (antedating [A] leading stimulus), hence creating a compound stimulus (AB).

This was followed by the same trailing object (X) as in training phase 1. In addition, two novel leading (object + object [CD]) and a trailing (object [Y]) objects were presented as a control condition. In the test phase, the leading stimulus of each condition (antedating [A] / blocked [B] / control [D]) was presented alone, followed by either the expected trailing object (based on the training phases), or an unexpected trailing object. Expected and unexpected object pairs were presented equally often to prevent any learning at this final test stage (see Figure 2.1c). In the test phase, control (D) trials were compared to blocked (B) trials to assess blocking while controlling for the amount of exposure. It should be noted that the amount of exposure to trailing object X and trailing object Y are not the same, given that trailing object Y was only introduced in the second learning phase. This difference is an inevitable feature in classic blocking paradigms. If we would have presented trailing object Y in isolation in an additional experimental phase or if we would have paired Y with another leading stimulus in training phase 1, this could have elicited latent inhibition (i.e., difficulty in learning associations as a result of pre-exposure, McLaren & Mackintosh, 2000). Thus, we opted for the classic blocking paradigm. Furthermore, the control trials in the test phase allowed us to assess whether new associations were learned during training phase 2.

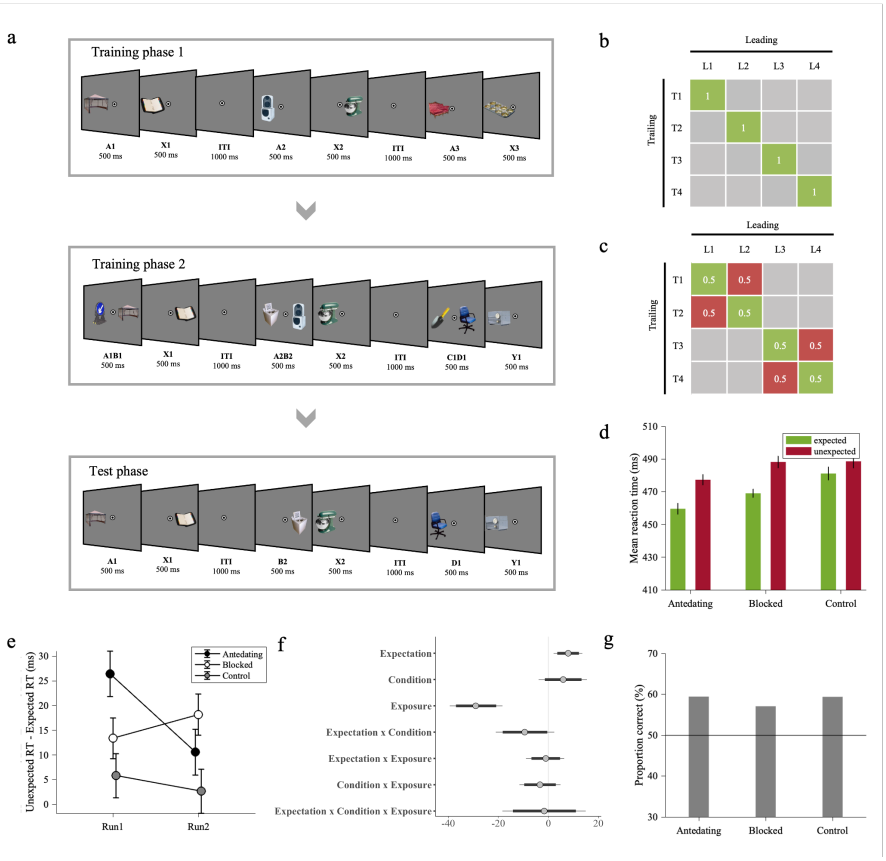


Figure 2.1. Experimental procedure and results of Experiment 1. (**a**) Experiment 1 comprised two training phases (training phase 1 and training phase 2) and a test phase. On every trial throughout the experiment, participants saw a pair of consecutively presented stimuli, i.e., a leading object followed by a trailing object. In training phase 1, the antedating leading object (i.e., A) was followed by a specific trailing object. In training phase 2, a novel blocked leading object (i.e., B) was presented in compound, along with the antedating (A) leading object (i.e., AB), and followed by the same trailing object from the antedating stimulus in training phase 1. In addition, we introduced novel control compound leading (i.e., CD) and trailing (i.e., Y) objects. In the test phase, antedating, blocked or control leading stimuli were followed by the associated (expected) or not associated (unexpected) trailing object. There were four different object pairs for ABX and CDY. Throughout the experiment, participants performed a categorization task on the trailing object. They reported, as fast as possible, whether the trailing object was electronic or non-electronic. (**b**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase 1 and training phase 2. Ls represent leading stimuli, and Ts represent trailing stimuli. There were 16 different leading objects and 8 different trailing objects coming from four different ABX and CDY pairs. (**c**) Statistical regularities depicted as image transition matrix with stimuli pairs in test phase. Green cells represent expected pairs, and red cells represent unexpected pairs. (**d**) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Reaction times were faster to expected than unexpected trailing objects in each condition. The reaction time difference between expected

and unexpected trials was greater in blocked than control trials, providing evidence for the absence of blocking effect and the augmentation of learning. (**e**) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. Please note that we split data into successive runs for visualization purposes only; data analysis was performed with number of trials as a continuous fixed factor (Exposure). The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning antedating condition. (**f**) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero. (**g**) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in all conditions indicating whether the trailing object was likely or unlikely given the leading object.

Data was collected during one single session per participant. Firstly, participants familiarized themselves with all trailing objects (both X and Y). In each trial, an object image was presented for 3500 ms, and participants had 1500 ms to categorize the object image as electronic or non-electronic (via a keyboard key press, keys counterbalanced across participants). Then, written feedback indicated the true category and the name of the object for 2000 ms (8 pairs × 2 trials / pairs = 16 trials in total). Afterwards, participants performed the experiment (i.e., training phase 1, training phase 2 and test phase). In each trial, the leading and trailing objects were presented for 500 ms successively with no inter-stimulus interval, followed by a 1500 ms inter-trial interval. Participants categorized the trailing object as electronic or non-electronic as fast as possible (via keyboard key press, keys counterbalanced across participants). Training phase 1 and training phase 2 started with a short practice period (practice training phase 1: 4 pairs × 4 trials / pairs = 16 trials in total; practice training phase 2: 8 pairs × 4 trials / pairs = 32 trials in total). After each practice, participants completed the training phases (training phase 1: 4 object pairs × 30 trials = 120 trials in total; training phase 2: 8 object pairs × 30 trials = 240 trials in total). In addition, animate detection and attention check trials (see above) were pseudo-randomly interspersed throughout the training phases without repetitions in successive trials. Afterwards, participants completed the test phase (12 pairs × 16 trials = 192 trials in total). Crucially, for each leading object, both expected and unexpected trailing objects belonged to the same category (electronic or non-electronic). This ensured that differences in RTs during object categorization would not arise by mere response adjustments costs, but instead reflected perceptual surprise to unexpected trailing objects.

Finally, at the end of the experiment participants performed a pair recognition task to probe their explicit knowledge of the statistical regularities. Before starting the recognition task, participants were informed about the presence of statistical regularities among leading and trailing images in the previous experimental phases (i.e., training phases 1 and 2), and they were asked to indicate whether the

trailing object was likely or unlikely given the leading stimulus according to what they saw during these previous phases. Participants familiarized themselves with the procedure via a brief practice (12 pairs × 2 trials / pairs = 24 trials in total) before completing the recognition task (12 pairs × 8 trials / pairs = 96 trials in total).

Exclusion and inclusion criteria

The online experiment was terminated if the percentage of correct responses during object categorization was below 80% (threshold was defined based on a preliminary pilot study) in any training or test phase (see 'Experimental design' and Figure 2.1a) or if the percentage of correct responses in attention check trials was below 80% in any of the experimental phases (see section 'Experimental design').

Prior to the main data analysis, we discarded trials with no responses, wrong responses, or anticipated responses (i.e., response time < 200 ms). We also rejected trial outliers (response times exceeding 3 MAD from mean RT of each participant) and subject outliers (participants whose RTs exceeded 3 MAD from the group mean). For the accuracy analysis of the pair recognition task, we rejected trial outliers in terms of response speed (response times exceeding 3 MAD from mean RT of each participant).

Data analysis

We analyzed the RT data in the test phase in order to test for incidental learning of predictable stimulus transitions: upon learning, participants were hypothesized to react faster to expected relative to unexpected trailing stimuli (Richter et al., 2018, Richter & de Lange, 2019). We did not statistically analyze the accuracy data in the test phase, given that the categorization task was not challenging, and performance was near ceiling levels (97% in Experiment 1 and 97% in Experiment 2). Furthermore, we analyzed the accuracy data in the pair recognition test to assess participants' explicit knowledge about learnt statistical regularities. For both analyses, we used a Bayesian mixed effect model approach. Data were analyzed using the *brm* function of the BRMS package (Bürkner, 2017) in R. Furthermore, in supplementary tables (see Supplementary information 1) we provide post-hoc Bayesian mixed effect models that follow significant interaction effects.

Analysis of RT data in test phase. Firstly, we modeled the behavioral data of the antedating condition, where one leading stimulus was followed by one trailing stimulus. This served as a sanity check to verify the baseline assumption that participants were able to learn the temporal association between the leading and trailing stimuli. The model of the antedating (A) condition included reaction time

as dependent variable and Expectation (unexpected / expected) as a fixed factor. To model the overall effect of time on task, we included Exposure as a continuous numeric predictor. Exposure was scaled between -1 and 1 to be numerically in the same range as the other factors, which aids model convergence. For the interpretation of the results, the model coefficient for Exposure represents the increase in RT from the first to the last exposure. Finally, we included the interaction between Exposure and Expectation in the model, to probe extinction of the learnt associations. Namely, during the test phase participants were exposed equally often to expected and unexpected stimulus pairs, potentially resulting in extinction of the RT advantage for expected stimuli over time. The model included a full random effect structure (i.e., a random intercept and slopes for all within-participant effects) to account for individual variance.

Secondly, we determined whether there was blocking by jointly modeling the blocked (B) and control (D) conditions. The model of blocked and control conditions included reaction time as a dependent variable and Expectation (unexpected / expected), Condition (control / blocked) and Exposure as fixed independent variables. We included the interaction between Expectation and Condition to test for the blocking effect. The contrasts of the factors Expectation and Condition were coded as successive difference contrasts. Exposure was a continuous predictor scaled between -1 and 1, as in the antedating condition analysis. Again, we also modeled extinction (Expectation × Exposure interaction) and its interaction with Condition to probe for potential differences in extinction between conditions. We adjusted the priors of the main effect of Expectation and Exposure and the prior of their interaction based on the posteriors of pilot experiments. Each prior was centered according to the median of the respective posterior estimate, and its standard deviation equated to the posterior estimate error times two to make the priors less informative. The prior for the Condition effect and its interaction with Expectation, i.e., blocking effect, was centered at zero. Note that specifying the priors in this way turns the estimates of Expectation and Exposure effects of Experiment 1 into the combined evidence from pilot experiments *and* Experiment *1*. Crucially, the pattern of results from Experiment 1 was exactly the same when not only the priors for the Condition effect but also for Expectation and Exposure were centered at zero. Further details and the complete model parametrization can be found in the R codes provided on the Donders Repository. The response time data was modelled using the ex-gaussian family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. We report posterior fixed effects model coefficients. Coefficients were accepted as convincing statistical evidence, analogously to statistically significant

in a frequentist framework, if the associated 95% posterior credible intervals were non-overlapping with zero.

Analyses of accuracy data in pair recognition test. Firstly, we determined whether accuracy was above chance level within each condition (antedating / blocked / control). Hence, we created three separate binomial mixed-effects models with response error as dependent variable. If accuracy was above chance level within each condition, we then determined whether there was a blocking effect in the explicit knowledge of implicitly learned associations. To do so, we created a binomial mixed-effects model with response error as binary dependent variable and Condition (blocked / control) as fixed factor. The models included a full random effect structure (i.e., a random intercept and slopes for the within-participant effects). The models were constructed using weakly informative priors centered at zero. All accuracy models were fit using Bernoulli family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. With respect to significance and amount of evidence we used the same criteria as for the RT data.

Results

Analyses of RT data in test phase. Firstly, we compared the reaction times of expected and unexpected trials in the antedating condition (see Table 2.2). We observed faster reaction times in expected (460 ms) than in unexpected (477 ms) trials (b = 10.81, CI = [5.04, 16.16], Cohen's d_z = 0.61, see Figure 2.1d), indicating successful learning of conditional probabilities and the consequent behavioral benefit of expectation in terms of response speed. In addition, we evaluated how this learning effect changed across exposure. Again, we observed an interaction effect between expectation and exposure (b = -9.01, CI = [-16.83, -1.18]), indicating that learning showed rapid extinction (expectation effect for run 1: 26 ms, run 2: 11 ms; see Figure 2.1e).

Table 2.2. Posterior fixed effects of the model of antedating condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*
Intercept	474.88	10.37	454.81 – 495.21
Expectation	10.81	2.83	5.04 – 16.16
Exposure	-23.39	3.70	-30.63 – -16.08
Expectation × Exposure	-9.01	4.00	-16.83 – -1.18

Next, we modeled the blocked and control conditions to test whether we found blocking (see Table 2.3 and Figure 2.1f). There was an interaction effect between expectation and condition (b = -9.48, CI = [-18.26, -0.45], Cohen's d_z= -0.26, see Figure 2.1b. We performed separate analyses for the blocked and control conditions to test for the presence of an expectation effect in each condition respectively. The reaction times in expected (481 ms) and unexpected (489) trials were not different from each other in the control condition (b = 4.36, CI = [-0.73, 9.51], Cohen's d_z= 0.20, see Table 2.S1). On the other hand, reaction times were clearly faster in expected (469 ms) than in unexpected (488 ms) trials of the blocked condition (b = 10.11, CI = [4.82, 15.16], Cohen's d_z= 0.65, see Table S2). Interestingly, this is exactly the opposite pattern of what would be expected under blocking, and rather supports better learning of the associations among blocked stimuli than control stimuli. Extinction was not different between blocked and control conditions (b = -1.63, CI = [-14.19, 11.00]; expectation effect in blocked condition for run 1: 13 ms, run 2: 18 ms; expectation effect in control condition for run 1: 6 ms, run 2: 3 ms; see Figure 2.1c).

Table 2.3. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*
Intercept	487.75	9.52	469.42 – 506.90
Expectation	7.92	2.18	3.57 – 12.23
Condition	5.87	3.71	-1.38 – 13.26
Exposure	-29.01	4.09	-36.93 – -20.88
Expectation × Condition	-9.48	4.49	-18.26 – -0.45
Expectation × Exposure	-1.05	2.92	-6.78 – 4.67
Condition × Exposure	-3.33	3.20	-9.63 – 2.97
Expectation × Condition × Exposure	-1.63	6.45	-14.19 – 11.00

Analysis of accuracy data in pair recognition test. Participants showed slightly above-chance level performance in indicating whether the trailing object was likely or unlikely given the leading object in the antedating (proportion correct = 59%; b = 0.39, CI = [0.26, 0.51]), blocked (proportion correct = 57%; b = 0.29, CI = [0.17, 0.42]) and control (proportion correct = 59%; b = 0.39, CI = [0.24, 0.54]) conditions (see Figure 2.1g). Response errors did not differ between the blocked and control conditions (b = -0.1, CI = [-0.08, 0.29]), indicating the absence of blocking effect for the explicit knowledge of incidentally learned associations.

Experiment 2

In Experiment 1, we observed a stronger reaction time benefit for B→X compared to control, indicating successful learning and the absence of forward blocking. We speculated that this pattern of results may be explained by the following process: upon learning the A→X association in the first training phase, attention may have shifted to the novel (and therefore potentially more salient) leading image B during the second training phase, thereby enhancing the learning of the B→X association. Importantly, this attentional mechanism is not at play in the related, but distinct paradigm of backward blocking (Shanks, 1985). Here, the order of training phases is reversed compared to forward blocking. Observers are first trained with AB→X association and presented in a subsequent training phase with the A→X association. As a result, both leading objects A and B are equally novel and salient during the first training phase and therefore should be learnt equally well. Therefore, we reasoned that backward blocking may allow us to study blocking without the potentially confounding factors related to novelty and salience. Crucially, this paradigm also allowed us to test for the first time whether retrospective revaluation takes place during incidental statistical learning.

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). Eighty-four participants performed the experiment. Thirty-three of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' below). One participant was excluded from the final data analysis due to overall excessively fast responses (i.e., 93% of responses being less than 200 ms). As a result, 50 participants

were included in the data analysis, as preregistered. This final number of included participants was based on the same sample size approach explained above.

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The design and procedure of Experiment 2 was identical in all respects to Experiment 1, apart from the fact that the order of elemental and compound training phases was reversed (see Table 2.1 and Figure 2.2a).

Data analysis

The data analysis of Experiment 2 was the same as for Experiment 1. Also here, we adjusted the priors of the main effect of Expectation and Exposure and the prior of their interaction based on the posteriors of the previous experiment, i.e., Experiment 1, because the stimuli and procedure regarding effects of Expectation and Exposure were exactly the same. Note that specifying the priors in this way turns the results of Experiment 2 with respect to Expectation and Exposure effects into the combined evidence from Experiments 1 *and* 2. Crucially, the pattern of results from Experiment 2 was exactly the same when the priors for Expectation and Exposure were also centered at zero.

Results

Analysis of RT data in test phase. First, we compared the reaction times of expected and unexpected trials in the companion condition to test whether repeated exposure to the pairs of the companion leading object A and trailing object X led to learning their temporal association (see Table 2.4). We observed faster reaction times in expected (477 ms) than unexpected (487 ms) trials (b = 8.90, CI = [3.53, 14.27], Cohen's d_z = 0.35, see Figure 2.2b), indicating successful learning of stimulus transition probabilities and the consequent behavioral benefit of expectation in terms of response speed. In addition, we tested whether this behavioral benefit remained stable during the test phase or tended to decrease as the exposure increased (i.e., extinction). We did not observe any interaction effect between Expectation and Exposure (b = 4.63, CI = [-12.41, 3.04]), indicating that learning did not show reliable extinction over time (expectation effect for run 1: 13 ms, run 2: 4 ms; see Figure 2.2c).

