Université Catholique de Louvain

Unconscious learning: behavioral evidence, relationship with attention and physiological markers

President- prof. Marcus Missal (UCL)
Promoter - prof. André Mouraux (UCL)
Copromoter - dr. Alexandre Zénon (UCL)
Committee member - dr. Jean Jacques Orban de Xivry (KUL)
Committee member – prof. Wim Vanduffel (KUL)
Committee member – prof. Floris de Lange (RU - Nijmegen)

Andrea Alamia
11/09/2017

To Etienne - a mentor, a supervisor, a guide I'm terribly missing.

To Alex - for being an awesome person and a brilliant scientist.

To my family - for being there no matter what.

Table of Contents

Overview

The main objective of this work is to demonstrate that the brain processes information unconsciously, how this affects visual attention and eye movements, and whether it is possible to track unconscious learning by means of physiological recordings (i.e. pupil size).

The first part of the work, which introduces the main topics, combines two sections: the first one is about unconscious learning, whereas the second one is a short overview regarding pupillary responses. In the first part, a short introduction about unconscious systems and a brief historical overview are followed by a survey about the main methodological issues related to the study of unconscious processes. These issues, both methodological and conceptual, have led few skeptics to question the very existence of unconscious processes, and are further discussed through a critical overview of the main paradigms used in the study of unconscious learning. Lastly, the most influential models of consciousness are introduced and considered in the light of the main experimental paradigms presented in the previous section. The second part of the introduction regards the main physiological measure considered in the last study: the pupil size. A very brief overview about the physiological mechanisms involved in driving the pupillary responses is followed by an overview of the cognitive processes that elicit a change in pupil size. Lastly, a brief synopsis of few studies about pupillary response and unconscious processing is outlined.

The second part of the thesis is composed of 4 studies (2 published and 2 in preparation). The first study (chapter II) addresses the criticisms raised against the existence of unconscious processing, providing robust evidence in favor of its existence. In a simple perceptual decision making task, participants unconsciously learn a simple association between the color of the stimuli and its motion direction. Accurate and comprehensive measures of awareness revealed that participants are truly unaware of such association, despite its effect on behavior during the task (i.e. accuracy and

RT). The studies presented in chapter III and IV investigate the relationship between unconscious learning and visual attention, as measured by means of eye movements. The second study employs the same perceptual paradigm introduced in the first study, in which we report robust unconscious learning, whereas the third study investigate the relationship between eye movements and unconscious knowledge in a statistical learning framework. Both studies report a small but significant effect suggesting that unconscious information affects visual attention, and eye movements. The last study (chapter V) dives in the search for physiological measures able to track unconscious processing. In the domain of implicit statistical learning, this last study, composed of 4 experiments, shows that 1) pupil dilates in response to unconsciously surprising events, 2) attention plays a role in such response, 3) it correlates the amount of awareness of the statistical rules with the pupillary response; 4) it reveals the neural correlates of such study by electrophysiological measures (i.e. EEG). The last experiment provides robust evidence (both in the time and in the time-frequency domain) in favor of the hypothesis that pupil tracks prediction error, even when unconsciously, opening the doors for further hypotheses and future studies.

The manuscript is concluded by a short discussion summarizing all the findings and suggesting further perspectives. Particularly, the hypothesis that pupillary response tracks precision-weighted prediction errors is discussed more in detail.

Chapter I: Introduction

Unconscious processing

Unconscious processing is a phenomenon that occurs in our everyday life, influencing several aspects of our behavior and choices. For example, the learning of our mother tongue is deeply rooted in unconscious mechanisms. Despite the existence of rigorous lexical and grammatical rules, a native speaker may have trouble in explaining how he formulates his speech in his native language. In Italian, for example, there are two different articles for masculine words, i.e. 'il' and 'lo'. Even though precise rules about how to use these articles exist (based on the graphemic neighboring, i.e. 'il' is used with words starting with certain consonants, 'lo' with words starting with a combination of two consonants), common native Italian speakers would fail in reporting them, while not making any mistakes in ecological situation of conversation. In other words, they exploit these rules unconsciously. In the following paragraph, a taxonomy of unconscious processes is briefly outlined, based on the current literature.

A taxonomy of unconscious processes

The conscious-unconscious dichotomy has always played an important role in modern psychology, both in the scientific community and in the clinical domain. Since Freud, many scientists have tried to unveil the mental processes that escape the scrutiny of consciousness (or awareness: for sake of simplicity, in the following the terms 'awareness' and 'consciousness' will be used interchangeably). To this day, the literature about unconscious processing is extremely vast, and it embraces a plethora of different approaches and methods that have been developed in neuropsychology and cognitive sciences over the last decades. The

ambitious objective of framing the massive literature about unconscious processing can be undertaken following different perspectives.

One approach may focus on the important distinction between learning and memory: on the one hand the processes in which new knowledge is acquired (i.e. learning), on the other hand the storing and retrieval of the acquired knowledge (i.e. memory). In the unconscious domain, the two processes can be labeled as implicit learning (Ashworth, 1998; Reber, 1967a) and implicit memory (Tulving & Schacter, 1990): in the first case participants are not aware that they are acquiring new information about the environment, in the second case participants are not aware of using information they have previously learnt.

Another way of framing the studies about unconscious processes points directly at the dichotomy between declarative and non-declarative knowledge (fig.1), and their associated brain region.