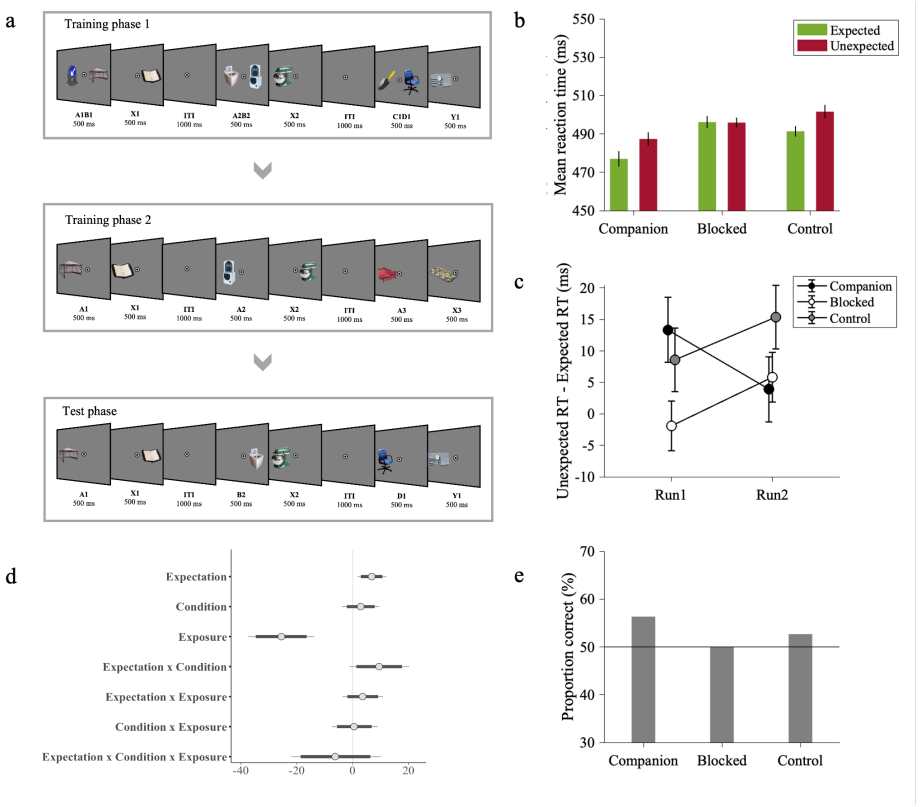


Figure 2.2. Experimental procedure and results of Experiment 2. (**a**) The design and procedure of Experiment 2 was identical in all respects to Experiment 1, apart from the order of training phases. (**b**) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (companion / blocked / control). Reaction times were faster to expected than unexpected trailing objects in companion and control conditions but not in blocked condition, providing evidence for the presence of backward blocking in statistical learning. (**c**) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. There was no extinction in learning in any conditions. (**d**) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero. (**e**) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in companion condition indicating whether the trailing object was likely or unlikely given the leading object.

Table 2.4. Posterior fixed effects of the model of companion condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	483.97	11.08	462.31 – 505.62
Expectation	8.90	2.75	3.53 – 14.27
Exposure	-14.12	4.38	-22.81 – -5.66
Expectation × Exposure	-4.63	3.93	-12.41 – 3.04

Next, we moved to our main question and tested for the presence of backward blocking (see Table 2.5 and Figure 2.2d). There was an interaction effect between expectation and condition (b = 9.45, CI = [1.34, 17.63], Cohen's d_z= 0.32, see Figure 2.2b). We performed separate analyses for the blocked and control conditions to test for the presence of an expectation effect in each condition respectively. The reaction times were faster in expected (491 ms) than in unexpected (501 ms) trials of the control condition (b = 8.44, CI = [3.60, 13.29], Cohen's d_z= 0.39, see Table S3). On the other hand, there was no evidence that reaction times in expected (496 ms) and unexpected (496) trials were different from each other in the blocked condition (b = 2.69, CI = [-2.08, 7.44], Cohen's d_z= -0.01, see Table S4). This pattern of results supports the presence of backward blocking. There was no extinction in blocked and control conditions (b = -6.22, CI = [-18.63, 6.38]; expectation effect in blocked condition for run 1: 0 ms, run 2: 6 ms; expectation effect in control condition for run 1: 9 ms, run 2: 15 ms; see Figure 2.2c).

Table 2.5. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	497.47	9.40	478.97 – 515.83
Expectation	6.81	1.98	2.94 – 10.65
Condition	2.86	2.54	-2.11 – 7.87
Exposure	-25.51	4.63	-34.68 – -16.49
Expectation × Condition	9.45	4.16	1.34 – 17.63
Expectation × Exposure	3.53	2.83	-1.95 – 9.11
Condition × Exposure	0.54	3.17	-5.57 – 6.80
Expectation × Condition × Exposure	-6.22	6.29	-18.63 – 6.38

Analyses of accuracy data in pair recognition test. Participants showed slightly above chance-level performance in indicating whether the trailing object was likely or unlikely given the leading stimulus in the companion (proportion correct = 56%; b = 0.25, CI = [0.10, 0.40])**,** but not in blocked (proportion correct = 50%; b = 0, CI = [-0.11, 0.11]) and control (proportion correct = 53%; b = 0.09, CI = [-0.04, 0.22]) conditions (see Figure 2.2e).

Discussion

Statistical learning allows us to detect and learn structure in the environment, with direct benefits for directing our limited processing resources more efficiently to optimize behavior. This results, for example, in more efficient behavioral processing (Fiser & Aslin, 2001, 2002; Hunt & Aslin, 2001; Saffran et al., 1996, 1999) and more efficient neural processing (Batterink & Paller, 2017; Henin et al., 2021; Richter et al., 2018; Richter & de Lange, 2019; Turk-Browne et al., 2009) for predictable than unpredictable events. While the benefits of statistical learning are obvious, the mechanisms of statistical learning itself are less clear. In separate experiments, we used respectively forward and backward blocking (Kamin, 1969; Shanks, 1985) to examine whether cue competition and retrospective revaluation, which have been observed during reinforcement learning, also apply to statistical learning. We found backward blocking, suggesting a retrospective revaluation process during the incidental extraction of statistical regularities.

In Experiment 1, participants learned the associations for the blocked (B) stimulus condition; in fact, learning was even stronger for B stimuli compared to control (D) condition, a phenomenon which is sometimes referred to as 'augmentation' (Batson & Batsell Jr., 2000; Beesley & Shanks, 2012; Vadillo & Matute, 2010). This pattern of results is opposite to the predictions of forward blocking and suggests the absence of forward blocking in statistical learning. One might argue that overall learning in the antedating and blocked conditions may not have been strong enough to generate forward blocking, given that the reduction in response speed was less than 20 ms. Such a small reaction time difference is, however, common in statistical learning (Richter et. al. 2018; Richter, & de Lange, 2019; Turk-Browne et. al., 2005) and similar in magnitude to RT benefits elicited by other cognitive factors such as probabilistic attentional cues (Posner, 1980).

We speculate that selective attention may provide a parsimonious explanation for the observed augmented learning in the blocked condition in Experiment 1.

Pearce-Hall model is one of the traditional models explaining forward blocking based on attention and prediction error. According to the model, during the second training phase, attention is divided equally to both antedating (A) and blocked (B) leading objects, and the outcome of blocked leading object is less surprising because the antedating leading object (A) already predicts the outcome (X). Thus, the association between B and X cannot be formed. On the other hand, in learning, stimuli whose consequences are initially unexpected may attract more attention (Holland & Schiffino, 2016; Pearce & Hall, 1980b), leaving open the possibility of the associability of B and X. Similarly, several recent studies show that attentional allocation may proceed in order to maximize learning. For example, observers preferentially attend to stimuli that are not completely predictable or unpredictable (Gottlieb et al., 2013; Kidd et al., 2012; Poli et al., 2020). In other words, their attention is drawn to stimuli that offer maximum information gain. In our experiment, the association between the antedating leading object (A) and the trailing object was learnt during the first training phase. Therefore, participants' attention may have shifted to the novel blocked (B) leading object during the second training phase, enhancing learning of the association between the blocked leading image and the trailing image. On the other hand, in the control (D) condition, two novel leading objects were presented in the second training phase. In line with overshadowing, these two leading objects may have competed for associative strength with the trailing object and hence their individual predictive power was reduced (Rescorla & Wagner, 1972). In Experiment 2, we aimed to eliminate this potentially attentional effect by applying a backward blocking procedure. Given that the blocked leading object (B) was presented together with a companion leading object (A) in training phase 1, both the companion leading object (A) and blocked leading object (B) were equally familiar and salient in the first phase of the study. As a result, we removed the potentially confounding factors related to novelty and salience, and crucially our results suggested that backward blocking occurs in statistical learning.

One may wonder whether the present forward and backward blocking experiments provide contradictory results regarding the presence of blocking in statistical learning. Here, it is worth noting that it is more difficult to obtain backward blocking than forward blocking, because more criteria need to be met to observe backward blocking (Van Hamme & Wasserman, 1994). Forward blocking only requires a strong A→X association, which is learned in the elemental training phase, to prevent learning the relationship between cue B and outcome X during the compound training phase. On the other hand, in backward blocking, a strong A→X association learned in the elemental training phase is not enough to observe blocking. In addition to that, the second important condition of backward blocking

is that cue A needs to be associated with cue B in order to decrease the associability of cue B in its absence, which is supported by previous studies (Luque et al., 2013; Melchers et al., 2006; Melchers et al., 2004).

By showing backward blocking in Experiment 2, our results suggest the presence of retrospective revaluation in statistical learning. Such retrospective revaluation cannot be explained by the traditional Rescorla – Wagner model, which assumes that the relevant cue must be present in order to change the associative strength (Kruschke, 2008; Miller & Witnauer, 2016; Rescorla & Wagner, 1972). However, a number of alternative models are able to explain this observation. Van Hamme and Wasserman (1994) proposed a modification of the traditional Rescorla – Wagner model, by allowing an update in the weight of an absent cue if the cue that is associated with the absent cue is present in that trial. Backward blocking can also be explained by a Bayesian generalization of the Rescorla – Wagner model, the Kalman filter (Gershman, 2015; Kruschke, 2008). In sum, our results in Experiment 2 can be explained by both the Van Hamme – Wasserman model and the Kalman filter, both of which claim that learning is based on prediction errors (Kruschke, 2008). At the computational level, this implies that statistical learning may be error-driven. At the implementation level, it supports the view that statistical learning may follow the principles of predictive coding (Hasson, 2017).

Critically, retrospective revaluation may be explained also by the probabilistic contrast model, which does not rely on prediction error (P. Cheng, 1997; P. W. Cheng & Novick, 1992). This model simply calculates how frequently events occur during learning. That is, X appears after either AB or A during training phases, and the probability of X increases after A and in the absence of B. As a result, observers associate A with X. Given that the probabilistic contrast model disregards the order of elemental (i.e., A->X) and compound (i.e., AB->X) training phases (De Houwer & Beckers, 2002), it explains both forward and backward blocking using the same approach. Although our backward blocking results can be explained by the probabilistic contrast model, the model fails to explain our forward blocking results. This supports the importance of the order of training phases in blocking (De Houwer & Beckers, 2002).

Furthermore, it is important to acknowledge that blocking may not arise due to learning deficits, as explained by the models reviewed above, but instead may depend on the failure to express cue – outcome associations at test, as explained by the so-called comparator hypothesis (Miller & Matzel, 1988; Miller & Witnauer, 2016). In other words, retrospective revaluation would not occur because of the

increase or decrease in the associative strength between cue and outcome, but rather because of a change in its expression at test. Although we observed backward blocking in statistical learning in Experiment 2, we do not know whether it is because of a learning deficit during training or because of a performance deficit observed at test. Thus, further studies are required to better understand the cause underlying backward blocking in statistical learning.

Crucially, learning regularities is usually thought to be incidental rather than intentional in statistical learning paradigms (Batterink et al., 2019; Batterink & Paller, 2017; Henin et al., 2021; Turk-Browne et al., 2005, 2009). However, this can nevertheless result in the development of explicit knowledge of the regularities (Batterink & Paller, 2017; Fiser & Aslin, 2002; Turk-Browne et al., 2005). Indeed, testing for explicit knowledge is often used to assess the development of explicit knowledge (Fiser & Aslin, 2002; Turk-Browne et al., 2005). In Experiment 1 and Experiment 2, we observed that people developed some minimal amount of explicit knowledge (on average 56% correct, with chance level of 50%) in the pair recognition test (i.e., how likely the trailing object was given the leading object). The temporal association between leading and trailing object was unknown to participants at the beginning of the experiment and participants were not instructed to learn these associations. Also, participants performed the categorization task at ceiling level (overall above 97%), suggesting that their categorization judgments were not affected by knowledge of the statistical structure between stimuli. Therefore, it appears likely that learning occurred incidentally, without strong explicit knowledge of the associations that were learnt. This is a clear difference between our studies and previous 'classic' blocking paradigms where learning occurs intentionally and in the presence of reinforcement (Aggarwal et al., 2020; De Houwer et al., 2002; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Mitchell et al., 2006).

Further, in the context of reinforcement learning, some highlight the key role of inferential reasoning for blocking to occur. Accordingly, learning associations between events does not depend on transitional probabilities but, instead, depends on the observers' belief about the nature of the relationship between cue and outcome (De Houwer & Beckers, 2002; Waldmann, 2000; Waldmann & Holyoak, 1992). Specifically, the intentional evaluation of causal associations between cues and outcomes (e.g., cue A is the cause of outcome X) appears necessary for forward blocking (De Houwer et al., 2005; De Houwer & Beckers, 2003; Vandorpe et al., 2005) and backward blocking (De Houwer & Beckers, 2002; Waldmann, 2000; Waldmann & Holyoak, 1992). As a result, there is so far evidence that conscious inferential reasoning contributes to backward blocking. To the best of our knowledge, the

present study is the first to examine backward blocking in incidental statistical learning. In our experiment, participants were not instructed about any possible relationship between leading and trailing objects, and they learned the associative relationship incidentally. Therefore, our finding supports that conscious inferential reasoning is not required for backward blocking to occur; instead, retrospective revaluation can happen during incidental statistical learning.

To sum up, while we did not find forward blocking, our results are compatible with the presence of backward blocking in statistical learning, a form of learning that develops incidentally and in the absence of rewarding outcomes or feedback. Our results are compatible with the Van Hamme – Wasserman model and Kalman filter, and thus support the idea that statistical learning may be error-driven, similar to reinforcement learning (though see the comparator hypothesis). Most importantly, our results suggest a retrospective revaluation process in statistical learning and thus support a functional similarity between statistical learning and reinforcement learning.

Supplementary information 1

Supplementary tables

Table S1. Posterior fixed effects of the post-hoc model of control condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	491.09	9.73	472.16 – 510.23
Expectation	4.36	2.59	-0.73 – 9.51
Exposure	-30.03	4.33	-38.34 – -21.45
Expectation × Exposure	-2.02	3.71	-9.26 – 5.21

Table S2. Posterior fixed effects of the post-hoc model of blocked condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	485.56	9.70	466.60 – 504.68
Expectation	10.11	2.65	4.82 – 15.16
Exposure	-27.38	3.94	-35.05 – -19.58
Expectation × Exposure	-0.95	3.68	-8.26 – 6.25

Table S3. Posterior fixed effects of the post-hoc model of control condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	500.48	9.47	481.97 – 519.28
Expectation	8.44	2.46	3.60 – 13.29
Exposure	-22.24	4.70	-31.37 – -12.93
Expectation × Exposure	0.38	3.57	-6.43 – 7.36

Table S4. Posterior fixed effects of the post-hoc model of blocked condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	499.01	9.42	480.60 – 517.78
Expectation	2.69	2.41	-2.08 – 7.44
Exposure	-22.72	4.38	-31.25 – -13.85
Expectation × Exposure	3.40	3.79	-4.11 – 10.83

Supplementary information 2

Supplementary Experiment 1

Method

Participants

The experiment was performed online using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 148 participants performed the experiment. 47 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria'), and one participant was excluded from the final data analysis due to excessively slow responses (RTs above 3 times the mean absolute deviation [MAD] from the group mean). As a result, one hundred participants (37 females; mean age 24.49, range 18-40 years) were included in the data analysis. This final number of included participants was preregistered based on previous research (Richter & de Lange, 2019; Schmidt & De Houwer, 2019) considering that online data would be noisier and, therefore, a larger number of participants would be required to maintain the same statistical power. The pre-selected sample size yielded 84% power to detect a small sized (Cohen's dz= 0.3) effect ($\alpha = 0.05$).

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The design and procedure of Supplementary Experiment 1 was identical in all respects to Experiment 1 (see Figure S1a). In line with the traditional blocking paradigm Supplementary Experiment 1 comprised two training phases (training phase 1 and training phase 2) and a test phase. During the two training phases, leading object were perfectly predictive of their respective trailing object (i.e. P(trailing | leading = 1) ; see Figure S1b). Yet during the final test stage Expected and unexpected object pairs were presented equally often to prevent any learning (see Figure S1c). The main difference between Supplementary Experiment 1 and Experiment 1 is related to the type of leading stimuli and stimulus location. Leading

stimulus was either a geometric shape or an everyday object. If the antedating leading stimulus was an object, then the blocked leading stimulus was a shape or vice versa. Both leading and trailing stimuli were presented at the center of the screen.

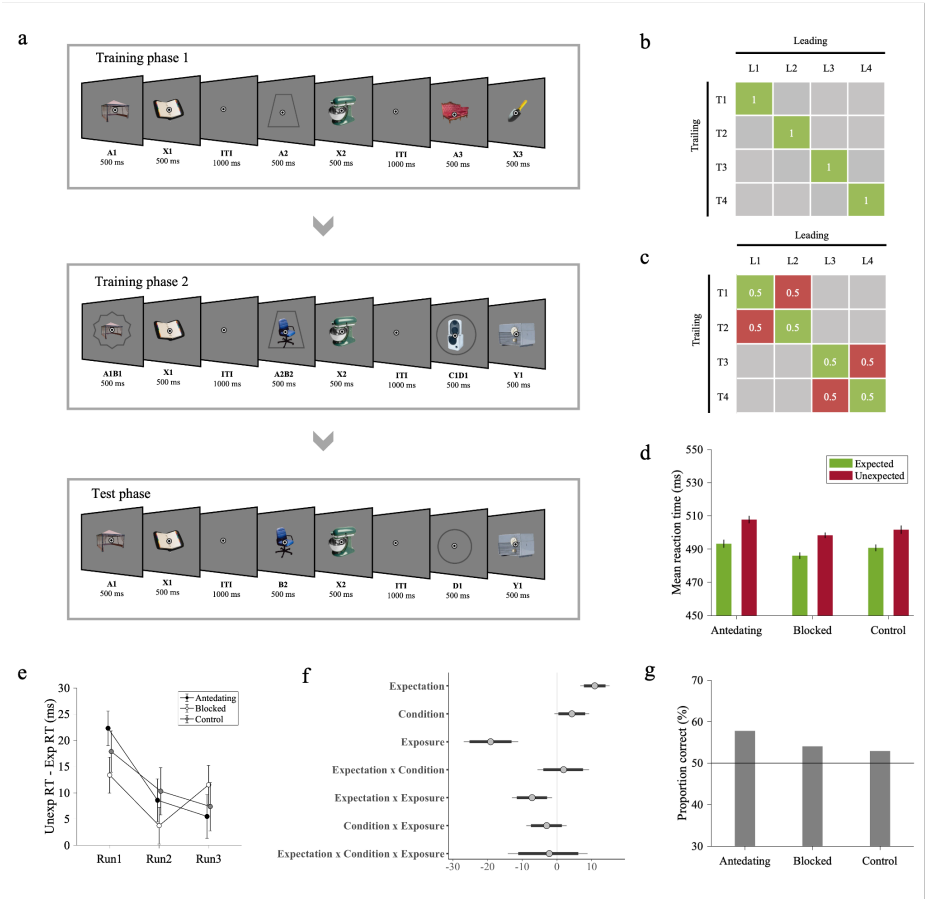


Figure S1. Experimental procedure and results of Experiment S1. (**a**) Experiment 1 comprised two training phases (training phase 1 and training phase 2) and a test phase. On every trial throughout the experiment, participants saw a pair of consecutively presented stimuli, i.e., a leading image followed by a trailing image. In training phase 1, the antedating leading stimulus (i.e., A), which could be either a shape or object, was followed by a specific trailing object. In training phase 2, a novel blocked leading stimulus (i.e., B) was presented in compound, along with the antedating (A) leading stimulus (i.e., AB), and followed by the same trailing object from the antedating stimulus in training phase 1. In addition, we introduced novel control compound leading (i.e., CD) and trailing (i.e., Y) stimuli. In the test phase, antedating, blocked or control leading stimuli were followed by the associated (expected) or not associated (unexpected) trailing object. Throughout the experiment, participants performed a categorization task on the trailing object. They reported, as fast as possible, whether the trailing object was electronic or non-electronic. (**b**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase 1 and training phase 2. Ls represent leading stimuli, and Ts represent

trailing stimuli. (**c**) Statistical regularities depicted as image transition matrix with stimuli pairs in test phase. Green cells represent expected pairs, and red cells represent unexpected pairs. (**d**) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Participants responded faster to expected than unexpected trailing objects in each condition. There was no difference between blocked and control conditions. (**e**) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. Please note that we split data into successive runs for visualization purposes only; data analysis was performed with number of trials as a continuous fixed factor (Exposure). Associations were rapidly extinguished during the test phase. Extinction was not different between conditions. (**f**) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero. (**g**) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in indicating whether the trailing object was likely or unlikely given the leading stimulus in all conditions.