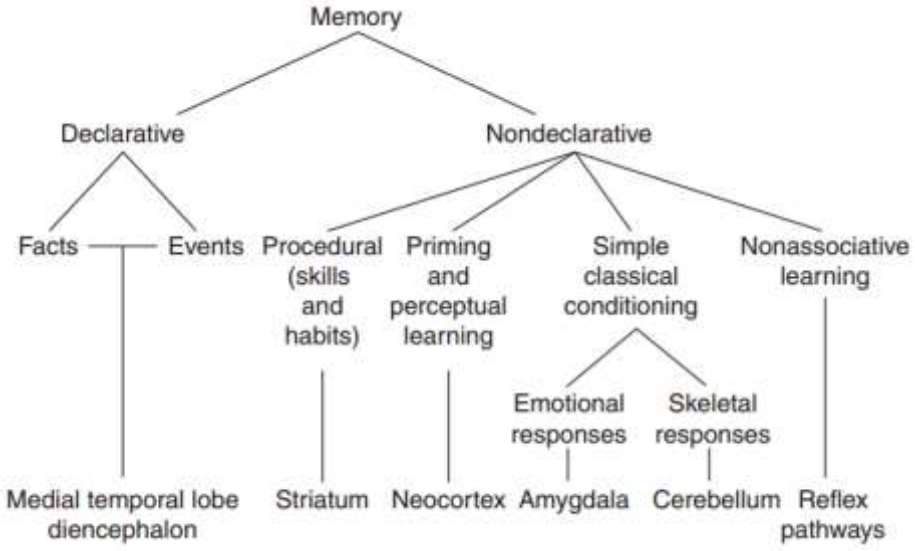


Figure 1 – Memory systems (Squire & Dede, 2015)

At first, it may be tempting to associate the labels ‘explicit’ and ‘implicit’ respectively to the declarative and non-declarative categories: as a matter of fact the review by Squire and Dede (Squire & Dede, 2015) states the definition of declarative knowledge as: “knowledge available for conscious recollection”, and points at the medial temporal lobe for the corresponding neuronal correlate; conversely, non-declarative knowledge can be further separated in procedural, priming and perceptual learning, conditioning and non-associative learning: each one of this category can be related to a specific brain region which is mainly responsible for each type of learning (see fig.1).

Another attempt to categorize the learning processes has been proposed by Katharina Henke (Henke, 2010), who suggested an approach free from the conscious/unconscious dichotomy, but rather based on the concept of processing modes. This original model distinguishes between three modes: rapid encoding of flexible associations, slow encoding of rigid associations and rapid encoding of single or unitized items. In other words, three variables play a key role in such taxonomy: speed of encoding (rapid versus slow), associative versus single-item encoding, and flexible versus rigid representation. The lack of the “consciousness variable” in the model is mostly motivated by the observation that awareness neither plays a discriminatory roles between modes, nor it’s a univocal discriminant from an anatomical and physiological point of view. Nevertheless, even though this approach is worth considering and raises innovative perspectives, the end point is a simple shift from the conscious-unconscious dichotomy to a different taxonomy.

Notwithstanding these attempts of categorizing the domain of learning and memory under the lens of consciousness, it is possible to disentangle further the world of unconscious processing in several, more experimental-specific domains. In cognitive psychology for a long while the term ‘unconscious

processing' has been equated with studies about subliminal learning (Greenwald, Klinger, & Schuh, 1995), in which a stimulus presented for few dozens of milliseconds induces a change in participants behavior, even if not perceived consciously. In this domain, the working hypothesis -or experimental question- was "how the brain processes information that has not been perceived consciously?", thus highlighting the role of consciousness in the perceptual domain (Bargh & Morsella, 2008). On the other side stands the study of unconscious processing considering supraliminal stimuli: within this framework it is possible to further characterize few lines of research. One first direction ties unconscious processing with automatic and habitual behavior (Dickinson, 1985; Ouellette & Wood, 1998) (i.e. procedural knowledge in the previous categorization – fig.1), in which participants are at present unaware of their current behaviors, since it's an automatic or habitual action, but they are nevertheless able to access the information driving their actions a posteriori. For example, typing a manuscript, or riding a bike usually does not require a conscious control over each muscles involved in the execution of the action: nevertheless, at a higher level, awareness and intentionality are still driving our acts. A second line of research with supraliminal stimuli originated from the investigation of unconscious processing in the domain of linguistic, and particularly grammar learning: this line of research was originally started by Arthur Reber in 1967 with a seminal study about the implicit learning of artificial grammars (Reber, 1967a), and it introduced for the first time the term 'implicit learning' paving the way for further studies in this domain beyond the linguistic framework.

A clinical perspective

Ultimately, another line of research to explore unconscious processes in the human brain relies on clinical studies. More specifically, of utmost

interest is the case of amnesic patients. The first tests date back to the second half of the previous century (E K Warrington & Weiskrantz, 1970; Elizabeth K. Warrington & Weiskrantz, 1968, 1982), and curiously reported no difference between amnesic patients and healthy participants in paradigms based on implicit priming effects. Remarkably, the same population of amnesic patients has been tested on more complex implicit learning paradigms (i.e. artificial grammar learning) confirming the lack of difference between amnesic patients and healthy subjects (Channon et al., 2002). On the same line, the results of a study from de Haan (de Haan, Young, & Newcombe, 1987) revealed that patients affected by severe prosopagnosia, i.e. unable to recognize faces of close family members, have normal implicit facial memory compared to a control population; and finally studies on dyslexic and aphasic patients revealed no impairments in implicit learning abilities in the clinical population (i.e. no difference in the processing of words presented too shortly for being read consciously) (Shallice & Saffran, 1986). In addition, a study published in 1987 by Nissen and Bullemer reported that patients affected by Alzheimer's syndrome have no impairment in implicit sequence learning tasks (Nissen & Bullemer, 1987a) when compared to healthy participants. All in all, the clinical approach shed some light on the mechanisms driving unconscious behaviors, revealing no substantial differences between healthy and pathological population, thus suggesting that conscious and unconscious processes possibly rely on different brain regions and involve different neuronal mechanisms.