Data analysis

The data analysis of Supplementary Experiment 1 was identical in all respects to Experiment 1, except for the priors and the additional analysis of RT data *split by stimulus type in test phase*. The models were constructed using weakly informative priors. The prior distributions for the effects of interest were Gaussian distributions with zero mean and standard deviation adjusted to expected effect sizes: 50 for Expectation, 70 for Condition and 30 for Exposure, 70 for the interaction between Expectation and Condition. Further details and the complete model parametrization can be found in the R codes provided on the Donders Repository.

Analysis of RT data split by stimulus type in test phase. We conducted a follow-up analysis that tested for the effect of the type of leading stimulus (shape / object). We reasoned that leading object stimuli may have attracted more attention than leading shape stimuli, given that they were visually more salient than the surrounding grey shapes, and their category was task-relevant, as the task required object categorization on the trailing image. Given that associative learning depends on attention (Kruschke, 2001; Pacton & Perruchet, 2008), it was therefore conceivable that leading objects, rather than shapes, developed a stronger temporal association with trailing objects. We fit the model of antedating condition and the model of blocked and control conditions as described above, but with the inclusion of leading Stimulus Type (shape / object) as additional fixed factor. The model included a full random effect structure (i.e., a random intercept and slopes for all within-participant effects). If the posterior credible intervals of the interaction effects between Expectation and leading Stimulus Type did not overlap with zero, we run separate models for shapes and objects respectively, in order to test for the blocking effect for each stimulus

type. The models were constructed using weakly informative priors centered at zero. All other analysis settings were as specified above.

Results

Analysis of RT data in test phase. First, we compared the reaction times of expected and unexpected trials in the antedating condition to test whether repeated exposure to leading-trailing pairs led to learning their temporal association (see Table S1). We observed faster reaction times in expected (493 ms) than unexpected (508 ms) trials (b = 11.23, CI = [6.80, 15.59], Cohen's d_z= 0.54, see Figure S1d), indicating successful learning of stimulus transition probabilities and the consequent behavioral benefit of expectation in terms of response speed. In addition, we tested whether this behavioral benefit remained stable during the test phase or dwindled, as would be expected by extinction. In line with the latter, we observed an interaction effect between Expectation and Exposure (b = -9.28, CI = [-15.26, -3.38]), indicating that learning showed rapid extinction (expectation effect for run 1: 22 ms, run 2: 9 ms, run 3: 6 ms; see Figure S1e).

Next, we moved to our main question and tested for the presence of blocking (see Table S2 and Figure S1f). The reaction time difference between unexpected and expected trials was not different between control (11 ms) and blocked (12 ms) conditions (b = 1.85, CI = [-3.95, 7.51], Cohen's d_z= -0.04, see Figure S1d). This pattern of results presents suggest for the absence of blocking. There was also no difference in how the reaction time benefit for expected items behaved over time (b = -2.29, CI = [-11.17, 6.13]; expectation effect in blocked condition for run 1: 13 ms, run 2: 4 ms, run 3: 12 ms; expectation effect in control condition for run 1: 18 ms, run 2: 10 ms, run 3: 7 ms; see Figure S1e).

Analyses of RT data split by stimulus type in test phase. In a follow-up analysis, we tested whether the type of leading stimulus (shape / object) affected statistical learning. In the antedating condition (see Table S3), the reaction time difference between unexpected and expected trials was larger for leading object (20 ms) compared to leading shape (9 ms) trials according to the posterior CI (b = -10.00, CI = [-18.57, -1.48]), which indicated that object-object associations were somewhat stronger than shape-object associations. While the difference in RT was larger for object-object associations than shape-object associations, separate follow-up models showed that the reaction time difference was significant when the leading stimulus was an object (b = 15.19, CI = [7.98, 22.46], see Table S1 and Figure S2a-e), and it was still significant (b = 5.44, CI = [0.83, 10.05], Table S2 and see Figure S2b-f).

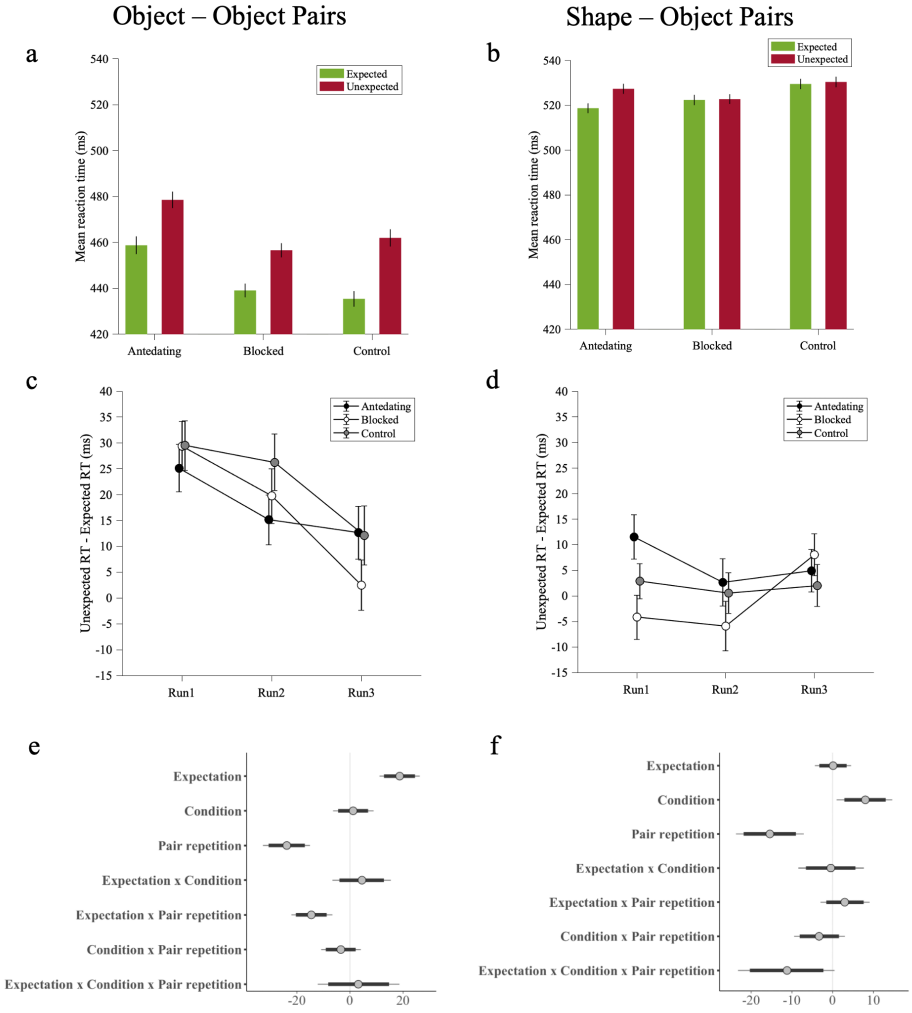


Figure S2. Results of Experiment S1 as a function of Stimulus Type. (**a-b**) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control) in leading objects (**a**) and leading shapes (**b**). The difference between expected and unexpected reaction times was larger for stimulus pairs with leading objects, compared to leading shapes. (**c-d**) Across participants' mean reaction time difference between expected and unexpected trials as a function of time in leading objects (**c**) and leading shapes (**d**). The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning only in leading objects. (**e-f**) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals in leading objects (**e**) and leading shapes (**f**). Estimates indicate significant results when they do not overlap with zero.

Across blocked and control conditions (see Table S4), the reaction time difference between unexpected and expected trials was also larger when the leading stimulus was an object (18 ms for B, 27 ms for D) compared to a shape (0 ms for B, 1 ms for D) (b = 18.40, CI = [11.52, 25.41]). Separate follow-up models showed that reaction times were faster in expected trials than in unexpected trials when the leading stimulus was an object (RT difference = 18 ms in blocked condition and 27 ms in control condition; b = 18.73, CI = [12.83, 24.5], see Table S3 and Figure S2a-e). This was not the case when the leading stimulus was a shape (RT difference = 0 ms in blocked condition and 1 ms in control condition; b = 0.11, CI = [-3.27, 3.44], Table S4 and see Figure S2b-f). Overall, the data suggest that shape – object associations could be learnt, but to a lesser extent than object – object associations. In particular, shape – object associations could be learnt only if a leading shape in isolation was followed by a trailing object (i.e., in the antedating condition), but not when the leading shape was concurrently paired with a leading object (in a compound stimulus) and then followed by the trailing object (i.e., in the blocked and control conditions). This pattern of results fits our prediction that leading objects attract more attention than shapes, given that they were visually more salient, and their category was task-relevant. As associative learning depends on attention (Kruschke, 2001; Pacton & Perruchet, 2008), this may have hampered associative learning between leading shapes and trailing objects. In other words, we found cue competition among the leading shape and object in the forms of overshadowing (Boddez et al., 2014; Pavlov, 1927; Schmidt & De Houwer, 2019), with the leading shape being overshadowed by the leading object. Finally, there was an interaction between Expectation, Condition and leading Stimulus Type (b = 4.09, CI = [-6.18, 15.80]), suggesting that the absence of blocking did not depend on leading Stimulus Type.

Analyses of accuracy data in pair recognition test. Participants showed slightly above chance-level performance in indicating whether the trailing object was likely or unlikely given the leading stimulus in the antedating (proportion correct = 58%; b = 0.32, CI = [0.23, 0.42]), blocked (proportion correct = 54%; b = 0.16, CI = [0.09, 0.24]) and control (proportion correct = 53%; b = 0.12, CI = [0.04, 0.19]) conditions (see Figure S1g). Response errors did not differ between the blocked and control conditions (b = -0.05, CI = [-0.15, 0.05]), indicating no blocking for the explicit knowledge of incidentally learned associations.

Supplementary Experiment 2

Supplementary Experiment 1 showed that the type of leading stimulus critically influenced statistical learning. Antedating and control leading shapes got less strongly associated with the trailing object than antedating and control leading objects. Moreover, blocked and control leading shapes could not compete with the concurrent leading objects for associative strength because they attracted less attention. This imbalance between shapes and objects may provide an alternative explanation for the lack of blocking that we observed. Therefore, in Supplementary Experiment 2 we made one modification to our paradigm and only presented objects as leading and trailing stimuli to remove any potential difference in attention between different leading stimuli, which might finally result in a blocking effect.

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 81 participants performed the experiment. 27 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' above). Four extra participants were excluded from the final data analysis: two showed accuracy below 50% chance level in test phase; two showed overall excessively slow responses (RTs above 3 MAD from the group mean). As a result, fifty participants (16 females; mean age 23.90, range 18-34 years) were included in the data analysis, as preregistered. This final number of included participants was derived from the following a priori power calculation: we aimed for 90% power to detect the effect size of Cohen's d_z= 0.5 derived in the antedating leading object condition of Experiment 1 ($\alpha = 0.05$).

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euros per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

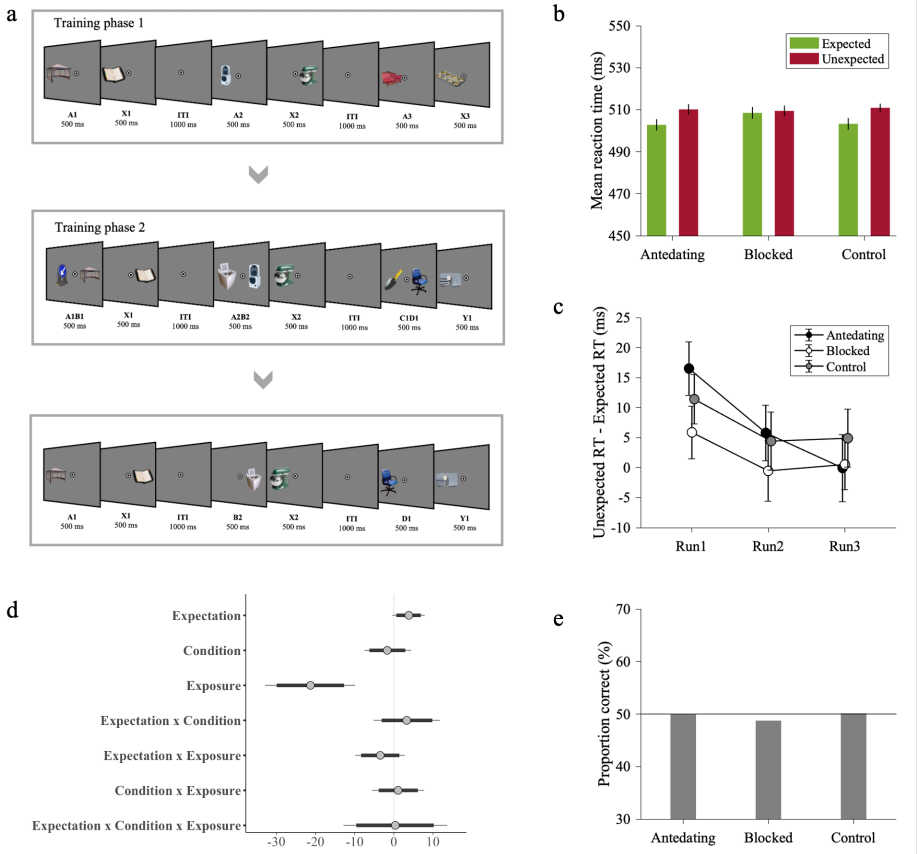


Figure S3. Experimental procedure and results of Experiment S. (**a**) The design and procedure of experiment 2 was identical in all respects to experiment 1, apart from the fact that the leading stimulus was an object presented in the left or right side of the fixation point, and it was followed by the trailing object presented in the left or right side of the fixation point. (**b**) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Reaction times were faster to expected than unexpected trailing objects in blocked and control conditions. There was no difference between blocked and control condition in terms of reaction time difference between expected and unexpected trials, providing evidence for the absence of blocking effect. (**c**) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning antedating condition. (**d**) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero. (**e**) Across participants' proportion correct responses in pair recognition test. Participants were not able to indicate above chance level whether the trailing object was likely or unlikely given the leading object in all conditions.

Experimental design

The design and procedure of Supplementary Experiment 2 was identical in all respects to Supplementary Experiment 1, apart from the type of leading stimuli and their location. Both leading and trailing stimuli were everyday objects. Leading and trailing objects were randomly presented on the left or right side of the central fixation point. Stimuli position (left / right) was counterbalanced with respect to Expectation (expected / unexpected) and Condition (antedating / blocked / control). In other words, leading and trailing objects appeared equally often on the left or right side of the central fixation point across trials. As a result, the expectation manipulation did not depend on spatial position. In addition, both hemi-fields were equally task-relevant, which fostered participants' attention to both sides.

Data analysis

The data analysis of Supplementary Experiment 2 was identical in all respects to Supplementary Experiment 1, except for omitting the factor Stimulus Type because this experiment featured only object stimuli (see Figure S3a).

Results

Analyses of RT data in test phase. First, we compared the reaction times of expected and unexpected trials in the antedating condition (see Table S5). We observed that reaction times for expected (503 ms) and unexpected (510 ms) trials, although showing a qualitative pattern similar to Experiment 1, were not significantly different from each other (b = 4.95, CI = [-0.07, 9.96], Cohen's d_z= 0.31, see Figure S3b). Therefore, unlike Experiment 1, our data do not provide robust support for learning of the conditional probabilities in condition A. There was however some statistical support for extinction, as the reaction time difference between expected and unexpected trials tended to decrease as the exposure increased (b = -8.17, CI = [-15.39, -0.91]) (expectation effect for run 1: 17 ms, run 2: 6 ms, run 3: 0 ms; see Figure S3c).

Next, we moved to our main question and compared reaction time differences between expected and unexpected stimulus pairs between B and C (see Table S6 and Figure S3d). The reaction time difference between unexpected and expected trials was not statistically different between control (8 ms) and blocked (1 ms) conditions (b = 3.34, CI = [-3.11, 9.85], Cohen's d_z= 0.24, see Figure S3b). Moreover, extinction was not different between B and C (b = 0.37, CI = [-9.60, 10.22]; expectation effect in blocked condition for run 1: 6 ms, run 2: -2, run 3: 0 ms; expectation effect in control condition for run 1: 11 ms, run 2: 4 ms, run 3: 5 ms; see Figure S3c).

Analysis of accuracy data in pair recognition test. Participants were not able to indicate above chance level whether the trailing object was likely or unlikely given the leading object in the antedating (proportion correct = 50%; b = 0, CI = [-0.15, 0.14]), blocked (proportion correct = 49%; b = -0.05, CI = [-0.17, 0.07]) or control (proportion correct = 50%; b = 0, CI = [-0.13, 0.14]) conditions (see Figure S3e).

Supplementary tables

Table S1. Posterior fixed effects of the model of antedating condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	502.44	8.42	485.21 – 518.66
Expectation	11.23	2.25	6.80 – 15.59
Exposure	-15.14	3.51	-22.08 – -8.19
Expectation × Exposure	-9.28	3.01	-15.26 – -3.38

Table S2. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	494.45	8.34	478.02 – 510.93
Expectation	10.88	1.6	7.76 – 13.98
Condition	4.30	1.95	0.38 – 8.10
Exposure	-19.10	3.08	-25.19 – -13.08
Expectation × Condition	1.85	2.91	-3.95 – 7.51
Expectation × Exposure	-7.19	2.24	-11.61 – -2.87
Condition × Exposure	-3.00	2.26	-7.49 – 1.40
Expectation × Condition × Exposure	-2.29	4.48	-11.17 – 6.13

Table S3. Posterior fixed effects of the model of antedating condition on reaction times split by stimulus type in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	502.27	8.42	485.68 – 518.72
Expectation	10.37	2.18	6.15 – 14.62
Leading stimulus type	56.95	5.6	46.13 – 67.99
Exposure	-15.35	3.52	-22.36 – -8.31
Expectation × Leading stimulus type	-10.00	4.37	-18.57 – -1.48
Expectation × Exposure	-7.26	2.61	-12.36 – -2.18
Leading stimulus type × Exposure	10.55	3.81	3.02 – 18.15
Expectation × Leading stimulus type × Exposure	1.12	5.38	-9.32 – 11.81

Table S4. Posterior fixed effects of the model of blocked and control conditions on reaction times split by stimulus type in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	494.13	8.29	477.45 – 510.54
Expectation	9.30	1.65	6.05 – 12.49
Condition	4.58	1.96	0.71 – 8.47
Leading stimulus type	-79.57	6.11	-91.88 – -67.48
Exposure	-19.54	3.03	-25.49 – -13.66
Expectation × Condition	1.97	2.60	-3.13 – 7.09
Expectation × Leading stimulus type	18.40	3.62	11.52 – 25.41
Condition × Leading stimulus type	-6.78	3.88	-14.36 – 0.89
Expectation × Exposure	-5.79	1.90	-9.53 – -2.06
Condition × Exposure	-3.45	1.98	-7.35 – 0.37
Leading stimulus type × Exposure	-8.64	2.93	-14.29 – -2.90
Expectation × Condition × Leading stimulus type	4.90	5.55	-6.18 – 15.80
Expectation × Condition × Exposure	-3.96	3.77	-11.36 – 3.41
Expectation × Leading stimulus type × Exposure	-17.54	3.70	-24.82 – -10.32
Condition × Leading stimulus type × Exposure	-0.21	3.74	-7.41 – 7.13
Expectation × Condition x Leading stimulus type × Exposure	14.37	7.56	-0.59 – 28.99

Table S5. Posterior fixed effects of the model of antedating condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	512.83	18.97	475.52 – 549.47
Expectation	4.95	2.51	-0.07 – 9.96
Exposure	-18.40	4.29	-26.74 – -10.02
Expectation × Exposure	-8.17	3.70	-15.39 – -0.91

Table S6. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.