In the following I will briefly describe the history of the unconscious: from its first steps to its actual scientific investigation. This historical stroll will lead us to the current debate, in which the very existence of unconscious learning is questioned on the ground of methodological and theoretical considerations. In conclusion, an overview about the main paradigms will be provided in the light of the methodological issues raised in the previous paragraph.

Brief history of the unconscious

The term "*unconscious*" has been introduced for the first time in "System of Transcendental Idealism", by the German philosopher Friedrich Schelling, and it was translated in current English for the first time by the poet Samuel Taylor Coleridge in his "Biographia Literaria". The concept of unconsciousness, though, has farther root in the history of human thinking: the concept of unconsciousness can indeed be found at every step of human History. At the beginning of the 16th century for example, the famous Italian doctor Paracelso was thought to base some of his successful treatments on unconscious aspects of cognition (besides the well-known placebo effect, largely -and often unconsciously- exploited by many self-claimed doctors back in the years). Going even farther in time, several centuries before Paracelso's activities, it is possible to read some of the ideas of the Greek philosopher Plato –such as the allegory of the cave or the idea of the Hyperuranion- in the light of unconscious processing: supposedly, he was resuming these ideas from even earlier Orphic and Pythagorean doctrines, whose precise origin is lost in the fogs of History. Interestingly, in non-Western cultures, we find distinct traces of reflections about unconscious thinking: the famous religious Hindu texts, known as Veda and dated around 2500BC in India, refers more than once to the lack of awareness in human behavior and its intentions. Similar thoughts are found also in ancient volumes from the Middle East and Chinese literature. Back in Europe in more recent times, and precisely at the dawn of the scientific revolution, Leibniz, one of the father of mathematical calculus with Isaac Newton, interpreted the idea of unconsciousness mostly as an argument against the tight materialism that was spreading in Europe. The idea of unconscious thinking was

therefore set in opposition with the *res cogitans* (literally a "thinking thing"), *a* Cartesian philosophical idea which identified consciousness and conscious thinking as the unique approach to investigate cognition. In section 17 of his *Monadology*, Leibniz wrote an interesting piece about perception and (un)consciousness, being a very early precursor of modern debates about the nature of consciousness and the so-called 'Homunculi approach' (Dennett, 2003):

> *"One is obliged to admit that perception and what depends upon it is inexplicable on mechanical principles, that is, by figures and motions. In imagining that there is a machine whose construction would enable it to think, to sense, and to have perception, one could conceive it enlarged while retaining the same proportions, so that one could enter into it, just like into a windmill. Supposing this, one should, when visiting within it, find only parts pushing one another, and never anything by which to explain a perception. Thus it is in the simple substance, and not in the composite or in the machine, that one must look for perception."*

After Leibniz, the idea of unconsciousness remained strictly related to the metaphysical realm, and it became the main core of the German philosophical movement known as idealism, whose main interpreters were Fichte and Schelling. It played a quite important role also in the philosophical thought of Kant, Schopenhauer and Nietzsche. All in all, the concept of some sort of thinking outside the scrutiny of consciousness was in part already an idea dwelling in the philosophical domain, but was never investigated under the lens of a scientific perspective. History needed to wait the dawn of the 20th century to see the first considerations about unconsciousness based – if not on scientific experiments- at least on clinical observations.

Even though William James already investigated in 1890 the role of unconscious and subconscious processes in his masterpiece "Principia of Psychology", and later other authors such as Arthur Schopenhauer, Alfred Binet and Pierre Janet (among others), it was Sigmund Freud the first

scientist who approached the concept of unconscious from a clinical perspective. Particularly, Freud made the unconscious the heart of his newborn theory known as Psychanalysis. In his theory of the mind, the unconscious plays a fundamental role, driving most of human behavior. Most of his clinical work aimed at finding ways to reach the deepest level of the human mind: interpretation of dreams, verbal lapses and idiosyncrasies were his weapons to unveil the roots of the psycho-pathologies that were affecting his patients. After him, Jung carried over his legacy about the study of the unconscious mind, going even farther in the interpretations. Among the several contributions to psychology, he introduced the idea of collective unconscious, referring to patterns and structures of the unconscious mind that are shared among all individuals. At the basis of the human collective unconscious there are instincts and archetypes, i.e. universal symbols charged with meanings and interpretations. He distinguished clearly the collective unconscious from the personal unconscious introduced by Freud. Despite both started a scientific path into the psychology of unconscious, they were both very much influenced by earlier philosophers (Freud can be related to many Schopenhauer ideas, whereas Jung is influenced by Schelling works), and lack most of the rigor and exactness that one would expect from a truly scientific investigation.

Finally, at the beginning of the 20th century, a genuine, truly scientific interest started to move its first steps in the study of the unconscious. One approach to investigate (un)consciousness stands in the study of habitual behavior, which focuses on how our actions and choices are affected by unconscious thinking, and specifically how those actions escape at time from the spotlight of focal awareness. More precisely, habits can be defined as the tendencies to repeat responses given a stable supporting context, and their main behavioral advantage consists in the fact that it can be performed

quickly and in parallel, with allocation of minimal mental resources (Ouellette & Wood, 1998). As a matter of fact, most of our daily actions are performed without conscious control: while answering the phone we don't ponder which hand to use, or while driving the car we execute chains of actions (composed by several, smaller chunks) automatically. Quite importantly though, in habitual behavior we still have **intentionality**: the action starts consciously and intentionally. According to this idea, Aarts and Dijksterhuis showed that habitual behavior is indeed triggered by objectives, and can be loosely labeled as goal-directed behavior (Aarts & Dijksterhuis, 2000).

Still bound to the concept of intentionality, another area from the unconscious domain has been investigated by Benjamin Libet (Libet, 1985). Surprisingly, he reported an electrophysiological measure known as "readiness potential" (RP), usually associated with the preparation of a voluntary movement. Notably, this potential was preceding the conscious decision of moving by 300ms - 400ms. In other words, he reported that the physiological marker of the onset of a voluntary movement preceded conscious intentions, thus being unconscious (or preconscious (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006)). Interestingly, if a no-go signal was displayed during the time window between the onset of the RP and the actual movement, the participant was still able to withdraw the movement: subjects could "veto" motor actions even if already started. This result does not surprise in the light of the work of Soon et al. (Soon, Brass, Heinze, & Haynes, 2008), where they showed that action-related activity in prefrontal and parietal cortex can be already recorded 10 seconds before the onset of the movement, leaving plenty of time to alter or interrupt the action. Even though these exciting results are opening perspective for fascinating speculations about free will and consciousness (Haggard, 2008), still a lively

debate is ongoing in the scientific community regarding their interpretation and validity (Roskies, 2006).

In 1977 a very influential and seminal paper (11508 citations in May 2017) has been published by Nisbett and Wilson (Nisbett & Wilson, 1977), in which they reviewed the literature about unconscious processes, focusing particularly on the "little or no introspective access to higher order cognitive processes" showed by their subjects. Their conclusion was that participants are generally unable to report the true causes of their behavior, and that they lack the ability to introspect thoroughly their cognitive processes. Of major interest is the tripartite classification they suggest in classifying unconscious learning paradigms:

1. participants may be unaware of the existence of the *stimulus* that affects their behavior (it is the case of subliminal processing);
2. participants may lack awareness about the *action* they are producing in response to a consciously perceived stimulus;
3. participants may be unaware that the stimulus has affected their response during the experiment (both the stimulus and the response are consciously perceived).

The crucial point in the work of Nisbett and Wilson consists in the paradigm shift from *intentional* unconscious processing, to *unintentional* (or incidental) unconscious processing. In the examples introduced in the previous paragraphs, i.e. habitual behavior and motor preparation (i.e. RP in electrophysiological signals), participants were able to access the information about their actions. In other words, participants had the intention to perform the action or the habits, even if the action is triggered unconsciously or performed outside the spotlight of attention. Here, Nisbett and Wilson pointed their attention to another type of unconscious processing:

those behavioral changes that occur without intentionality from the executer. This work will be seminal in the interpretation of a new domain, still under the conceptual framework of unconscious processing: **implicit learning**.

In the handbook of implicit learning (Ashworth, 1998) there are as many as 11 definitions, withdrawn from different authors and different studies. This number gives the idea regarding the confusion around the topic, mostly in the last decade. Two simple but clear definitions that can help us characterizing implicit learning come from Seger (Seger, 1994) and Cleeremans (Cleeremans, Destrebecqz, & Boyer, 1998). The first one states:

"Implicit learning is non-episodic learning of complex information in an incidental manner, without awareness of what has been learned".

Similarly Cleeremans, few years later:

"Learning is implicit when we acquire new information without intending to do so, and in such a way that the resulting knowledge is difficult to express."

In both definitions two important concepts emerge: the *complexity* of the information learnt, and the *incidental* nature of its acquisition. The first concept is what prevents the information to reach the level of awareness: its complexity prevents the information to reach consciousness during the withdrawal phase (i.e. when the subjects is asked to report the information that affected her behavior). The second notion recalls the point raised by Nisbett and Wilson, i.e. the fact that the learning occurs without intentionality from the participant. In other words, participants learn some information that affect their behavior without willing to do so. Moreover, this information is complex and non-trivial (as a simple association can be). Such complexity raised a plethora of methodological issues related to the measure of

awareness. The current methods used to measure (un)conscious knowledge, and their limitations, are the topics of the next paragraph.

Measuring the unconscious: methodological approaches and related issues

The main issue of investigating the different aspects of consciousness, can be identified in the lack of a clear and broadly shared definition of the term ‘consciousness. Depending on such definition, different approaches can be used in order to measure participants’ awareness in different experimental designs. Regardless of the operative definition chosen, the vast majority of the studies in implicit and subliminal learning tends to apply an approach based on a dissociation logic (Erdelyi, 1986). The idea is that participants perform two sessions: first an exposure phase, in which they implicitly learn the rules or the structure present in the experimental design; secondly they undergo a test phase, in which participants’ knowledge (supposedly implicit) is assessed. Ideally, to claim the occurrence of unconscious learning, it should be possible to observe a measurable change in the behavior in the first phase due to the contingencies, but a performance at chance level during the testing session. The crucial point lays in the nature of both tasks (during the first and second phases), which critically depends on the considered operative definitions of consciousness. In a remarkable review, Destrebecqz and Peigneux framed the problem within the sequence learning domain (but easily generalizable

to other designs), suggesting different approaches to define consciousness (Destrebecqz & Peigneux, 2005). In the following, starting from these definitions we review the respective methodological characteristics and limitations (table 1).