Predictors	***Estimate***	***Est. Error***	***CI (95%)***
Intercept	515.30	19.42	476.89 – 533.55
Expectation	3.82	1.61	0.64 – 6.90
Condition	-1.68	2.33	-6.23 – 2.94
Exposure	-21.29	4.38	-29.92 – 12.70
Expectation × Condition	3.34	3.28	-3.11 – 9.85
Expectation × Exposure	-3.47	2.51	-8.35 – 1.42
Condition × Exposure	1.12	2.58	-3.86 – 6.15
Expectation × Condition × Exposure	0.37	5.08	-9.60 – 10.22

Chapter 3

What type of associations modulates statistical learning?

Abstract

Statistical learning refers to the acquisition of statistical regularities without an intention or an instruction to learn. Previous studies in statistical learning show that observer learn the relationship between events based on the strong predictive relationship or conditional probability (i.e., $CP = P(X|A) = \frac{P(X\&A)}{P(A)}$). However, the studies on more deliberative forms of learning state that learning is based on the unique predictive relationship by means of either ΔP (i.e., $\Delta P = P(X|A) - P(X|\sim A)$) or Dual Factor Heuristic (i.e., $DFH = \sqrt{P(X|A) \times P(A|X)}$). In this study, we want to better understand what type of relationship governs statistical learning. Across three consecutive experiments, we found that participants learned the relationship between object images based on their unique predictive relationship rather than their strong predictive relationship; however, we could not reach a conclusion about which type of unique predictive relationship determines statistical learning.

Contributing authorsz
Ilayda Nazli, Floris P. de Lange

Introduction

The world is full of repetitive structures. Therefore, agents can form associations between events to make predictions about the future, facilitating perception and decision-making. Observers can detect and acquire the regularities from the environment automatically and incidentally in the absence of rewarding/punishing outcome or feedback. This form of learning is known as statistical learning (Batterink et al., 2019; Frost et al., 2019; Saffran et al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). Previous studies in statistical learning show that observers can learn associations between events, i.e. how often X appears with or following A (i.e., conditional probability). On the other hand, studies on contingency learning and causal reasoning, which involve recognizing the regularities and cause-and-effect relationship (Shanks, 2010), suggest that rather than associations between events, observers learn the unique predictive association or how often X appears with A as well as how often X appears without A. According to the study of Leshinskaya and Thompson-Schill (2021), during incidental statistical learning participants learned associations between events not based on their conditional probability but rather on their unique predictive relations by means of *ΔP*. Unfortunately, their study did not tell us anything about the role of *DFH* in statistical learning. Here we extend their study to better understand how unique predictiveness of stimuli influence statistical learning by focusing on which forms of uniqueness influence statistical learning.

The association between two events can be represented by 2×2 matrix as shown in Figure 1.2. This figure shows the association between A and X: A is followed by X. A and -A respectively represent the occurrence and non-occurrence of leading stimulus A, and X and -X respectively represent the occurrence and non-occurrence of trailing stimulus X. The letters in the cells (i.e., a, b, c, d) represent the relative frequencies of the presence and absence of A and X: a cell shows the number of 'A is followed by X (A → X)' observations, b cell shows the number of 'A is followed by a different trailing stimulus (A → Y)' observations, c cell shows the number of 'X follows a different leading stimulus (B → X)' observations and d cell shows the occurrence of neither A nor X (B → Y). These four cells together determine the association strength, but differently for different metrics. One commonly used index is conditional probability which relies only on the a and b cells (i.e., $P(X|A) = \frac{P(X\&A)}{P(A)} = \frac{a}{a+b}$). Suppose that in most occurrences of A, X appears following A. This makes P(X | A) at high rate representing a strong relationship between A and X. In this case we expect observers to learn the A → X association strongly. However, imagine that X also appears in the absence of A, for example

following a different leading stimulus B quite often. In this case X is not unique given A; hence, it is not ideal for observers to learn the A → X association strongly. Given that conditional probability relies on a and b cells only, it cannot capture this type of relationship. To do this, c and d cells need to be included in the association index. One of the most well-known index capturing uniqueness is *ΔP* (Allan & Jenkins, 1980). According to *ΔP*, learning is based not only on how often X follows A but also on how often X appears in the absence of A (i.e., $\Delta P = P(X|A) - P(X|{\sim}A) = \frac{a}{a+b} - \frac{c}{c+d}$). Therefore, in the example above where both A and B are the strong predictive of X, *ΔP* of X given A is low leading to weak A → X association.

Observers using strategy to form associations focus equally on the occurrence and non-occurrence of events and process the relative frequencies of the presence and absence of A and X systematically and rationally (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005). However, it may be that observers focus on these four cells differentially (Béghin et al., 2021; Matute et al., 2015). They may mainly focus on the occurrence of events and disregard the d cell (Béghin et al., 2021; I. Hattori et al., 2017). For this reason Hattori and Oaksford (2007) proposed a different index called Dual Factor Heuristic (i.e., $DFH = \sqrt{P(X|A) \times P(A|X)} = \frac{a}{\sqrt{(a+b)\times(a+C)}}$). As opposed to the rational and analytic *ΔP*, observers using *DFH* tend to focus on the occurrence of events while disregarding the d cell and tend to process the relative frequencies of the presence and absence of A and X rapidly and with low effort (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005).

In the current study, we aimed to understand whether statistical learning is more sensitive to uniqueness rather than conditional probability (Experiment 1) and which forms of uniqueness (i.e., *ΔP* or *DFH* governs statistical learning (Experiment 2 and 3). On every trial, we presented participants with two consecutive visual objects and asked them to categorize the trailing object as either electronic or non-electronic. Unbeknownst to participants, we manipulated the unique prediction of trailing object from leading object such that the trailing object follows a certain leading object mostly and a different leading object or objects occasionally. After learning, we evaluated statistical learning by presenting participants with expected and unexpected object pairs and measuring how fast they categorized the trailing object. Successful learning was indexed by faster reaction times to expected relative to unexpected trailing object (Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et al., 2005). In Experiment 1, we showed that participants learned high unique pairs stronger than low unique pairs although both had conditional

probability of 1. In Experiment 2 and 3, we could not reach a conclusion about whether *ΔP* or *DFH* governs statistical learning. Yet, we showed that participants learned unique pairs strongly although they had conditional probability of 0.5. This suggest that although previous studies assume that conditional probability governs statistical learning, the uniqueness of association between events may be more important for statistical learning.

Experiment 1

Method

Preregistration

All experiments were preregistered on the Open Science Framework ("https://osf.io/jaubr" for Experiment 1; "https://osf.io/gcrjk" for Experiment 2; "https://osf.io/e6gck" for Experiment 3).

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 77 participants performed the experiment. 27 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' below). As a result, 50 participants were included in the data analysis, as preregistered. This final number of included participants was derived from the following a priori power calculation: we aimed for 90% power to detect a medium effect size (Cohen's $d_z = 0.5$), as derived from a Supplementary Experiment 1 (N=100).

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (10 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The experimental procedure consisted of one training phase followed by one test phase (see Figure 3.1a). In training phase, leading objects L1 and L2 were always followed by trailing object T1, and leading object L3 was followed by trailing object T2.

There were 2 sets of these pairs and they were repeated 28 times. The problem here was that T1 appeared more often than T2 due to the nature of the experimental manipulation. Therefore, we had another set of leading – trailing object pairs with the same relationship pattern (i.e., L4→T3, L5→T3, L6→T4) but presented twice more often (i.e., the pairs were repeated 56 times). The amount of pair repetition in T1 and T2 were the same, but only T3 was unique to its preceding leading object. On the other hand, although L4→T3 was not unique as opposed to L3→T2, the amount of exposure to L4→T3 was higher. This allowed us to compare the effect of exposure and uniqueness leading to four different conditions: low uniqueness-low exposure (i.e., two sets of L1→T1 pairs), high uniqueness-low exposure (i.e., two sets of L3→T2 pairs), low uniqueness-high exposure (i.e., two sets of L4→T3 pairs) and high uniqueness-high exposure (i.e., two sets of L6→T4 pairs, see Figure 3.1b). In the test phase, the leading object of each condition was followed by either the expected or unexpected trailing objects. Crucially, for each leading object, both expected and unexpected trailing objects belonged to the same category (electronic or non-electronic). This ensured that differences in RTs during object categorization would not arise due to response adjustments costs, but instead reflected perceptual surprise to unexpected trailing objects.

In each experimental trial, participants were exposed to a pair of objects presented in quick succession: a leading image was followed by a trailing image. There were 20 everyday objects and 4 animals which were randomly chosen from a pool of 80 stimuli (Brady et al., 2008) per participant in order to eliminate the potential effects induced by individual image features at the group level. A fixation point was presented in the center of the screen throughout the experiment. 50% of objects were electronic (consisting of electronic components and/or requiring electricity to function), and 50% were non-electronic. There were 18 pairs of objects in total. Object pairs were repeated in order to manipulate expectation. In other words, during the training phase, leading objects were followed by the same trailing object (i.e. P(trailing | leading = 1)), thus making the identity of the trailing object expected given the leading object over exposure in training phase.

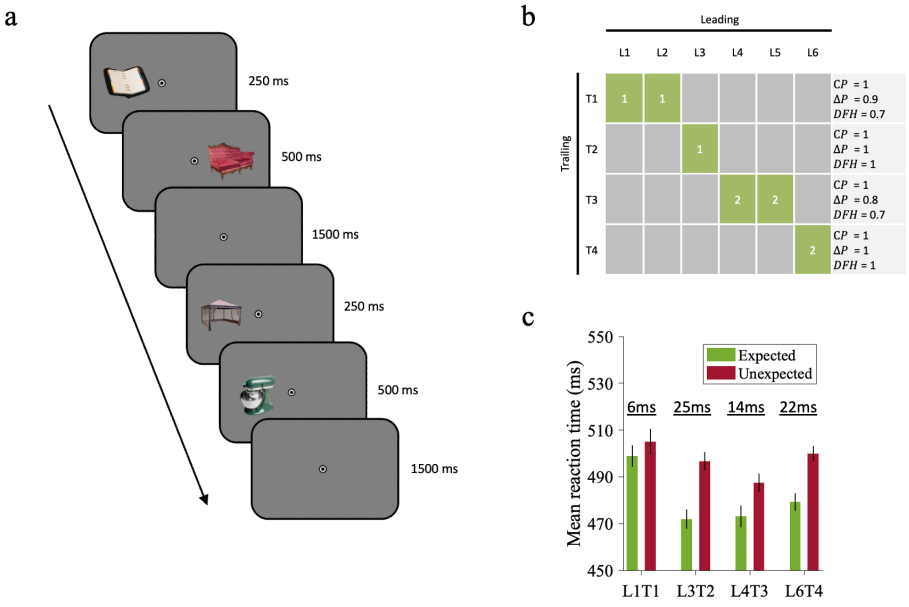


Figure 3.1. Experimental procedure and results of Experiment 1. (**a**) Experiment 1 comprised one training phase and a test phase. On every trial throughout the experiment, participants saw a pair of consecutively presented stimuli, i.e., a leading object followed by a trailing object. Throughout the experiment, participants performed a categorization task on the trailing object. They reported, as fast as possible, whether the trailing object was electronic or non-electronic. (**b**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase. Ls represent leading stimuli, and Ts represent trailing stimuli. (**c**) Across participants' mean reaction times as a function of Expectation and Condition. The reaction time difference between expected and unexpected trials was greater in high unique object pairs (i.e., L3T2 and L6T4) than low unique object pairs (i.e., L1T1 and L4T3).

In addition to this deterministic association between each leading and trailing objects, we manipulated uniqueness such that half of the trailing objects followed one single leading object making the relevant pair high unique whereas the other half followed two different leading objects making the relevant pair low unique. Participants were not informed about this deterministic association, and they were not instructed to learn this association at the beginning of the experiment. Therefore, the pair associations were likely learned incidentally. In test phase, expected and unexpected object pairs were presented equally often to prevent any learning at this final test stage. Throughout the experiment, the task of participants was to categorize the trailing object as electronic or non-electronic as fast and accurate as possible. We concentrated on the reaction time (RT) data from this task:

once temporal statistical regularities were learned incidentally, leading objects could be used to predict the correct categorization response before the trailing object appeared, which typically leads to implicit RT benefits (i.e., faster response to expected trailing object). In addition to the categorization task, in certain trials of training phases (20% of categorization-task trials per participant), animals were presented as leading images and followed by trailing objects randomly, thus preventing participants to develop predictive relationship between leading and trailing images. In these trials, the task of participants was to press a specific button as soon as they saw a leading animal. The aim of this animate detection task was to ensure that participants paid attention to the leading stimuli, such that the association would be learnt better. Leading and trailing images were randomly presented on the left and right side of the central fixation point. The position (left / right) of images were counterbalanced across trials in each experimental phase, such that the same image appeared equally often on the left and right side. This makes the expectation manipulation independent from the spatial position of objects. In addition, both hemi-fields were equally relevant, which fostered participants' paying attention to both sides. Lastly, there were attention check trials where participants were simply asked to press a specific key based on a message on screen (e.g., "Press left-arrow key"). The aim of these trials (20% of all trials per participant) was to monitor participants' vigilance (see 'Exclusion and inclusion criteria'). The trial order was pseudo-randomized. That is, the pairs were distributed equally over time, and the successive pairs were not identical. Therefore, any difference between expected and unexpected occurrences cannot be explained in terms of familiarity, adaptation, or trial history.

Data was collected during one single session per participant. Firstly, participants familiarized themselves with all objects and animals. In each trial, an image was presented for 2200 ms in the center of the screen, and participants had 1000 ms to categorize the image as electronic, non-electronic or animal via a key press. Then, written feedback indicated the true category and the name of the object for 1200 ms. All images were presented 2 times. Afterwards, participants performed the experiment (i.e., training phase and test phase). In each trial, the leading and trailing stimuli were presented for 500 ms successively with no inter-stimulus interval, followed by a 1500 ms inter-trial interval. Training phase with a short practice period with the pairs that were not presented in the main phase. After the practice, participants completed the training phases. There were 504 categorization task trials and 80 animate detection task trials in training phase. Afterwards, participants completed the test phase. There were 128 categorization task trials (i.e., each pair was repeated 16 times).

Exclusion and inclusion criteria

The online experiment was terminated if the percentage of correct responses during object categorization was below 80% (threshold was defined based on a preliminary pilot study) in any training or test phase (see 'Experimental design' and Figure 3.1a) or if the percentage of correct responses in attention check trials was below 80% in any of the experimental phases (see section 'Experimental design').

Prior to the main data analysis, we discarded trials with no responses, wrong responses, or anticipated responses (i.e., response time < 200 ms). We also rejected trial outliers (response times exceeding 2 SD from mean RT of each participant) and subject outliers (participants whose RTs exceeded 2 SD from the group mean). For the accuracy analysis of the pair recognition task, we rejected trial outliers in terms of response speed (response times exceeding 2 SD from mean RT of each participant).

Data analysis

We analyzed the RT data in the test phase in order to test for incidental learning of predictable stimulus transitions. We hypothesized that once learning occurred, participants reacted faster to expected relative to unexpected trailing object in high uniqueness. We did not statistically analyze the accuracy data in the test phase. This was because the categorization task was not challenging, which was supported by the performance near ceiling levels (mean accuracy was 86% in Experiment 1, 98% in Experiment 2, 97% in Experiment 3). We used a Bayesian mixed effect model approach. Data were analyzed using the *brm* function of the BRMS package (Bürkner, 2017) in R.

Analysis of RT data in test phase. The model included reaction time as dependent variable and Expectation (unexpected / expected), Uniqueness (low / high) and Exposure (low / high) as a fixed factor. To model the overall effect of time on task, we included Pair repetition as a continuous numeric predictor. Pair repetition was scaled between -1 and 1 to be numerically in the same range as the other factors, which aids model convergence. For the interpretation of the results, the model coefficient for Pair repetition represents the increase in RT from the first to the last exposure. Finally, we included the interaction between Pair repetition and Expectation in the model, to probe extinction of the learnt associations. Namely, during the test phase participants were exposed to expected and unexpected object pairs equally often, potentially resulting in extinction of the RT advantage for expected objects over time. We included the interaction between Expectation and Uniqueness to test for the effect of uniqueness. We also modeled the 3-way

interaction between Expectation, Uniqueness and Exposure to test whether the effect of uniqueness on statistical learning varies with different amount of exposure.

The model included a full random effect structure (i.e., a random intercept and slopes for all within-participant effects). The contrasts of the factors Expectation, Uniqueness and Exposure were coded as successive difference contrasts. We adjusted the priors of the main effect of Expectation and Pair repetition and the prior of their interaction based on the posteriors of Experiment 1 of Nazlı et. al. (2022). Each prior was centered according to the median of the respective posterior estimate, and its standard deviation equated to the posterior estimate error times two to make the priors weakly informative. The response time data was modelled using the exgaussian family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. Coefficients were accepted as statistically significant if the associated 95% posterior credible intervals were non-overlapping with zero.

Results

Analysis of RT data in test phase. We observed main effect of expectation (b = 12.91, CI = [7.99, 17.74]) indicating overall successful learning and the consequent behavioral benefit of expectation in terms of response speed. Next, we moved to our main question and tested for the effect of uniqueness. There was an interaction effect between expectation and uniqueness (b = 11.81, CI = [3.39, 20.36]) indicating that high unique pairs were learned stronger than low unique pairs. We did not observe 3-way interaction between Expectation, Uniqueness and Exposure (b = -13.73, CI = [-35.33, 8.2]) although the pattern of reaction time benefit (i.e., unexpected RT – expected RT) implied the possibility of the fact that the amount of exposure may compensate for low uniqueness (i.e., 6 ms RT benefit in low uniqueness-low exposure, 25 ms RT benefit in high uniqueness-low exposure, 14 ms RT benefit in low uniqueness-high exposure and 22 ms RT benefit in high uniqueness- high exposure, see Figure 3.1c). In addition, we tested whether this behavioral benefit remained stable during the test phase or tended to decrease as the exposure increased (i.e., extinction). We did not observe any interaction effect between Expectation and Exposure (b = -5.59, CI = [-11.18, 0.02]), indicating that learning did not show extinction over time.