Definition	Tasks / Methods	Criticisms
Ability to verbally access the acquired knowledge	Questionnaire, verbal reports	Information and Sensitivity criteria. No retrieval context
Ability to recollect the acquired knowledge	Generation and recognition tasks. Direct and indirect methods.	Exclusiveness assumption
Acquisition of meta-knowledge	Guessing criterion. Zero-correlation criterion	Depends on subjective interpretation of instructions.

Table 1 – Operative definitions of consciousness, based on Destrebecqz & Peigneux 2005

One first approach would be to define consciousness as the ability to *verbally* access the acquired knowledge: one way to investigate whether participants are aware of the rules consists in asking them to explicitly report such rules or patterns, for example by means of questionnaires. On the one hand this approach has been argued to be essential to probe the existence of conscious knowledge (Overgaard, 2001), but on the other hand several authors have criticized this method on the ground of several considerations. On this regard, a seminal paper published in 1994 by Shanks and St. John (D. Shanks & Stjohn, 1994) deeply undermined the reliability of verbal reports in measuring (un)conscious knowledge by introducing two fundamental criteria: the *information* and the *sensitivity* criteria.

The information criterion suggests that the information responsible for the behavioral changes during the exposure phase should be the same that is assessed during the testing phase. In other words, the measure of awareness should probe the same information than the one driving the behavioral change during the first phase (i.e. the exposure phase). On the other hand, the sensitivity criterion recommends that the test employed to probe awareness is sensitive enough to all the relevant information: all the conscious knowledge used by the participants to perform the task should be probed by the test. Regarding this second criterion, verbal reports fail on the basis of two points: first, participants may fail in reporting all the knowledge they have (fragmentary or incomplete knowledge of the rules can be omitted by participants (Juslin, Winman, & Olsson, 2000)); second, the questionnaire sets a different *retrieval context* than the experimental design (D. Shanks & Stjohn, 1994). The idea of retrieval context is fundamental to address –at least in part- the sensitivity criterion, thus obtaining more reliable results in measuring awareness: simply put, the exposure phase (i.e. when participants acquired the knowledge) and the testing phase (i.e. when the awareness of acquired knowledge is measured) should be as similar as possible, in order to allow participants during the test phase to express all the knowledge they acquired during the exposure phase.

Starting from these considerations and the concept of retrieval context, it is possible to introduce a second, more stringent, operative definition of consciousness (Destrebecqz & Peigneux, 2005): consciousness allows *recollection* of the acquired knowledge, that is the ability to access the acquired knowledge. Having in mind this novel definition, it is possible to introduce two further tasks to probe awareness: the *generation* and the *recognition* tasks. In the generation task, participants are asked to generate a stimulus that would follow the rules acquired during the exposure phase. Conversely, in the recognition task participants have to recognize whether a

stimuli was presented during the previous phase or not. Clearly, both tasks may easily address the information and the sensitivy criteria in a quite convincing way: in both cases the design of the tasks (test phase) share the same context and stimuli as the original one (exposure phase), setting a similar retrieval context. Interestingly, many studies have showed that participants are indeed able to express most of the acquired knowledge through the recognition and generation tasks (Perruchet & Amorim, 1992), questioning the truly implicit nature of many paradigms. On the other hand though, Jimenez and colleagues have questioned whether those tests depend solely on conscious knowledge, or tap also in the unconscious knowledge possessed by participants (Jimenez, Mendez, & Cleeremans, 1996). Theoretically, it is thus possible to frame such problem considering two extremes: on the one side the need to fullfil the sensitivity and the information criteria, having the task as close as possible to the original one, in order to have the same retrieval conditions; on the other extremity the so-called *exclusiveness assumption* (Reingold & Merikle, 1988), which states that the measure of awareness should be sensitive *only* to relevant conscious information, and shouldn't be polluted by unconscious knoweldge. In order to sidestep the challenge of providing a sensitive test of awareness which is at the same time not contaminated by any sort of implicit knowledge, Reingold and Merikle introduced in their work a novel approach (Reingold & Merikle, 1988). In line with the idea of defining consciousness as the ability to retrieve information, they suggested the comparison of *direct* and *indirect* tasks: in the direct tasks participants have to use the implicitly learnt information to perform a novel task (as in a generation task), whereas in the indirect case participants have to actively not use it. In the example of the sequence learning task, in which participants learn implicitly a motor sequence, the indirect task would require participants to produce a novel sequence that is *not* similar to the one learnt during the exposure phase.

Crucially, the direct and the indirect tasks are matching completely except in their instructions: in one case they may use the information learnt implicitly, in the other case they have to use that information explicitly. Theoretically, the comparison of the results on both tasks allows to disentangle the conscious and the unconscious knowledge. This approach has been succesfully used in subliminal perception (Greenwald et al., 1995), unconscious memory (Merikle & Reingold, 1991) and sequence learning (Jimenez et al., 1996). On the same lines Jacoby introduced in 1991 the Process Dissociation Procedure (PDP) (Jacoby, 1991). Despite the originality of the method and the theoretical interest, several controversies have been raised against these methods: particularly, no convincing model has been suggested so far about how unconscious and conscious information are processed in the human brain.

It is possible to introduce a third definition of consciousness from the work by Cheesman and Merikle (Cheesman & Merikle, 1984) in the field of subliminal learning. The idea is to consider two thresholds (one objective and one subjective) to measure stimuli awareness. Perception is defined as being below the objective threshold when the subliminal stimulus is identified at chance level; conversely, the subjective threshold is met when participants are affected by the subliminal stimuli but they report their performance at chance level. Motivated by this work, Dienes and Berry (Zoltan Dienes & Berry, 1997) proposed the same criteria for implicit learning: learning could be described as unconscious when participants claim they are responding at random but they are actually performing above chance level (i.e. following the rules). In other words, the third operative definition of consciousness (Destrebecqz & Peigneux, 2005) introduces the idea of *meta-knowledge*: conscious learning is associated with the acquisition of meta-knowledge, that is participants know they learnt something and that they are using such knowledge. In view of this definition, Dienes and colleagues (Z Dienes &

Perner, 1999; Zoltán Dienes, Altmann, Kwan, & Goode, 1995) introduced two criteria: the *guessing criterion* and the *zero-correlation criterion*. The first one simply states that knowledge is unconscious when participants' performance is above chance, but they claim they are responding at random. The zero-correlation criterion introduces the use of confidence levels, i.e. how confident participants are about their responses, and it states that knowledge is unconscious when confidence levels and performance are uncorrelated. Crucially though, both criteria are affected by the same problem we discussed about verbal reports. As mentioned in the original paper by Cheesman and Merikle (Cheesman & Merikle, 1984), caution should be used when applying these approaches, since both depend on subjective reports and subjective interpretations of instructions. It could happen that participants have a larger interpretation of the term 'guess' than the experimenter, or they may think they are guessing (e.g. in a generation task) when they are actually not guessing (D. Shanks & Stjohn, 1994). Interestingly, a review from Juslin and colleagues (Juslin et al., 2000) provides evidence in this direction suggesting how participants are frequently underconfident when facing a variety of perceptual judgement tasks. In conclusion, subjective measures alone are not sufficient to reliably test participants' awareness.

An interesting approach similar to the zero-correlation criterion but based on signal detection theory (SDT) has been proposed by Kunimoto and colleagues (Kunimoto, Miller, & Pashler, 2001). In his approach, rather than providing a continous value for their confidence about the response, participants are asked to provide a binary confidence judgment, to indicate whether their confidence was low or high. If participants are aware of their knowledge, they should report high confidence for correct answers, and low confidence for wrong answers. It is then possible to create a 2x2 table crossing high/low confidence with correct/incorrect responses. Importantly,

from such table it is possible to extract an index (i.e. meta d', or discriminability index), that is a measure of sensitivity independent from the biases of each participant. Interestingly, Kunimoto and colleagues compared their method with the one used by Cheesman and Merikle (Cheesman & Merikle, 1984), and they found that Cheesman and Merikle method -based on continous confidence judgments - underestimated awareness, thus concluding that their procedure was contaminated by response biases. Nevertheless, Kunimoto was still able to report evidences for subliminal perception using its method.

All in all, measuring unconscious knowledge is extremely challenging. Several methodological issues concern the reliable measure of the unconscious information that drives the behavior of participants, and, starting from these considerations, many researchers have expressed their skepticism about the true existence of unconscious processing (Newell & Shanks, 2014; Nieuwenstein, Blom, Morey, & Wicherts, 2015; D. Shanks & Stjohn, 1994). Besides the problem related to the measure of the implicit knowledge, it is worth mentioning the statistical problems that the study of unconscious learning introduces, such as the introduction of biases, the interpretation of negative results in a frequentist framework and the problem of small sample sizes, which inevitably leads to clamorous failure in replicability (Hendrickx, De Houwer, Baeyens, Eelen, & Van Avermaet, 1997).

Main unconscious learning paradigms, from a critical perspective

Many paradigms have been explored during the last decades to study implicit learning. In this section, the most important will be overviewed, under the critical lens of the methodological issues considered in the previous paragraph.

The first paradigm in which implicit learning has been investigated is the **Artificial Grammar Learning** framework, in the seminal work published by Arthur Reber in 1967 (Reber, 1967b). Originally, artificial grammar learning was a field already explored in the domain of linguistics, particularly by earlier studies performed by George Miller (Miller, 1958), in which he investigated how humans acquired knowledge about grammatical rules. Nevertheless, it was the work of Arthur Reber that paved the way for the new discipline of implicit learning. Reber was the first one who suggested that the rules and the fluency in a language (i.e. language learning) are acquired in an implicit fashion, without awareness playing any role in the learning of the grammatical rules. In his experiments, participants underwent two different sessions. In the first session, they were exposed to a dataset of strings, which –unbeknownst to them- were generated according to some statistical rules. In figure 1 an example of rules is provided: the strings are composed following probabilistic transitions from one state (i.e. one letter) to another, in a Markovian fashion. Participants were asked to study and memorize the strings during the first phase. Before the beginning of the second phase, participants were revealed the existence of some sort of rules (i.e. the artificial grammar) behind the generation of the strings: during the second phase they were asked to classify a new set of strings, based on what they have (implicitly) learnt in the previous phase.