Discussion

Experiment 1 shows that participants learned high unique object pairs stronger than low unique object pairs indicated by greater reaction time benefit in high uniqueness. This was despite the fact that the conditional probability of both high

and low unique object pairs was equivalent to 1. This finding provide additional support that observers are more sensitive to unique predictive relations between events rather conditional probabilities in visual statistical learning (Leshinskaya & Thompson-Schill, 2021).

Although we showed that uniqueness governs statistical learning, an open question is which forms of uniqueness influence statistical learning. For this, we computed and of each condition: was 0.94 and *DFH* was 0.71 in low uniqueness-low exposure, was 0.88 and *DFH* was 0.71 in low uniqueness-high exposure and both and *DFH* were 1 in high uniqueness-low exposure and high uniqueness- high exposure. Given that and *DFH* were high in the two low uniqueness conditions, it was not possible to make a clear interpretation. Therefore, to better understand the influence of different forms of uniqueness, we run Experiment 2 in which we kept *DFH* constant while varying .

Experiment 2

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 73 participants performed the experiment. 19 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' below). 4 participants were excluded from the final data analysis due to overall excessively slow or fast responses. As a result, 50 participants were included in the data analysis, as preregistered.

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (10 euros per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The design and procedure of Experiment 2 was identical in all respects to Experiment 1 apart from this critical manipulation in training phase

(see Figure 3.2a): Leading object L1 was followed by trailing object T1. Two sets of L1→T1 pairs generated control condition (i.e., *CP*=1, *ΔP*=1 and *DFH*=1) which was used as a sanity check to see whether participants were able to learn one-to-one associations. Leading objects L2 and L3 were followed by trailing object T2. These pairs generated high *ΔP* condition (i.e., *CP*=1, *ΔP*=0.89 and *DFH*=0.71). Leading object L4 was followed by trailing objects T2 and T3. These pairs generated low *ΔP* condition (i.e., *CP*=0.50, *ΔP*=0.50 and *DFH*=0.71). While keeping *DFH* constant and varying *ΔP*, we aimed to understand which forms of uniqueness governs visual statistical learning. If we observe a stronger RT benefit in high *ΔP* condition than low *ΔP* condition, we can argue that drives statistical learning. If we observe similar RT benefit in each condition, we can argue that *ΔP* drives statistical learning.

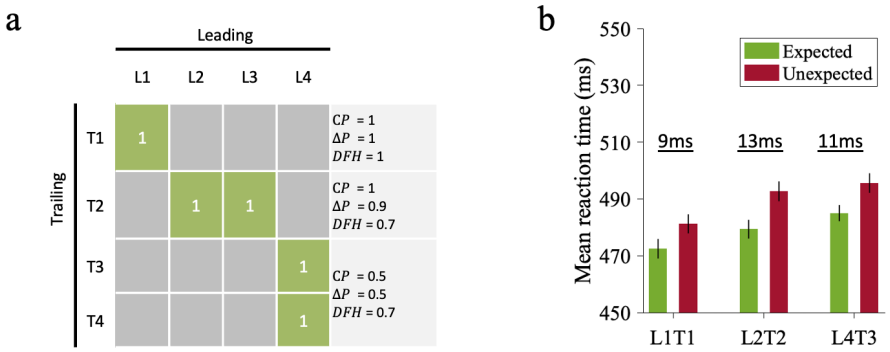


Figure 3.2. Experimental procedure and results of Experiment 2. (**a**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase. Ls represent leading stimuli, and Ts represent trailing stimuli. (**b**) Across participants' mean reaction times as a function of Expectation and Condition. The reaction time differences between expected and unexpected trials were written in the figure. Reaction times were faster to expected than unexpected trailing objects in each condition. There was no reaction time benefit difference between high *ΔP* (i.e., L2T2) and low *ΔP* (i.e., L4T3) trials.

Data analysis

The data analysis of Experiment 2 was identical in all respects to Experiment 1 apart from the following: the model included Expectation (unexpected / expected) and Condition (control / high *ΔP* / low *ΔP*) as a fixed factor. And we adjusted the priors of the main effect of Expectation and Pair repetition and the prior of their interaction based on the posteriors of Experiment 1. Each prior was centered according to the median of the respective posterior estimate, and its standard deviation equated to the posterior estimate error times two to make the priors weakly informative.

Results

Analysis of RT data in test phase. We observed main effect of expectation (b = 9.63, CI = [5.58, 13.64]), indicating overall successful learning and the consequent behavioral benefit of expectation in terms of response speed. Next, we moved to our main question and tested for which forms of uniqueness influence statistical learning. There was no difference in RT benefit both high *ΔP* and low *ΔP* (b = 3.23, CI = [-7.24, 13.73], 13 ms RT benefit in high and 11 ms RT benefit in low *ΔP*, see Figure 3.2b). And we did not observe any interaction effect between Expectation and Pair repetition (b = -3.25, CI = [-9.02, 2.52]), indicating that learning did not show extinction over time.

Discussion

In Experiment 2, while keeping *DFH* constant we varied *ΔP* to understand which forms of uniqueness influence visual statistical learning. We observed that participants learned pairs in low *ΔP* condition (i.e., *CP*=0.50, *ΔP*=0.50 and *DFH*=0.71) and in high *ΔP* condition (i.e., *CP*=1, *ΔP*=0.89 and *DFH*=0.71) equally well. This finding provides further support that uniqueness is more important than conditional probability for observers to learn regularities in object pairs automatically. Accordingly, participants learned pairs with *CP* of 0.50 as strong as pairs with *CP* of 1. Furthermore, the results imply that *DFH* may influence visual statistical learning rather than *ΔP* because participants learned pairs with *ΔP* of 0.50 as strong as pairs with *ΔP* of 0.89. To better understand which forms of uniqueness influence statistical learning, we sought to replicate the study with a slightly different design matrix with conditions in which *ΔP* and *DFH* were directly pitted against each other, by creating a modulation in opposite directions between the two in Experiment 3.

Experiment 3

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 180 participants performed the experiment. 69 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria' below). 11 participants were excluded from the final data analysis due to overall excessively slow or fast responses.

As a result, 100 participants were included in the data analysis, as preregistered. This final number of included participants was derived from the following a priori power calculation: we aimed for 90% power to detect a medium effect size (Cohen's d_z= 0.28) as derived from the interaction between low uniqueness-low exposure and high uniqueness-low exposure conditions of Experiment 1.

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (10 euros per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

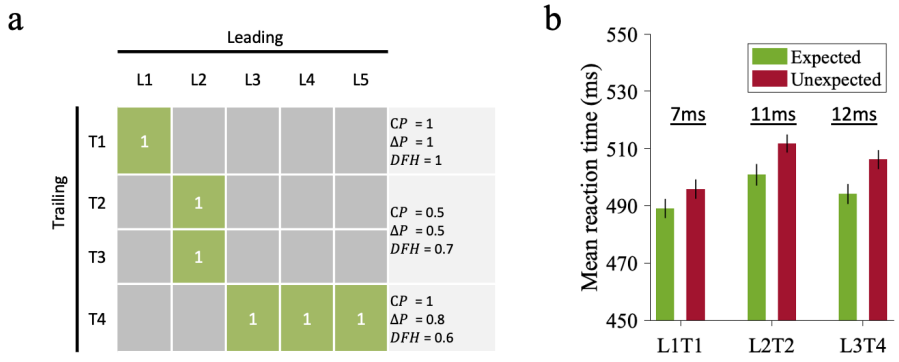


Figure 3.3. Experimental procedure and results of Experiment 3. (**a**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase. Ls represent leading stimuli, and Ts represent trailing stimuli. (**b**) Across participants' mean reaction times as a function of Expectation and Condition. The reaction time differences between expected and unexpected trials were written in the figure. Reaction times were faster to expected than unexpected trailing objects in each condition. There was no reaction time benefit difference between the three conditions, which varied in partly opposite directions in terms of *ΔP* and *DFH* (see panel a).

Experimental design

The design and procedure of Experiment 3 was identical in all respects to Experiment 1 and 2 apart from this critical manipulation in training phase (see Figure 3.3a): Leading object L1 was followed by trailing object T1. These pairs generated control condition (i.e., *CP*=1, *ΔP*=1 and *DFH*=1) which was used as a sanity check. Leading object L2 was followed by trailing objects T2 and T3. These pairs generated high *DFH* condition in which *DFH* is greater than *ΔP* (i.e., *CP*=0.50, *ΔP*=0.50 and *DFH*=0.71). Leading objects

L3, L4 and L5 was followed by trailing object T4. These pairs generated high *ΔP* condition in which *ΔP* is greater than *DFH* (i.e., *CP*=0.50, *ΔP*=0.82 and *DFH*=0.58). If we observe a stronger RT benefit in high *DFH* condition than high *ΔP* condition, we can argue that *DFH* drives statistical learning. If we observe a stronger RT benefit in high *ΔP* condition than high *DFH* condition, we can argue that *ΔP* drives statistical learning.

Data analysis

The data analysis of Experiment 3 was identical in all respects to Experiment 2 apart from the following: We adjusted the priors of the main effect of Expectation and Pair repetition and the prior of their interaction based on the posteriors of antedating condition in Experiment 2. Each prior was centered according to the median of the respective posterior estimate, and its standard deviation equated to the posterior estimate error times two to make the priors weakly informative.

Results

Analysis of RT data in test phase. We observed main effect of expectation (b = 9.85, CI = [5.91, 13.77]) indicating overall successful learning and the consequent behavioral benefit of expectation in terms of response speed. Next, we moved to our main question and tested for which forms of uniqueness influence statistical learning. There was no difference in RT benefit both high *ΔP* and high *DFH* (b = -4.06, CI = [-11.68, 3.56]), 11 ms RT benefit in high *ΔP* and 12 ms RT benefit in high *DFH*, see Figure 3.3b). Moreover, we observed an interaction effect between Expectation and Pair repetition (b = -5.09, CI = [-9.38, -0.25]), indicating that learning showed extinction over time.

Discussion

In Experiment 3, we observed that participants learned pairs in high *DFH* condition (i.e., *CP*=0.50, *ΔP*=0.50 and *DFH*=0.71) and in high *ΔP* condition (i.e., P=0.50, *ΔP*=0.82 and *DFH*=0.58) equally well. First, this finding provides further support that visual statistical learning is more sensitive to uniqueness rather than conditional probability because again participants learned pairs with *CP* of 0.50 as strong as pairs with *CP* of 1 Secondly, we observed that participants learned pairs with *DFH* of 0.71 as strong as pairs with *ΔP* of 0.82. Therefore, we cannot reach a conclusion about which forms of uniqueness determines the association strength during statistical learning.

General Discussion

Statistical learning enables us to use our limited processing resources more efficiently to optimize behavior by learning the repeated structure in the environment. The open question is what kind of structure is learned during statistical learning. Previous studies show that observers learn the strong predictive relationship between events or conditional probability. Yet, the recent study of Leshinskaya and Thompson-Schill (2021) showed that observers learn the association between events based on their unique predictive relationship rather than their conditional probability. Here, we aimed to better understand how unique predictiveness of stimuli influence statistical learning by focusing on which forms of uniqueness is used during statistical learning.

In Experiment 1, there were low unique pairs with *CP* of 1 and high unique pairs with *CP* of 1, and participants learned high unique pairs stronger than low unique pairs. In Experiment 2 and 3, there were high unique pairs with *CP* of 0.5 and *CP* of 1, and participants learned these pairs equally well. Thus, our findings imply that observers are more sensitive to unique predictive relationship between events rather than conditional probabilities during statistical learning. With Experiment 2 and Experiment 3, we aimed to understand which form of uniqueness (i.e., *ΔP* or *DFH*) is learned during statistical learning. However, the data remains inconclusive, and we cannot reach a conclusion about that. One possible explanation of this inconclusive result can be related to individual differences. Previous studies on uniqueness state that observers show individual differences in strategy use suggesting that some observers tend to rely on *ΔP* and some on *DFH* while forming associations between events (Béghin et al., 2021; Markovits et al., 2012). Thus, it may be possible that we did not observe any RT benefit difference between conditions in Experiment 2 and Experiment 3 conditions due to the individual differences cancelling each out on average.

Overall, our findings imply that during visual statistical learning observers learn the associations between events not based on their conditional probabilities but rather on their unique predictive relations. However, it remains an open question whether *ΔP* or *DFH* governs statistical learning. Future work may address this outstanding question by creating a stronger contrast between these different forms of uniqueness between stimuli.

Chapter 4

Does the uniqueness of visual associations modulate visual activity?

Abstract

Learning associations between events is crucial for us to make predictions about the future. In Chapter 3, we showed that observers learn unique predictive associations between events. In the present study, we set out to investigate what type of unique relationship (i.e., *ΔP* or *DFH*) governs statistical learning, and whether and how this modulates neural activity throughout the visual hierarchy, using fMRI in human volunteers. Participants were exposed to pairs of object images presented side-by-side based in a unique spatial arrangement, in which certain objects were presented in one spatial configuration mostly (i.e., expected spatial arrangement), while occasionally presented in a mirror reversed spatial arrangement (i.e., unexpected spatial arrangement). Unfortunately, we failed behavioral and neural evidence of statistical learning. Specifically, we did not find faster behavioral responses to expected compared with unexpected object pairs. Also, we did not observe the well documented suppression of neural responses to expected compared with unexpected object pairs within the ventral visual stream. We discuss several potential reasons for the lack of statistical learning in our specific paradigm.

Contributing authors

Ilayda Nazli, Matthias Ekman, Floris P. de Lange

Introduction

The environment is full of structural regularities and observers can detect and acquire these regularities automatically, which is known as statistical learning (Batterink et al., 2019; Frost et al., 2019; Saffran et al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). Learning regularities or forming associations between events allows us to make predictions about the future. What type of associative information is learned can be represented by a 2×2 matrix. Figure 1.2 summarizes the relationship between stimulus A and stimulus X. In particular, a cell shows the number of times that A and X appear together (i.e., A is followed by X or A→X), b cell shows the number of times that A appears without X and with a different stimulus (i.e., A→Y), c cell shows the number of times that X appears without A and with a different stimulus (i.e., B→X), and d cell shows the number of times that different stimuli other than A and X appear (i.e., B→Y). These four cells play roles in different association formulas. The most examined formula in statistical learning is conditional probability, which expresses how often X appears given A ($P(X|A) = \frac{P(X\&A)}{P(A)} = \frac{a}{a+b}$). However, observers track not only how often X appears given A but also if A can predict X uniquely and independently. Suppose that A is followed by X at a high rate. According to conditional probability, the associative link between A and X is strong. However, X appears following B quite often. In this case X is not unique given A; hence, it is not ideal for observers to form a strong link between A and X. This unique relationship can be captured by the well-known formula of *ΔP* ($\Delta P = P(X|A) - P(X|\sim A) = \frac{a}{a+b} - \frac{c}{c+d}$ Allan & Jenkins, 1980) and also by Dual Factor Heuristic ($DFH = \sqrt{P(X|A) \times P(A|X)} = \frac{a}{\sqrt{(a+b)\times(a+C)}}$, Hattori & Oaksford, 2007). The main difference between *ΔP* and *DFH* is based on the use of d cell. Observers using *ΔP* strategy process both the occurrence and non-occurrence of events systematically and rationally (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005) whereas observers using *DFH* strategy process only the occurrence of events rapidly and with low effort (Béghin et al., 2021; Hattori et al., 2017; Hattori & Oaksford, 2007; Markovits et al., 2012; Verschueren et al., 2005).

A recent study by Leshinskaya and Thompson-Schill (2021) showed that observers learned the unique relationship between events during visual statistical learning. In our online experiments in Chapter 3, we found that participants learned the unique pairs although their conditional probability was at chance level, implying that observers are more sensitive to the unique predictive relationship between events during statistical learning. However, the results of the previous online experiments remain inconclusive to clarify what type of uniqueness governs

statistical learning. This may be explained by the nature of online data collection. It is common in online studies that participants showed careless and inattentive behavior (Al-Salom & Miller, 2019; Brühlmann et al., 2020). Thus, here we replicated the study in a controlled lab environment to better understand which forms of uniqueness is learned during statistical learning. To further enhance the effect, we also manipulated the expectation in the spatial domain rather than the temporal domain because expectation in the former domain yields stronger behavioral and neural effects (He et al., 2022). On every trial, we presented participants with two colorful objects presented side-by-side. Unbeknownst to participants, we manipulated the unique spatial arrangement of objects: left-objects were presented on the left and right-objects were presented on the right mostly (i.e., expected spatial arrangement) and they were presented in the opposite side occasionally (i.e., unexpected spatial arrangement). One drawback of manipulating expectations in the spatial domain is the loss of the natural cause-and-effect relationship observed in the temporal domain. Uniqueness, particularly ΔP, is based on causal relationships and requires a clear temporal sequence. To avoid confounding results when shifting from the temporal to the spatial domain, we maintained a temporal relationship by presenting one object earlier than the other. After learning, we evaluated statistical learning by presenting participants with expected and unexpected spatial arrangements. Successful learning was indexed by faster reaction times to expected relative to unexpected spatial arrangement (Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et al., 2005) and by suppressed neural response to expected relative to unexpected spatial arrangement (He et al., 2022; Richter et al., 2018; Richter & de Lange, 2019). In brief, our results show no behavioral and neural response benefit to expected rather than unexpected pairs, suggesting the absence of statistical learning. We discuss the methodological aspects that may have contributed to these results.

Method

Preregistration

All experiments were preregistered on the Open Science Framework (https://osf.io/nczgx/).

Participants

28 participants were recruited from the Radboud research participation system. All participants were right-handed and MRI-compatible, had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric disorder.

They provided written informed consent and received financial reimbursement (15 euro per hour for fMRI session and 10 euro per hour for behavioral sessions) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

Behavioral Session 1 on Day 1. In each experimental trial, participants were exposed to a pair of objects (see Figure 4.1a). A fixation point was presented at the center of the screen throughout the experiment. The object on the one side was presented for 250 ms throughout the experiment. Then the other object appeared, and two objects were presented together for 500 ms, followed by a 1500 ms inter-trial interval. The position (left / right) of objects presented earlier were counterbalanced across participants. The object on the left was presented earlier for the half of the participant while the object on the right was presented earlier for the other half. The reason of this difference in stimulus presentation time is that *ΔP* manipulation is directional and requires temporal relationship between events. Presenting one of the objects slightly earlier gave that object some cue properties and enables us to compute *ΔP* of that object given the other object. There were 21 everyday objects randomly chosen from a pool of 80 stimuli (Brady et al., 2008) per participant in order to eliminate the potential effects induced by individual image features at the group level. 50% of objects were electronic (consisting of electronic components and/or requiring electricity to function), and 50% were non-electronic. Object image size was 5° × 5° visual angle, and images were presented 4° visual angle left and right from the central fixation dot on a mid-gray background. A 12×9 matrix was used to manipulate uniqueness (see Figure 4.1d), with 12 objects presented on the left and 9 objects presented on the right or vice versa. L1 was followed by R1 and R2, L2 was followed by R3 and R4, and L3 was followed by R5 and R6. These pairs generated high *DFH* condition (i.e., *CP*=0.5, *ΔP*=0.50 and *DFH*=0.71). L4, L5 and L6 were followed by R7, L4, L5 and L6 were followed by R8, and L10, L11 and L12 were followed by R9. These pairs generated high *ΔP* condition (i.e., *CP*=1, *ΔP*=0.89 and *DFH*=0.58). On 80% of the trials, left-objects were presented on the left and right-objects were presented on the right. This was the expected spatial arrangement. On 20% of the trials, left-objects were presented on the right and right-objects were presented on the left. This was the unexpected spatial arrangement. As a result, the expectation manipulation was independent from the response bias and ensured that differences in data would not arise due to the response adjustments costs, but instead reflected perceptual surprise to unexpected spatial arrangement. Throughout the Behavioral Session on Day 1, participants indicated if the category

of the objects (i.e., electronic or non-electronic) were the same or not as fast and accurately as possible. Each pair was repeated 60 times (i.e., 48 expected spatial arrangement and 12 unexpected spatial arrangement). There were 900 trials in total (i.e., 288 expected spatial arrangement and 72 unexpected spatial arrangement in high *DFH* condition and 432 expected spatial arrangement and 108 unexpected spatial arrangement in high *ΔP* condition).

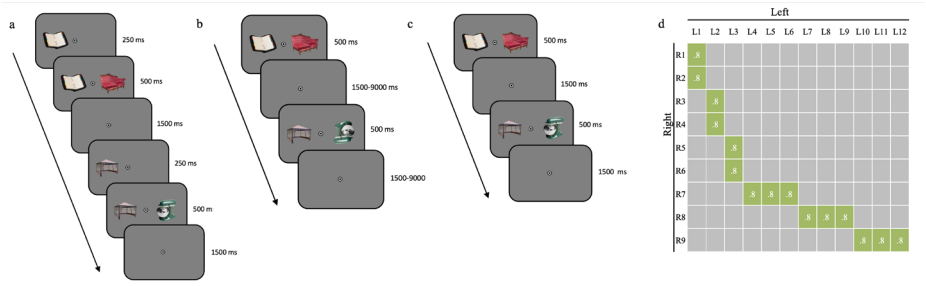


Figure 4.1. Experimental design and procedure. (**a**) On Behavioral Session 1 on Day 1, left objects was presented for 250 ms alone and together with right objects for 500 ms, followed by 1500 ms inter-trial interval. Participants were asked to indicate if the category of the two objects were the same or not as fast and accurate as possible. (**b**) On fMRI Session on Day 2, two objects were presented at the same time for 500 ms, followed by 1500-9000 ms inter-trial interval. Participants were asked to detect the image presented upside-down. (**c**) On Behavioral Session 3 on Day 2, two objects were presented at the same time for 500 ms, followed by a 1500 ms inter-trial interval. Participants were asked to indicate if the category of the two objects were the same or not as fast and accurate as possible. (**d**) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase. Ls represent left stimuli, and Rs represent right stimuli.

Behavioral Session 2 on Day 2. This was the short reminder task before the fMRI session. The design and procedure of Behavioral Session 2 was identical in all respects to Behavioral Session 1 apart from the number of trials. Each pair was repeated 10 times (i.e., 8 expected spatial arrangement and 2 unexpected spatial arrangement). There were 150 trials in total (i.e., 48 expected spatial arrangement and 12 unexpected spatial arrangement in high *DFH* condition and 72 expected spatial arrangement and 18 unexpected spatial arrangement in high *ΔP* condition).

fMRI Session on Day 2. The design and procedure of fMRI Session was identical in all respects to Behavioral Session 1 apart from this critical manipulation in design and task (see Figure 4.1b): Two objects were presented at the same time for 500 ms, followed by 1500-9000 ms inter-trial interval. On 16% of the trials, one of the two objects was presented upside-down. Participants were instructed to press the button as soon as they detected the flipped image.

Localizer. Following the main task, participants underwent a functional localizer to define object-selective regions for each participant. The functional localizer involved the same object images as in the previous tasks and their globally phase-scrambled versions. In a block design, the images were presented on the one side for 500 ms, followed by a 500 ms inter-trial interval. Each object was presented 6 times. Participants were instructed to fixate on the fixation point and press a button whenever it turned yellow.

Behavioral Session 3 on Day 2. The design and procedure of Behavioral Session 3 was identical in all respects to Behavioral Session 1 apart from these critical manipulations in design (see Figure 4.1c): Two objects were presented at the same time. And on 50% of the trials, expected spatial arrangement was presented.

fMRI Data Acquisition

Functional and anatomical MRI data were acquired on a 3T Prisma and PrismaFit scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. The data acquisition protocol included a T1-weighted anatomical and five functional runs. The anatomical scan was acquired with a Magnetization Prepared Rapid Acquisition Gradient Echo sequence (MP-RAGE; TR = 2300 ms, TI = 1100 ms, TE = 3 ms, flip angle = 8°, 1 × 1 × 1 mm3 isotropic). The five functional runs comprised of four main task runs and one localizer run. Functional images were acquired using a whole-brain T2*-weighted multiband-4 sequence (TR = 1000 ms, TE = 34 ms, flip angle = 75°, 2 × 2 × 2 mm3, 66 slices).

Data analysis

Behavioral data analysis. Reaction time (RTs) data from the object categorization task of Behavioral Session 1 on Day 1 and Behavioral Session 3 on Day 2 were analyzed. The data of Behavioral Session 2 on Day 2 was not analyzed due to the low number of trials in this short reminder session. The trials with no responses, wrong responses, fast responses (i.e., reaction time < 200 ms) and outliers (reaction times exceeding 2 SD from mean RT of each participant) were excluded from the analysis. Reaction times data of Behavioral Session 1 and 3 were entered into a 2 Expectation (expected – unexpected) x 2 Condition (*DFH - ΔP*) repeated measures ANOVA using JASP Team (2022, Version 0.16.3).

fMRI data preprocessing. MRI data were preprocessed using FSL (version 6.00; FMRIB Software Library, Smith et al., 2004). We applied brain extraction using BET, motion correction using MCFLIRT, spatial smoothing (Gaussian kernel of FWHM = 5 mm) and temporal high-pass filtering (120 s). All analyses were carried out in native

subject space. We used FSL FLIRT to register functional images to the anatomical image and the anatomical image to the MNI152 T1 2 mm template brain using linear registration (12 degrees of freedom).

fMRI data analysis. FSL FEAT was used to fit voxel-wise general linear models (GLM) to each participants' run in an event-related approach. The functional localizer was modelled with 6 regressors of interest (left/right, scrambled/unscrambled, expected/unexpected) and a set of 24 motion regressors. To define ROIs for object-selective LOC, for each participant, we first contrasted scrambled vs. unscrambled object presentations and then selected the 200 most active voxel from that contrast and then constrained to anatomic LOC derived from Freesurfer's Desikan-Killiany cortical atlas. The runs of the main task were modelled using 4 regressors of interest, namely *DFH* expected, *DFH* unexpected, *ΔP* expected, *ΔP* unexpected and 24 additional motion regressors. The contrast of interest for the whole-brain analysis compared BOLD activity during unexpected minus expected trials (i.e., expectation suppression), *DFH* minus *ΔP*, unexpected minus expected *DFH* trials and unexpected minus expected *ΔP* trials. FSL's fixed effects analysis was used to combine data across all runs. Whole-brain analysis across-participants was carried out using FSL's mixed effect model (FLAME 1). All ROI analyses were conducted in participants' native space. In addition to the functionally defined LOC region, we also used two a priori defined ROIs, namely primary visual cortex (V1) and temporal occipital fusiform cortex (TOFC) based on a previous study by Richter et. al. (2018).

Results

Analysis of Reaction Time Data. We analyzed reaction time data during Behavioral Session 1 and 3 to evaluate how behavioral benefits of expectations varies by the type of uniqueness. During Behavioral Session 1 on Day 1, a two-way within subjects ANOVA revealed no significant effect of expectation ($F(1, 27) = 1.67$, $p = 0.21$), a significant effect of condition ($F(1, 27) = 18.00$, $p < 0.001$; participants were 39 ms faster in *DFH* than *ΔP*) and no significant interaction effect ($F(1, 27) = 1.50$, $p = 0.23$; 7 ms RT benefit in high *DFH* and 1 ms RT benefit in high *ΔP*, see Figure 4.2a). During Behavioral Session 3 on Day 2, a two-way within subjects ANOVA revealed no significant effect of expectation ($F(1, 27) = 1.77$, $p = 0.19$), condition ($F(1, 27) = 0.32$, $p = 0.58$) and interaction ($F(1, 27) = 1.50$, $p = 0.23$, 10 ms RT benefit in high *DFH* and 1 ms RT benefit in high *ΔP*, see Figure 4.2b). Overall, we did not observe any behavioral benefit of expectation effect (i.e., faster response to

expected than unexpected pairs), which may imply that participants did not learn the most likely spatial arrangement of the object images.

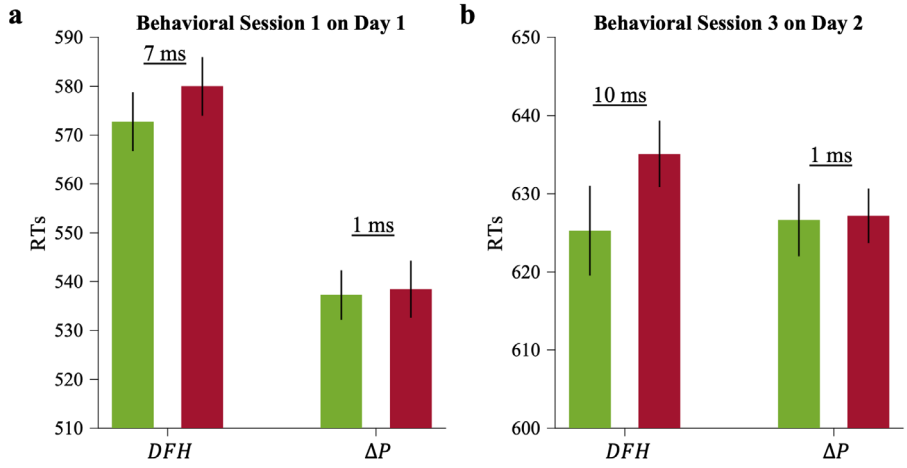


Figure 4.2. Pre- and Post-Scanning Behavioral Results. Across participants' mean reaction times as a function of Expectation and Condition. Displayed are parameter estimates + /- within subject SE for responses to expected (green) and unexpected (red) object images on Behavioral Session 1 (see panel a) and Behavioral Session 3 (see panel b). There was no significant reaction time benefit in high *DFH* and high *ΔP* conditions.

Analysis of fMRI Data. We analyzed BOLD responses within our ROIs (i.e., V1, object-selective LOC and TOFC) to evaluate how neural benefits of expectations varies by the type of uniqueness. In V1, a two-way within subjects ANOVA revealed no significant effect of expectation ($F(1, 27) = 0.00$, $p = 0.95$), condition ($F(1, 27) = 0.95$, $p = 0.34$) and interaction ($F(1, 27) = 0.03$, $p = 0.88$, see Figure 4.3a). In LOC, a two-way within subjects ANOVA revealed no significant effect of expectation ($F(1, 27) = 0.62$, $p = 0.44$), a significant effect of condition ($F(1, 27) = 6.21$, $p = 0.02$) and no significant interaction effect ($F(1, 27) = 0.02$, $p = 0.90$, see Figure 4.3b). In TOFC, a two-way within subjects ANOVA revealed no significant effect of expectation ($F(1, 27) = 0.15$, $p = 0.71$), a significant effect of condition ($F(1, 27) = 8.56$, $p = 0.01$) and no significant interaction effect ($F(1, 27) = 0.07$, $p = 0.80$, see Figure 4.3c). Overall, we did not observe expectation suppression (i.e., larger BOLD response to unexpected than expected pairs) supporting the failure of participants' learning the spatial relationship between objects. We observed larger BOLD response to *DFH* than *ΔP*. This may be explained by the fact that there were more *ΔP* trials than *DFH* trials in order to create contrast between *ΔP* P and *DFH* values for a clear comparison. However, this imbalance in the number of trials

made participants more familiar with *ΔP* P conditions. Increased familiarity with a stimulus leads to reduced neural responses (Fritsche et al., 2020; Manahova et al., 2020).

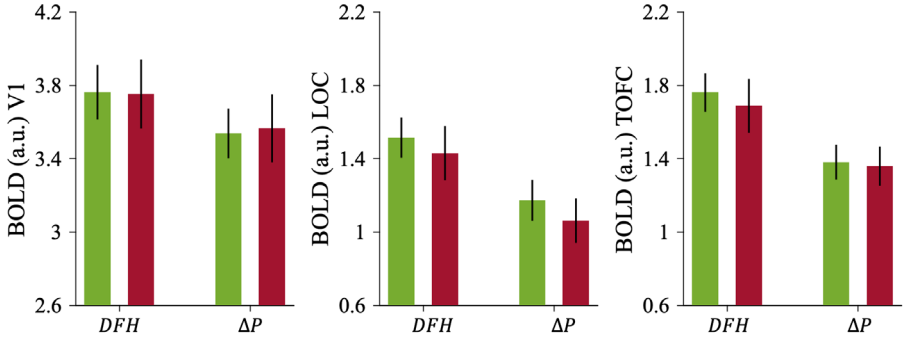


Figure 4.3. fMRI Results. Expectation suppression within V1 and object-selective LOC and TOFC. Displayed are parameter estimates + /- within-subject SE for responses to expected and unexpected object pairs. In all ROIs, BOLD responses to unexpected object pairs were not stronger than to expected object pairs.

Discussion

Statistical learning allows us to efficiently utilize our limited resources by extracting the repeated structure in the environment. One crucial open question is what kind of structure observers learn. Although previous studies show that observers learn the strong predictive relationship, recent findings imply that observers learn unique predictive relationships between events (Leshinskaya & Thompson-Schill, 2021). In this study, we aimed to understand what kind of unique relationship is learned during statistical learning. However, our results remain inconclusive to answer this question. Potentially, this could be explained by an absence of learning. Successful statistical learning is demonstrated by behavioral and neural response benefit. Observers are faster to react to expected than unexpected events (Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et al., 2005) and show suppressed neural response to expected than unexpected events (He et al., 2022; Richter et al., 2018; Richter & de Lange, 2019). In our study, there is a trend of faster response to expected than unexpected spatial arrangement in high *DFH* condition; however, this difference was numerically very small and statistically non-significant. Similarly, there was not any activity modulation in any of our ROIs. These results suggest that participants did not learn the most likely spatial arrangement of the object images.

We speculate that the lack of learning may be explained by the expectation manipulation in the spatial domain. In the unexpected spatial arrangement, we swapped the spatial locations of the same object images. Therefore, the correct responses for both expected and unexpected spatial arrangements were the same. This approach makes the expectation manipulation independent of response bias and ensures that differences in reaction time data were due to perceptual surprise at unexpected spatial arrangements rather than response adjustment costs. Yet, participants may have learned the relationship between objects based on the identity of the objects while disregarding their spatial arrangement. Instead of swapping the location of objects, we could have replaced one of the objects with another object from the same category, which could have also prevented the response bias to unexpected pairs. Moreover, this could have enhanced the learning effect by better attracting participants' attention, which is necessary for learning to takes place (Richter & de Lange, 2019). Indeed, He et.al. (2022) found both behavioral and neural expectation effects when they changed the identity of the image in the unexpected spatial arrangement.

Comparing design parameters from the previous chapter with those in the current chapter may help us understand why participants did not learn the relationship between object images in the current chapter. First, we can consider the amount of exposure to object pairs during training. In the previous chapter, each expected object pair was repeated 16 times across three online experiments. In the current chapter, each expected object pair was repeated 60 times. Although participants were exposed to object pairs less frequently in the previous chapter, they still showed a behavioural benefit of expectation. Therefore, the lack of expectation benefit in the current chapter cannot be explained by the amount of exposure. The second difference relates to the nature of the relationship between object images during the training phase. In the previous chapter, participants were first exposed to deterministic relationships during the training phase, whereas in the current chapter, they were exposed to probabilistic relationships. Recent research suggests that when participants are exposed to probabilistic rather than deterministic relationships, statistical learning becomes weaker (Richter, 2021). This exposure to probabilistic relationships during the training phase may have weakened the learning effects in the current chapter. Therefore, if participants had been presented with deterministic relationships in the first behavioral session on day 1, we might have observed a learning effect.

To conclude, we could not find any behavioral and neural response benefit to expected rather than unexpected pairs, suggesting the absence of statistical

learning. Therefore, it remains an open question whether *ΔP* or *DFH* governs statistical learning. To obtain more insightful results, future studies could create unexpected spatial arrangements by changing the identity of objects instead of swapping their locations. Additionally, participants could be trained on a deterministic relationship before introducing violations to that relationship during the fMRI session.

Chapter 5

Discussion

Throughout this thesis, I have investigated the underlying mechanisms of statistical learning and the type of relationship extracted during statistical learning. In several experiments, I used incidental statistical learning where the relationship between object pairs was learned without intentional effort and explicit instruction and I observed behavioral benefits of statistical learning (i.e., faster responses to expected objects). In the following discussion, I will attempt to answer the key questions raised in the introduction based on the presented data in the previous chapters.

Blocking in statistical learning

Blocking is a crucial phenomenon in reinforcement learning, demonstrating the role of prediction error. In forward blocking, the outcome X is first paired with the cue A. Subsequently, the outcome X is paired with the compound cues A and B. Through repeated exposure to A→X, the relationship between cue A and outcome X is learned, minimizing the prediction error. As a result, the association between cue B and outcome X cannot be learned (Rescorla & Wagner, 1972). In backward blocking, the outcome X is initially paired with the compound cues A and B. Later, the outcome X is paired only with cue A. Through repeated exposure to AB→X, the prediction error decreases. When cue A alone precedes outcome X, the associative relationship between cue A and outcome X is updated, further reducing the prediction error. This retrospective update weakens the association between cue B and outcome X (Kalman, 1960; Kruschke, 2008; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994).

In the introduction of my thesis, I raised the question whether statistical learning is based on prediction errors. To answer this question, in **chapter 2** I utilized both forward and backward blocking paradigms of reinforcement learning in the context of statistical learning. In the first experiment, I applied a forward blocking paradigm to the classical statistical learning task. While I did not observe forward blocking, I instead found augmentation of learning. Specifically, I observed that the associative strength between the blocked stimuli and the trailing stimuli (which was hypothesized to become weaker) was learned equally well (or even better) as the association with the antedating stimuli. I speculate that attention may provide a parsimonious explanation for the augmented learning of the blocked stimuli. Studies suggest that stimuli whose consequences are initially unexpected attract more attention (Holland & Schiffino, 2016; Pearce & Hall, 1980b) and that attentional allocation maximizes learning, with observers focusing on stimuli that are neither completely predictable nor unpredictable (Gottlieb et al., 2013; Kidd

et al., 2012; Poli et al., 2020). In our forward blocking experiment, participants learned the association between the antedating leading object and the trailing object during the first training phase. Thus, their attention may have shifted to the novel and therefore potentially more salient blocked leading object in the second phase, thereby enhancing the association between the blocked leading object and the trailing object. On the other hand, in the control condition, two novel leading objects equally competed for associative strength with the trailing object and hence their individual predictive power was reduced (Rescorla & Wagner, 1972). In Experiment 2, I used backward blocking to control the novelty and salience and thereby eliminate this potential attentional effect. Our results indicated that backward blocking occurs in statistical learning, supporting that statistical learning may be error-driven and thus suggesting a functional similarity between statistical learning and reinforcement learning.