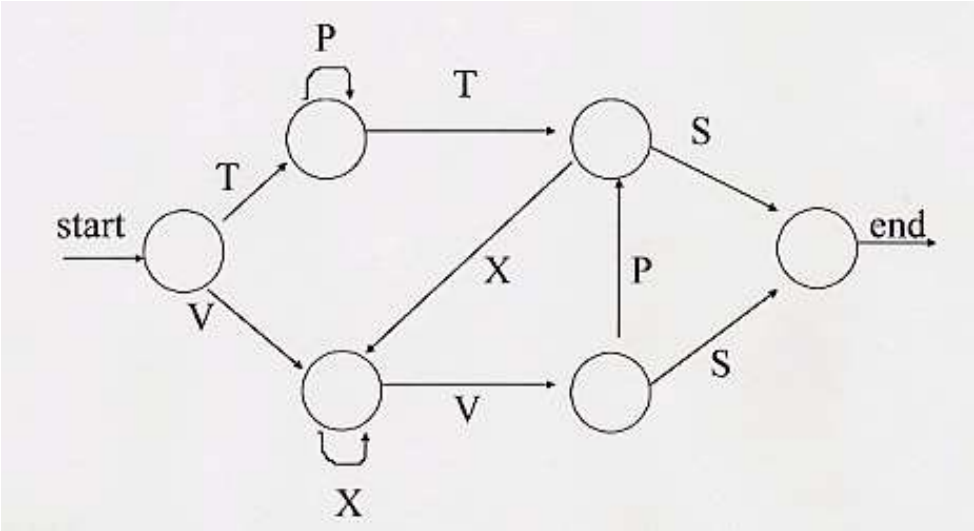


Figure 2 – artificial grammar learning

Arthur Reber reported that participants could classify above chance the strings of the new dataset according to whether they had been generated following the rules of the grammar or not. Crucially, when questioned about the rules, participants were not able to verbalize explicitly the transition probabilities that generated the strings. Eventually, Reber concluded that participants could perform the classification of the new strings by using the implicit knowledge they acquired during the first phase. Many further studies with the same paradigm have been reported in the literature, replicating and expanding such results. For example, a study from Dienes and colleagues (Zoltán Dienes et al., 1995) showed that participants, when exposed to 2 different grammars during the first phase, lacked meta-knowledge about the rules (i.e. the learning was implicit), but were nevertheless able to strategically control which grammar to apply during the test phase, thus showing some degree of strategic control over their knowledge (Zoltán Dienes et al., 1995). Other studies have investigated the learning of artificial

grammars in clinical populations (i.e. amnesic or dyslexic patients) (Channon et al., 2002; Pavlidou, Williams, & Kelly, 2009) and in younger subjects (Pavlidou & Williams, 2014; Witt & Vinter, 2012). Even though these results opened the door of a new field, and paved the way for subsequent studies in implicit learning, many criticisms have been raised regarding the methodologies and the conclusions. Specifically, considering the methodological issues reviewed in the previous paragraph, it is possible to detect a failure in addressing the information criterion (i.e. the information responsible for the behavioral changes during the exposure phase should be the same probed during the testing phase) (D. Shanks & Stjohn, 1994). Verbal reports may indeed fail in probing all the relevant knowledge that may be responsible for the behavior of the participants: an alternative interpretation suggests that participants may respond on the basis of a subset of rules, or other statistics present in the grammar - such as the frequency of occurrence of each letter- that alone could drive a response significantly different from chance (Dulany, Carlson, & Dewey, 1984). In other words, the questionnaire may not be sensitive enough to measure awareness in such a complex task. Therefore, other measures have been considered to probe awareness during an artificial grammar learning experiment, for example confidence ratings (Zoltán Dienes & Altmann, 1997), which allow to apply the zero-correlation criterion, (i.e. implicit learning occurs when there is no correlation between behavior and confidence judgements, and participants have no meta-knowledge of the rules) (Z Dienes & Perner, 1999). Interestingly, in a study from Tunney and Shanks (Tunney & Shanks, 2003), the authors showed in a set of 4 experiments that classification accuracy was poorly correlated with confidence judgements when the judgements were reported on a continuous scale; conversely, when these were reported on a binary scale (high vs low) (Kunimoto et al., 2001) participants showed a correlation between the two measures, thus

revealing some sort of explicit knowledge (as defined by the zero-correlation criterion). In other words, this study reveals the importance of the sensitivity of the tests in measuring awareness, and how difficult it may be for participants themselves to report their knowledge. In conclusion, despite the plethora of studies on the topic, it is not audacious to say that the truly and exclusive implicit nature of the learning in such tasks has not been conclusively proven yet.

Another well-known paradigm used to study implicit learning (and more generally motor control) is **sequence learning**. The first studies investigating sequence learning date back to the beginning of the previous century, with an interesting review from Lashley summarizing the ideas, already in 1951 (Lashley, 1951). The first time a Serial Reaction Time Task appeared in the implicit learning literature was in 1987 by Nissen and Bullemer (Nissen & Bullemer, 1987b). In this task, participants have to respond to an array of visual cues with a specific motor response, typically a key press (fig. 3). Each visual cue is univocally associated with a motor response, so that each cue is mapped with a different key. Unbeknownst to participants, the sequences of visual cues (and therefore the motor responses) are driven by probabilistic (or in some cases deterministic) rules.