Exploring the Role of Prediction Error in Statistical Learning: Insights from Blocking Phenomena

In typical blocking experiments, associations are learned either when the outcome is a reward (Aggarwal et al., 2020; Aggarwal & Wickens, 2020; Sharpe et al., 2017; Steinberg et al., 2013) or when performance-related feedback is provided (Blanco et al., 2014; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Luque et al., 2018; Mitchell et al., 2005, 2006). There are few studies examining forward blocking using incidental learning. For instance, Beesley and Shanks (2012) did not observe forward blocking in contextual cueing experiments, where participants incidentally learned the spatial relationship among distractors and targets in a visual search task. However, this procedure deviates from classic forward blocking paradigms that requires the temporal relationship between cue and outcome (Aggarwal et al., 2020; Aggarwal & Wickens, 2020; Blanco et al., 2014; De Houwer et al., 2005; De Houwer & Beckers, 2003; Kruschke & Blair, 2000; Le Pelley et al., 2005, 2007; Luque et al., 2018; Mitchell et al., 2006; Steinberg et al., 2013; Vandorpe et al., 2005). Two subsequent experiments (Morís et al., 2014; Schmidt & De Houwer, 2019) observed forward blocking of temporal associations only for material that was intentionally learned, but not for incidentally learned stimulus associations. Such learning conditions differ significantly from typical statistical learning scenarios, where observers extract regularities without intention (Batterink et al., 2019; Frost et al., 2019; Sherman et al., 2020; Turk-Browne et al., 2010). To address the paradigm-related issues in these studies, I adapted the classical statistical learning paradigm to the forward blocking paradigm. Despite these improvements, I did not observe

forward blocking in statistical learning, consistent with previous incidental learning studies. While a few studies have investigated forward blocking in incidental learning, less is known about backward blocking in incidental learning. There is evidence of backward blocking in infants (Sobel & Kirkham, 2006, 2007), who clearly did not follow explicit task instructions. To the best of our knowledge, no study has examined and found backward blocking in incidental learning with adults.

The traditional Rescorla-Wagner model which highlight the role of prediction error in learning, explain forward blocking this way: the previously learned the cue A – outcome X association prevents the formation of an associative link between the cue B and the outcome X. This happens because the antedating cue already minimizes the prediction error during the initial exposure to the cue-outcome pairs. However, the Rescorla-Wagner model, which assumes that the relevant cue must be present to change its associative strength (Kruschke, 2008; Miller & Witnauer, 2016; Rescorla & Wagner, 1972), cannot explain backward blocking. However, this phenomenon can be explained by the Van Hamme and Wasserman model, which assigns non-zero salience to the absent cue by assuming its representation can be retrieved by the presentation of a previously competing cue (i.e., the cue A). Another explanation is provided by the Kalman filter, which updates the weights of all possible cues simultaneously (Gershman, 2015; Kalman, 1960; Kruschke, 2008). Thus, these revised models enable to update the associative strength between the cue A – outcome X association, thus leading the weakening of the cue B – outcome X association retrospectively. By showing backward blocking in statistical learning, our results suggest that statistical learning may rely on prediction error as in reinforcement learning.

In addition to the Rescorla-Wagner model and its modified versions, which assume that the associability of events is directly modulated by prediction errors, attentional models also provide an explanation for backward blocking. First, the Pearce-Hall model posits that the attention paid to events is directly influenced by prediction error. In the second phase of backward blocking, where one cue is presented alone, the prediction error is high for this cue because it is now a new predictor of the outcome. Consequently, attention is shifted to this cue. Conversely, the absence of the blocked cue reduces its prediction error and thus its associability. As a result, backward blocking occurs. On the other hand, the Mackintosh model explains backward blocking based solely on attention, without requiring prediction error. It states that attention paid to events is modulated by their predictiveness. In the second phase of backward blocking, the presentation of one cue alone makes it a more reliable predictor of the outcome. Thus, attention shifts towards this cue, and

the reduced attention to the blocked cue decreases its associative strength with the outcome. In both the Pearce-Hall and Mackintosh models, attention plays a crucial role in explaining backward blocking. The difference lies in that the Pearce-Hall model requires prediction errors to determine how attention is allocated to events. Therefore, from the perspective of the Pearce-Hall model but not the Mackintosh model, demonstrating backward blocking in statistical learning suggests that statistical learning is based on prediction error.

Further, in the context of reinforcement learning, some researchers emphasize the role of inferential reasoning for blocking to occur. They argue that the deliberate assessment of causal relationships between cues and outcomes is essential for both forward blocking (De Houwer et al., 2005; De Houwer & Beckers, 2003; Vandorpe et al., 2005) and backward blocking (De Houwer & Beckers, 2002; Waldmann, 2000; Waldmann & Holyoak, 1992). So far, the studies examining blocking in adults clearly instruct their participants to learn the relationship between events. However, in our experiment, participants were not informed of any potential relationship between leading and trailing objects, and they learned the associative relationship incidentally. Therefore, our findings suggest that conscious inferential reasoning is not necessary for backward blocking to occur; instead, backward blocking can occur during incidental statistical learning.

In sum, while I did not find forward blocking, our results are compatible with the presence of backward blocking in statistical learning, a form of learning that develops incidentally and in the absence of rewarding outcomes or feedback. Based on the Van Hamme – Wasserman model, Kalman filter and Pearce-Hall model, our result suggests a functional similarity between statistical learning and reinforcement learning and support the idea that statistical learning may be error-driven.

Unique associations determine statistical learning

In the introduction of my thesis, I raised the questions whether statistical learning is more sensitive to uniqueness rather than conditional probability and which form of uniqueness (i.e., *ΔP* or *DFH*) is learnt during statistical learning. Our results from **chapter 3** speak to these questions. In each study I manipulated the relationship between object pairs such that *CP*, *ΔP* and *DFH* of object pairs can be compared to better understand which one of them governs statistical learning. In Experiment 1, both low and high unique pairs had a *CP* of 1, but participants learned high unique

pairs more effectively. In Experiments 2 and 3, high unique pairs with *CP*s of 0.5 and 1 were learned equally well. These results suggest that during statistical learning, observers learned the relationship between objects based on their unique predictive relationships between events rather than conditional probabilities (Leshinskaya & Thompson-Schill, 2021).

In Experiment 2 and 3, I attempted to identify which form of uniqueness (*ΔP* and *DFH*) governs statistical learning. For this, I varied *ΔP* while keeping *DFH* constant. I found that participants learned pairs with low *ΔP* and high *ΔP* equally well, suggesting that *DFH* may influence statistical learning rather than *ΔP*. To better understand and to replicate the findings, in Experiment 3, I directly made a direct comparison *ΔP* and *DFH* by creating a modulation in opposite directions between the two and found that participants learned pairs with high *DFH* and high *ΔP* equally well. Therefore, our findings remain inconclusive about which forms of uniqueness determines statistical learning.

Therefore, in **chapter 4**, I specifically focused on better understanding which form of uniqueness governs statistical learning. First, I created a stronger contrast between different forms of uniqueness to make a clear comparison between them. Second, to enhance the learning effect, the expectation manipulation was shifted from the temporal to the spatial domain to further enhance the learning effect, as spatial expectations yield stronger behavioral and neural effects (He et al., 2022). However, this approach loses the natural cause-and-effect relationship seen in the temporal domain, which is crucial for uniqueness. To address this, I maintained a temporal relationship by presenting one object earlier than the other. Despite these improvements in our experimental paradigm, I could not observe any learning effect. Although there was a trend of faster responses to expected spatial arrangements in the high *DFH* condition, this was not statistically significant. Additionally, I did not observe such trend in activity pattern of any of our ROIs. These findings suggest that participants did not learn the most likely spatial arrangement of the object images.

Exploring the Role of Uniqueness in Statistical Learning: Insights from ΔP and DFH Metrics

Metrics of uniqueness, particularly *ΔP*, are often used to examine causal reasoning during tasks where participants are explicitly encouraged to learn and make causal judgments about relationships between events (Griffiths & Tenenbaum, 2005). For

instance, the blicket detector paradigm is frequently used to study how individuals intentionally learn relationships to draw accurate causal conclusions (Beckers et al., 2009; Griffiths et al., 2011; Jiang & Lucas, 2024; McCormack et al., 2009; Sobel et al., 2004). In this classic experimental setup, participants are introduced to the blicket machine and asked to identify the blickets (i.e., causes) that activate the machine. Studies using the blicket detector paradigm have shown that causal inferences are influenced by *ΔP* in both children (Sobel et al., 2004) and adults (Griffiths et al., 2011; Jiang & Lucas). Contrasting these studies, where participants were explicitly encouraged to learn causal relationships, the effects of causal learning were observed in infants (Sobel & Kirkham, 2006, 2007), who did not follow explicit task instructions but instead attuned themselves to statistical regularities through passive exposure. This suggests that statistical learning may be sensitive to unique predictive relationships rather than conditional probabilities, as indicated by previous studies (Fiser & Aslin, 2002).

Leshinskaya and Thompson-Schill (2021) demonstrated that participants failed to learn the relationship between events with high conditional probabilities when the ΔP between them was low. This suggests that statistical learning is governed by unique predictive relationships defined by ΔP. Similarly, two subsequent experiments found that observers learned relationships between events with high conditional probabilities as well as those with low conditional probabilities in both monkeys (Ramachandran et al., 2016) and humans (Richter et al., 2018), due to the same value of *DFH*, further supporting the role of unique predictive relationships in statistical learning.

In a series of behavioral and fMRI experiments described in **chapters 3** and **4**, I aimed to better understand whether statistical learning is governed by uniqueness and which metrics of uniqueness determine statistical learning. Our results provide further evidence that unique predictive relationships govern statistical learning. However, our findings remain inconclusive about the specific type of unique relationship that determines learning. We discuss potential limitations of our experimental setups that may have caused these inconclusive results and suggest improvements for future research in the following sections.

Limitations

In **chapter 2** and **chapter 3**, participants performed the experiments online. Due to the nature of online data collection where the commitment of participants can be lower than in traditional laboratory settings (Al-Salom & Miller, 2019; Brühlmann et al., 2020), the rejection rates were high in these experiments. This may raise concerns about the generalizability of our results. Based on a simple task where the general population is expected to perform at ceiling levels, I excluded participates who underperformed, likely due to not reading instructions carefully, not understanding the task, or not paying enough attention. Such exclusions are common in online experiments, where about half of the participants often exhibit careless and inattentive behavior (Al-Salom & Miller, 2019; Brühlmann et al., 2020). Therefore, the consequences of statistical learning appear to be limited to a subset of participants who demonstrated high motivation and adequate attention to the stimuli, which is essential for supporting statistical learning (Richter & de Lange, 2019).

There are also limitations related to the uniqueness manipulation in **chapter 3** and **chapter 4**. First, *ΔP* and *DFH* share the same components, meaning that strengthening or weakening one affects the other. This prevented us from finding optimal conditions with a strong enough contrast between *ΔP* and *DFH* to make a clear comparison. Second, the amount of exposure to objects differed between the *ΔP* and *DFH* conditions. There were more trials in the high *ΔP* conditions than in the high *DFH* conditions, making participants more familiar with *ΔP* conditions. It is known that familiar stimuli lead to stronger and faster processing and improved behavioral performance (Fritsche et al., 2020; Manahova et al., 2020). Therefore, this imbalance in familiarity may have confounded our results, potentially reducing the observed effects due to the overall faster processing of object pairs in the *ΔP* conditions.

Future directions

The experiments in chapters 3 and 4 demonstrated that statistical learning is governed not by the leading object strongly predicting the trailing object, but by the leading object uniquely and independently predicting the trailing object. However, these experiments did not clarify what type of uniqueness determines learning. Future studies could use a modified version of the paradigm from chapters 3 and 4 to address this question. To obtain more insightful results, future studies could improve the design matrix to create a stronger contrast between *ΔP* and

DFH conditions. In experiment 2 of chapter 3, *ΔP* was varied while *DFH* was kept constant. Reversing this design might show the role of *DFH* independently from *ΔP*.

Additionally, to investigate the role of uniqueness in the temporal domain, the paradigm from chapter 3 could be improved by changing the location of object presentation. In chapter 3, objects were presented on the left or right side of the screen for consistency with chapter 2. However, supplementary experiment 1 of the blocking project highlighted the importance of stimulus location in learning. This experiment presented leading and trailing objects at the center of the screen, resulting in stronger associations between objects and a reaction time benefit approximately twice as large as in other experiments. Therefore, future studies could present leading and trailing objects at the center to boost learning and thereby obtain stronger results.

To better understand the role of uniqueness in the spatial domain, the paradigm from chapter 4 could also be improved. In chapter 4, I swapped the spatial locations of object pairs to create unexpected spatial arrangements. However, participants may have focused on the identity of the objects rather than their spatial arrangement (He et al., 2022). Future studies could create unexpected spatial arrangements by changing the identity of the objects instead of swapping their locations.

Conclusions

In conclusion, throughout this thesis I sought to unravel the mechanisms of statistical learning and the types of relationships that influence it. First, I investigated whether statistical learning is error-driven by employing both forward and backward blocking paradigms. Our forward blocking experiment did not show the expected blocking effect but rather an augmentation of learning, likely due to attentional shifts towards novel stimuli. However, backward blocking was observed, supporting the idea that statistical learning might be error-driven and thus indicating a functional similarity between statistical learning and reinforcement learning. Second, I examined whether statistical learning is more sensitive to unique predictive relationships rather than conditional probabilities. Our results indicated that participants learned object pairs based on unique predictive relationships. However, our experiments remain inconclusive about which specific form of uniqueness governs statistical learning. Our findings contribute to the understanding of statistical learning by suggesting that it may be error-driven, akin to reinforcement learning. However, further research is needed to clarify the specific metrics of uniqueness that drive statistical learning.

References

Aggarwal, M., Akamine, Y., Liu, A. W., & Wickens, J. R. (2020). The nucleus accumbens and inhibition in the ventral tegmental area play a causal role in the Kamin blocking effect. *The European Journal of Neuroscience*, *52*(3), 3087–3109. https://doi.org/10.1111/ejn.14732

Aggarwal, M., & Wickens, J. R. (2020). The Kamin Blocking Effect in Sign and Goal Trackers. *bioRxiv*.

Agliari, E., Aquaro, M., Barra, A., Fachechi, A., & Marullo, C. (2023). From Pavlov Conditioning to Hebb Learning. *Neural Computation*, *35*(5), 930–957. https://doi.org/10.1162/neco_a_01578

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *34*(1), 1–11. https://doi.org/10.1037/h0081013

Al-Salom, P., & Miller, C. J. (2019). The Problem with Online Data Collection: Predicting Invalid Responding in Undergraduate Samples. *Current Psychology*, *38*(5), 1258–1264. https://doi.org/10.1007/s12144-017-9674-9

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Batson, J. D., & Batsell Jr., W. R. (2000). Augmentation, not blocking, in an A+/AX+ flavor-conditionaing procedure. *Psychonomic Bulletin & Review*, *7*(3), 466–471. https://doi.org/10.3758/BF03214358

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*, *11*(3), 482–503. https://doi.org/10.1111/tops.12420

Beckers, T., Vandorpe, S., Debeys, I., & De Houwer, J. (2009). Three-year-olds' retrospective revaluation in the blicket detector task. Backward blocking or recovery from overshadowing? *Experimental Psychology*, *56*(1), 27–32. https://doi.org/10.1027/1618-3169.56.1.27

Beesley, T., & Shanks, D. R. (2012). Investigating cue competition in contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 709–725. https://doi.org/10.1037/a0024885

Béghin, G., Gagnon-St-Pierre, É., & Markovits, H. (2021). A dual strategy account of individual differences in information processing in contingency judgments. *Journal of Cognitive Psychology*, *33*(4), 470–481. https://doi.org/10.1080/20445911.2021.1900200

Blanco, F., Baeyens, F., & Beckers, T. (2014). Blocking in human causal learning is affected by outcome assumptions manipulated through causal structure. *Learning & Behavior*, *42*(2), 185–199. https://doi.org/10.3758/s13420-014-0137-y

Boddez, Y., Haesen, K., Baeyens, F., & Beckers, T. (2014). Selectivity in associative learning: A cognitive stage framework for blocking and cue competition phenomena. *Frontiers in Psychology*, *5*, 1305. https://doi.org/10.3389/fpsyg.2014.01305

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, *2*, 100022. https://doi.org/10.1016/j.metip.2020.100022

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Cheng, P. (1997). From Covariation to Causation: A Causal Power Theory. *Psychological Review*, *104*, 367–405. https://doi.org/10.1037//0033-295X.104.2.367

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365–382. https://doi.org/10.1037/0033-295x.99.2.365

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., Robbins, T. W., Bullmore, E. T., & Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*, *44*(5), 877–888. https://doi.org/10.1016/j.neuron.2004.11.022

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *55*(4), 289–310. https://doi.org/10.1080/02724990244000034

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *The Quarterly Journal of Experimental Psychology Section B*, *56*(4b), 345–357.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(3), 965–985. https://doi.org/10.1080/02724980143000578

De Houwer, J., Vandorpe, S., & Beckers, T. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior*, *33*(2), 239–249. https://doi.org/10.3758/BF03196066

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*(5), 1175–1185. https://doi.org/10.1093/cercor/bhn161

Ferrari, A., Richter, D., & de Lange, F. P. (2022). Updating Contextual Sensory Expectations for Adaptive Behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(47), 8855–8869. https://doi.org/10.1523/JNEUROSCI.1107-22.2022

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037//0278-7393.28.3.458

Fiser, J., & Lengyel, G. (2019). A common probabilistic framework for perceptual and statistical learning. *Current Opinion in Neurobiology*, *58*, 218–228. https://doi.org/10.1016/j.conb.2019.09.007

Fiser, J., & Lengyel, G. (2022). Statistical Learning in Vision. *Annual Review of Vision Science*, *8*, 265–290. https://doi.org/10.1146/annurev-vision-100720-103343

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Fritsche, M., Lawrence, S. J. D., & de Lange, F. P. (2020). Temporal tuning of repetition suppression across the visual cortex. *Journal of Neurophysiology*, *123*(1), 224–233. https://doi.org/10.1152/jn.00582.2019

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Gershman, S. J. (2015). A Unifying Probabilistic View of Associative Learning. *PLOS Computational Biology*, *11*(11), e1004567. https://doi.org/10.1371/journal.pcbi.1004567

Gershman, S. J. (2017). Context-dependent learning and causal structure. *Psychonomic Bulletin & Review*, *24*(2), 557–565. https://doi.org/10.3758/s13423-016-1110-x

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758–770. https://doi.org/10.1038/s41583-018-0078-0

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*(8), 1407–1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384. https://doi.org/10.1016/j.cogpsych.2005.05.004

Hasson, U. (2017). The neurobiology of uncertainty: Implications for statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160048. https://doi.org/10.1098/rstb.2016.0048

Hattori, I., Hattori, M., Over, D. E., Takahashi, T., & Baratgin, J. (2017). Dual frames for causal induction: The normative and the heuristic. *Thinking & Reasoning*, *23*(3), 292–317. https://doi.org/10.1080/13546783.2017.1316314

Hattori, M., & Oaksford, M. (2007). Adaptive Non-Interventional Heuristics for Covariation Detection in Causal Induction: Model Comparison and Rational Analysis. *Cognitive Science*, *31*(5), 765–814. https://doi.org/10.1080/03640210701530755

He, T., Richter, D., Wang, Z., & de Lange, F. P. (2022). Spatial and Temporal Context Jointly Modulate the Sensory Response within the Ventral Visual Stream. *Journal of Cognitive Neuroscience*, *34*(2), 332–347. https://doi.org/10.1162/jocn_a_01792

Hebb, D. O. (1949). *The Organization of Behaviour (Hoboken, NJ*. Wiley.

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8), eabc4530. https://doi.org/10.1126/sciadv.abc4530

Holland, P. C., & Schiffino, F. L. (2016). Mini-Review: Prediction errors, attention and associative learning. *Neurobiology of Learning and Memory*, *131*, 207–215. https://doi.org/10.1016/j.nlm.2016.02.014

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680. https://doi.org/10.1037/0096-3445.130.4.658

Jiang, C., & Lucas, C. G. (2024). Actively Learning to Learn Causal Relationships. *Computational Brain & Behavior*, *7*(1), 80–105. https://doi.org/10.1007/s42113-023-00195-0

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In ba campbell & rm church (eds.), Punishment and aversive behavior (pp. 279-296). *New York: Appleton-Century-Crofts*.

Kaposvari, P., Kumar, S., & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex*, *28*(1), 250–266.

Keiflin, R., Pribut, H. J., Shah, N. B., & Janak, P. H. (2019). Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. *Current Biology: CB*, *29*(1), 93-103.e3. https://doi.org/10.1016/j.cub.2018.11.050

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Klein-Flügge, M. C., Wittmann, M. K., Shpektor, A., Jensen, D. E. A., & Rushworth, M. F. S. (2019). Multiple associative structures created by reinforcement and incidental statistical learning mechanisms. *Nature Communications*, *10*(1), 4835. https://doi.org/10.1038/s41467-019-12557-z

Kruschke, J. K. (2001). Toward a Unified Model of Attention in Associative Learning. *Journal of Mathematical Psychology*, *45*(6), 812–863. https://doi.org/10.1006/jmps.2000.1354

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior*, *36*(3), 210–226. https://doi.org/10.3758/LB.36.3.210

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636–645. https://doi.org/10.3758/bf03213001

Le Pelley, M. E., Beesley, T., & Suret, M. B. (2007). Blocking of human causal learning involves learned changes in stimulus processing. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1468–1476.

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 56.

Leshinskaya, A., & Thompson-Schill, S. L. (2021). *Statistical learning reflects inferences about unique predictive relations*. PsyArXiv. https://doi.org/10.31234/osf.io/c3jpn

Luque, D., Flores, A., & Vadillo, M. A. (2013). Revisiting the role of within-compound associations in cue-interaction phenomena. *Learning & Behavior*, *41*(1), 61–76. https://doi.org/10.3758/s13420-012-0085-3

Luque, D., Vadillo, M. A., Gutiérrez-Cobo, M. J., & Le Pelley, M. E. (2018). The blocking effect in associative learning involves learned biases in rapid attentional capture. *The Quarterly Journal of Experimental Psychology*, *71*, 522–544. https://doi.org/10.1080/17470218.2016.1262435

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298. https://doi.org/10.1037/h0076778

Manahova, M. E., Spaak, E., & de Lange, F. P. (2020). Familiarity Increases Processing Speed in the Visual System. *Journal of Cognitive Neuroscience*, *32*(4), 722–733. https://doi.org/10.1162/jocn_a_01507

Markovits, H., Forgues, H. L., & Brunet, M.-L. (2012). More evidence for a dual-process model of conditional reasoning. *Memory & Cognition*, *40*(5), 736–747. https://doi.org/10.3758/s13421-012-0186-4

Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, *6*, 888. https://doi.org/10.3389/fpsyg.2015.00888

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–346.

McCormack, T., Butterfill, S., Hoerl, C., & Burns, P. (2009). Cue competition effects and young children's causal and counterfactual inferences. *Developmental Psychology*, *45*(6), 1563–1575. https://doi.org/10.1037/a0017408

McCormack, T., Simms, V., McGourty, J., & Beckers, T. (2013). Blocking in children's causal learning depends on working memory and reasoning abilities. *Journal of Experimental Child Psychology*, *115*(3), 562–569. https://doi.org/10.1016/j.je*CP*.2012.11.016

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*(3), 211–246. https://doi.org/10.3758/BF03200258

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*(1), 25–53. https://doi.org/10.1080/02724990344000042

Melchers, K., Lachnit, H., & Shanks, D. R. (2006). The comparator theory fails to account for the selective role of within-compound associations in cue-selection effects. *Experimental Psychology*, *53*(4), 316–320. https://doi.org/10.1027/1618-3169.53.4.316

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Miller, R. R., & Matzel, L. D. (1988). The Comparator Hypothesis: A Response Rule for The Expression of Associations. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 22, pp. 51–92). Academic Press. https://doi.org/10.1016/S0079-7421(08)60038-9

Miller, R. R., & Witnauer, J. E. (2016). Retrospective Revaluation: The Phenomenon and Its Theoretical Implications. *Behavioural Processes*, *123*, 15–25. https://doi.org/10.1016/j.beproc.2015.09.001

Mitchell, C. J., Killedar, A., & Lovibond, P. F. (2005). Inference-based retrospective revaluation in human causal judgments requires knowledge of within-compound relationships. *Journal of Experimental Psychology. Animal Behavior Processes*, *31*(4), 418–424. https://doi.org/10.1037/0097-7403.31.4.418

Mitchell, C. J., Lovibond, P. F., Minard, E., & Lavis, Y. (2006). Forward blocking in human learning sometimes reflects the failure to encode a cue–outcome relationship. *Quarterly Journal of Experimental Psychology*, *59*(5), 830–844.

Morís, J., Cobos, P. L., Luque, D., & López, F. J. (2014). Associative repetition priming as a measure of human contingency learning: Evidence of forward and backward blocking. *Journal of Experimental Psychology: General*, *143*(1), 77.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(1), 80–96. https://doi.org/10.1037/0278-7393.34.1.80

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (pp. xv, 430). Oxford Univ. Press.

Pearce, J. M., & Hall, G. (1980a). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pearce, J. M., & Hall, G. (1980b). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Ramachandran, S., Meyer, T., & Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *Journal of Neurophysiology*, *115*(1), 355–362. https://doi.org/10.1152/jn.00091.2015

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. *Classical Conditioning: Current Research and Theory*.

Richter, D. (2021). *Prediction throughout visual cortex: How statistical regularities shape sensory processing* [Radboud University]. https://hdl.handle.net/2066/230613

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *eLife*, *8*, e47869. https://doi.org/10.7554/eLife.47869

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *The Journal of Neuroscience*, *38*(34), 7452–7461.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *The European Journal of Neuroscience*, *35*(7), 1190–1200. https://doi.org/10.1111/j.1460-9568.2011.07986.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

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. https://doi.org/10.1016/S0010-0277(98)00075-4

Schmidt, J. R., & De Houwer, J. (2019). Cue Competition and Incidental Learning: No Blocking or Overshadowing in the Colour-Word Contingency Learning Procedure Without Instructions to Learn. *Collabra: Psychology*, *5*(1), 15. https://doi.org/10.1525/collabra.236

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Shanks, D. R. (1985). Forward and Backward Blocking in Human Contingency Judgement. *The Quarterly Journal of Experimental Psychology Section B*, *37*(1b), 1–21. https://doi.org/10.1080/14640748508402082

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, *61*, 273–301. https://doi.org/10.1146/annurev.psych.093008.100519

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., Niv, Y., & Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), 735–742. https://doi.org/10.1038/nn.4538

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20. https://doi.org/10.1016/j.cobeha.2020.01.015

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208-219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, *10*(3), 298–306. https://doi.org/10.1111/j.1467-7687.2007.00589.x

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303–333. https://doi.org/10.1207/s15516709cog2803_1

Spicer, S., Wills, A., Jones, P., Mitchell, C., & Dome, L. (2021). Representing uncertainty in the Rescorla-Wagner model: Blocking, the redundancy effect, and outcome base rate. *Open Journal of Experimental Psychology and Neuroscience*, 14–21. https://doi.org/10.46221/ojepn.2021.6623

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966–973.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Vadillo, M. A., & Matute, H. (2010). Augmentation in contingency learning under time pressure. *British Journal of Psychology*, *101*(3), 579–589. https://doi.org/10.1348/000712609X477566

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151. https://doi.org/10.1006/lmot.1994.1008

Vandorpe, S., De Houwer, J., & Beckers, T. (2005). Further evidence for the role of inferential reasoning in forward blocking. *Memory & Cognition*, *33*(6), 1047–1056.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*(3), 239–278. https://doi.org/10.1080/13546780442000178

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(1), 53–76. https://doi.org/10.1037//0278-7393.26.1.53

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology. General*, *121*(2), 222–236. https://doi.org/10.1037//0096-3445.121.2.222

Nederlandse samenvatting

Leren stelt ons in staat ons begrip van de wereld te ontwikkelen en te verfijnen door associaties te vormen tussen systematisch gerelateerde gebeurtenissen in onze omgeving. Deze associaties, afgeleid van regelmatigheden in ruimte en tijd, helpen ons toekomstige gebeurtenissen te voorspellen, reacties voor te bereiden en ons aan te passen aan onze omgeving. Dit automatische en vaak onbewuste proces, bekend als statistisch leren, stelt ons in staat deze regelmatigheden te herkennen door middel van meerdere blootstellingen zonder de intentie of moeite om te leren. Statistisch leren verbetert de informatieverwerking, wat resulteert in snellere en nauwkeurigere reacties op verwachte stimuli en omvat onderdrukte neurale reacties op verwachte stimuli.

In Hoofdstuk 2 onderzochten we welk leermechanisme ten grondslag ligt aan statistisch leren en of statistisch leren foutgestuurd is zoals bekrachtigingsleren. Om dit te doen, leende ik de voorwaartse en achterwaartse blokkering paradigma's van bekrachtigingsleren. In het voorwaartse blokkering paradigma wordt aanvankelijk uitkomst X gekoppeld aan cue A. Later wordt een nieuwe cue B geïntroduceerd naast cue A, die beide leiden tot dezelfde uitkomst X. De eerder vastgestelde associatie tussen cue A en uitkomst X voorkomt het vormen van een nieuwe associatie tussen cue B en uitkomst X omdat cue A de voorspellingsfout tijdens de eerste blootstelling al minimaliseert. Het enige verschil in het achterwaartse blokkering paradigma is dat uitkomst X eerst wordt gekoppeld aan zowel cue A als cue B en later X alleen wordt gekoppeld aan cue A. Het koppelen van cue A met uitkomst X vermindert de voorspellingsfout, wat op zijn beurt de associatie tussen cue B en uitkomst X verzwakt. We pasten klassieke statistische leertaken aan naar blokkering paradigma's in een reeks online experimenten. Bij elke proef kregen de deelnemers paren afbeeldingen te zien. Zonder dat zij dit wisten, waren bepaalde afbeeldingen voorspellers van andere, zodat elke volgende afbeelding kon worden voorspeld op basis van de voorafgaande afbeelding. Na het leren evalueerden we statistisch leren door deelnemers verwachte en onverwachte afbeeldingsparen te tonen en hun reactietijd te meten voor categorisatiebeoordelingen van de volgende afbeelding als elektronisch of niet-elektronisch. In het eerste experiment met een voorwaarts blokkering paradigma, observeerden we geen voorwaartse blokkering maar eerder een versterking van leren. In tegenstelling tot de verwachtingen werden de geblokkeerde stimuli net zo effectief geleerd als de aanvankelijk gepresenteerde stimuli. Dit onverwachte resultaat kan worden verklaard door aandacht mechanismen. In het voorwaartse blokkering experiment verschoof de aandacht van de deelnemers waarschijnlijk naar de nieuwe geblokkeerde stimuli

tijdens de tweede fase, wat hun leren verbeterde. Om de nieuwigheid te beheersen en dit aandachtseffect te elimineren, werd in het tweede experiment achterwaartse blokkering toegepast. De resultaten bevestigden achterwaartse blokkering in statistisch leren, wat wijst op een functionele gelijkenis tussen statistisch en bekrachtigingsleren en ondersteunt het idee dat statistisch leren kan worden aangedreven door voorspellingsfouten.

In Hoofdstuk 3 onderzochten we welke soorten associatieve relaties worden geëxtraheerd tijdens statistisch leren en welke metrics hun extractie sturen. In statistische leerstudies is de belangrijkste metric die wordt gebruikt om leren te moduleren de conditionele waarschijnlijkheid. Echter, causale leerstudies tonen aan dat leren plaatsvindt op basis van de unieke voorspellende relatie in plaats van conditionele waarschijnlijkheid. Er zijn twee verschillende vormen van uniciteit: meer rationeel en analytisch *ΔP* en sneller en heuristisch *DFH*. Om te onderzoeken of statistisch leren gevoeliger is voor uniciteit of conditionele waarschijnlijkheid en welke vormen van uniciteit (*ΔP* of *DFH*) worden geleerd, voerden we een reeks online experimenten uit waarbij paren afbeeldingen werden gepresenteerd met een relatie, zodat elke volgende afbeelding kon worden voorspeld op basis van de voorafgaande afbeelding. Na het leren evalueerden we statistisch leren door deelnemers verwachte en onverwachte afbeeldingsparen te tonen en hun reactietijd te meten voor categorisatiebeoordelingen van de volgende afbeelding. Experiment 1 toonde aan dat deelnemers hoog unieke paren effectiever leerden dan laag unieke paren, ondanks dat beide een conditionele waarschijnlijkheid van 1 hadden. Experimenten 2 en 3 vonden dat hoog unieke paren met *CP*'s van 0,5 en 1 even goed werden geleerd, wat aangeeft dat unieke voorspellende relaties statistisch leren meer aandrijven dan conditionele waarschijnlijkheden. In Experiment 2, bij het specifiek onderzoeken van *ΔP* en *DFH*, leidde variërende *ΔP* terwijl *DFH* constant bleef tot gelijk leren van lage en hoge *ΔP* paren, wat suggereert dat *DFH* mogelijk een grotere rol speelt. Echter, een directe vergelijking in Experiment 3 toonde aan dat paren met hoge *DFH* en hoge *ΔP* even goed werden geleerd. Dus, onze bevindingen zijn niet eenduidig over welke vorm van uniciteit voornamelijk invloed heeft op statistisch leren.

In Hoofdstuk 4 wilden we de soort uniciteit identificeren die statistisch leren bepaalt. Om dit te bereiken, verhoogden we het contrast tussen verschillende vormen van uniciteit en verschoven we de verwachting manipulatie van het temporele naar het ruimtelijke domein om het leereffect te versterken. Onze aanpak omvatte zowel gedrags experimenten uitgevoerd in een gecontroleerde labomgeving als fMRI experimenten. Ondanks deze methodologische verbeteringen observeerden

we geen significant leereffect. Hoewel er een trend was naar snellere reacties op verwachte ruimtelijke rangschikkingen in de hoge *DFH* conditie, was dit niet statistisch significant. Bovendien werden er geen significante trends gevonden in de activiteits patronen van onze ROIs. Deze resultaten geven aan dat deelnemers niet de meest waarschijnlijke ruimtelijke rangschikking van de object afbeeldingen leerden.

In mijn dissertatie onderzocht ik de mechanismen van statistisch leren en de relaties die het beïnvloeden. De eerste reeks experimenten onderzocht of statistisch leren foutgestuurd is, gebruikmakend van zowel voorwaartse als achterwaartse blokkering paradigma's. Hoewel voorwaartse blokkering niet het verwachte blokkeringseffect toonde, werd achterwaartse blokkering waargenomen, wat wijst op een functionele gelijkenis tussen statistisch leren en bekrachtigingsleren, en ondersteunt het idee dat statistisch leren foutgestuurd kan zijn. De tweede reeks experimenten onderzocht of statistisch leren gevoeliger is voor unieke voorspellende relaties dan voor conditionele waarschijnlijkheden. De resultaten gaven aan dat deelnemers object paren leerden op basis van unieke voorspellende relaties, hoewel de specifieke vorm van uniciteit onduidelijk blijft. Over het algemeen suggereren de bevindingen dat statistisch leren foutgestuurd kan zijn, maar verder onderzoek is nodig om de precieze metrics van uniciteit die het aandrijven te bepalen.

Acknowledgements

Completing my PhD would not have been possible without the support, advice, and help of many people. I am deeply grateful to all of you.

First and foremost, I would like to express my deepest gratitude to Floris. The existence of this thesis is primarily due to your guidance. I have learned so much from your extensive scientific knowledge, tireless enthusiasm, and genuine interest. You have been the ideal supervisor, supporting and guiding me every step of the way. It was an honor to work with you and to be a part of the Predictive Brain Lab family.

Ambra, thank you for your invaluable help and support. From you, I have learned essential skills in running experiments and meticulously writing manuscripts.

Christoph, I am deeply thankful for your guidance on GLMMs and your unwavering support whenever I needed it.

Matthias, your involvement in my last research project was crucial. Thank you for teaching me the intricate details of running and analyzing fMRI experiments. I hope I can recall your explanations during my defense. Completing my final research project and finishing my PhD would have been impossible without you.

I also want to thank all my current and former lab members at the Predictive Brain Lab: Alya, Ashley, Britta, Chuyao, Claire, David, Dota, Eelke, Ella, Eva, Floortje, Ingmar, Jakub, Joey, Judit, Lea, Kim, Lieke, Maartje, Mandy, Marisha, Micha, Yamil, and Qifei. Thank you all for the engaging meetings, motivational and emotional support, and all the fun times.

Paul, thank you for teaching me so much about MRI and for all your support in data acquisition.

Finally, I would like to thank my mom, dad, and best friends for their unconditional love and emotional support during the tough times. The comfort, confidence, and happiness you provide are indescribable. Having you by my side has been invaluable.

About the author

Ilayda Nazli was born on April 23, 1994, in Malatya, Türkiye. She earned her bachelor's degree in Psychology from Middle East Technical University in 2017 and her master's degree in Psychology from Ihsan Dogramaci Bilkent University in 2019. In her master's thesis, she investigated the effect of orientation-related prior probability information on contrast perception. In 2019, she received a scholarship from the Ministry of Education in Türkiye. A year later, she joined the Predictive Brain Lab at the Donders Institute to conduct the research presented in this dissertation.

Research data management

This research followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles. The paragraphs below specify in detail how this was achieved.

Ethics

This thesis is based on the results of human studies, which were conducted following the principles of the Declaration of Helsinki and were approved by the local ethics committee (CMO region Arnhem-Nijmegen, The Netherlands; CMO2014/288). The research was funded by ERC Starting Grant 101000942 "Surprise", awarded to Floris P. de Lange and by the Ministry of National Education of the Republic of Türkiye, awarded to Ilayda Nazli.

Findable and Accessible

The table below details where the data and research documentation for each chapter can be found. All data archived as a Data Sharing Collection remain available for at least 10 years after the termination of the studies.

Chapter	DAC	DSC	License
2	DAC_3018054.01_887	DSC_3018054.01_700	RU-DI-HD-1.0
3	DAC_3018054.02_151	-	
4	DAC_3018054.03_926	-	

DAC = Data Acquisition Collection, DSC = Data Sharing Collection

Informed consent was obtained on paper or online form following the Centre procedure. The forms are archived in the central archive of the Centre for 10 years after the termination of the studies.

Interoperable and Reusable

The raw data are stored in the DAC and DSC in their original form. Data remains usable in the future by using long-lived files (e.g. .nii, .csv, .txt) to improve data access and reusability for the DAC and DSC. Results reported in this thesis are reproducible following the provided description of the experimental setup, raw data and analysis scripts in the DAC and DSC.

Privacy

The privacy of the participants in this thesis has been warranted using random individual subject codes. A pseudonymization key linked this random code with the personal data. This pseudonymization key was stored on a network drive that was only accessible to members of the project who needed access to it because of their role within the project. The pseudonymization key was stored separately from the research data.

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: http://www.ru.nl/donders/graduate-school/phd/

DONDERS INSTITUTE

Radboud University Radboudumc