Participants implicitly learn the transition probabilities between stimuli, improving their reaction times throughout the experiment. Importantly, when a truly random sequence is provided in place of the learnt one, the reaction times increase significantly, suggesting that participants use their knowledge of the transition probabilities. When tested on the nature of the rules driving the transitions (probabilistic case) or about the sequence per se (deterministic case) most of the participants fail in reporting explicit knowledge of the rules, despite the changes in RT suggesting that they implicitly exploit the information.

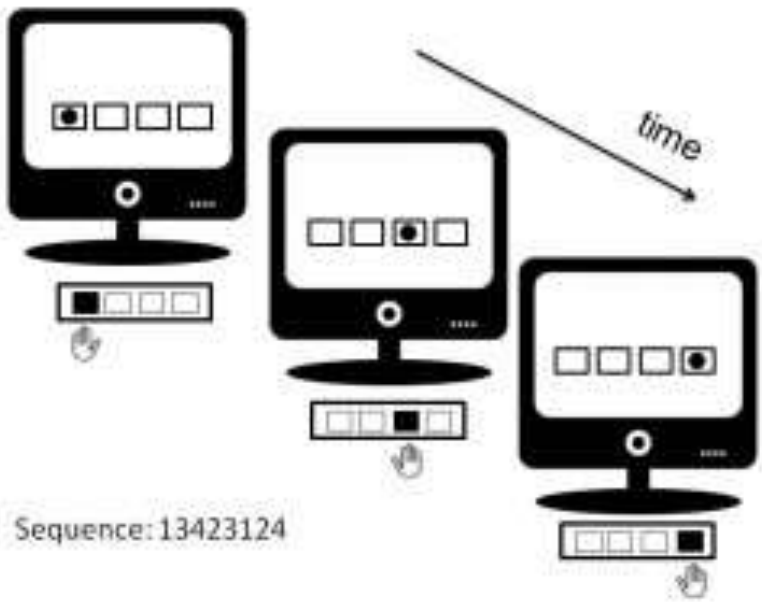


Figure 3 - Sequence learning design

Several studies have investigated how implicit is the knowledge of the rules (Zoltan Dienes & Berry, 1997; Song, Howard Jr., & Howard, 2007), but other studies have provided controversial results using generative methods (D. R. Shanks, 2003; D. R. Shanks, Rowland, & Ranger, 2005), or the Process Dissociation Procedure introduced in the previous paragraph (Fu, Fu, & Dienes, 2008). Moreover, the sequence learning paradigm has been used also to investigate the relationship between implicit learning and attention allocation. Already in their study, Nissen and Bullemer (Nissen & Bullemer, 1987) showed that attracting participants' attentional resources away by adding a secondary orthogonal task critically impaired learning: they concluded that attention does play an important role in the occurrence of implicit learning. Conversely, other authors interpreted such results in the

light of interference between the two tasks (Stadler, 1995), rather than in terms of attentional resources. Cohen provided opposite results (Cohen, Ivry, & Keele, 1990), advocating the independency between attention and implicit learning, but other authors failed in replicating Cohen results (Frensch, Lin, & Buchner, 1998). All in all, the implicit nature of sequence learning and its relationship with attention remains unclear. In the second and third study (respectively chapter III and IV) the main models about unconscious learning and attention will be reviewed in details.

In the late 90s, Chun and colleagues introduced another paradigm to investigate implicit learning: **contextual cueing** (M M Chun & Jiang, 1998; Marvin M. Chun, 2000). The idea behind this paradigm is that context information influences perception, both in a visual search task (contextual cueing) and in object recognition (object cueing). The hypothesis is that contextual information defines some constraints that may provide computational benefit to visual processing, thus facilitating perceptual processing. The nature of these constraints could be either probabilistic, i.e. changing the probability of finding an item in a specific location, or contextual, i.e. it's more likely to find a cup than a lawnmower in a kitchen. Therefore, Chun and Jiang suggested that saliency maps, i.e. topographically arranged maps that represent visual saliency of a corresponding visual scene, are highly influenced by prior information. In their design (see figure 4), participants were asked to perform a visual search task in which they had to identify whether the target (the letter 'T') was on the right or on the left side of the screen. In the visual field, there were several distractors (the letters 'L') displayed throughout the screen. Interestingly, the location of the distractor and of the target was not randomized: during the experiment participants were facing the same 12 'configurations', interleaved with truly randomized configuration. Unbeknownst to the participants, the locations of the target correlated with the 12 set of configurations of the distractors. In

their study, Chun and Jiang observed that participants were faster at localizing the target when the configuration of distractors was familiar than when it was novel, thus concluding that participants were exploiting the contextual information to perform the task. Crucially, participants were not aware of the existence of such configurations, and were not able to verbalize them explicitly at the end of the experiment. As for the previous paradigms, also the implicit nature of the learning during a contextual cueing task has been put into question. Specifically, Shanks and colleagues tested the explicit knowledge of participants with “extended and concurrent tests”, by means of generative and recognition tasks (Smyth & Shanks, 2008). Crucially, they reported in 2 experiments that participants had explicit access to their knowledge when probed with tests that were statistically more powerful than the ones used by Chun and colleagues (i.e. with more trials). They concluded that contextual cueing depends on explicit learning. A subsequent methodological paper, focused on the more typical statistical limitation in testing awareness, has also questioned the implicit nature of learning in a contextual cueing task (Vadillo, Konstantinidis, & Shanks, 2015).

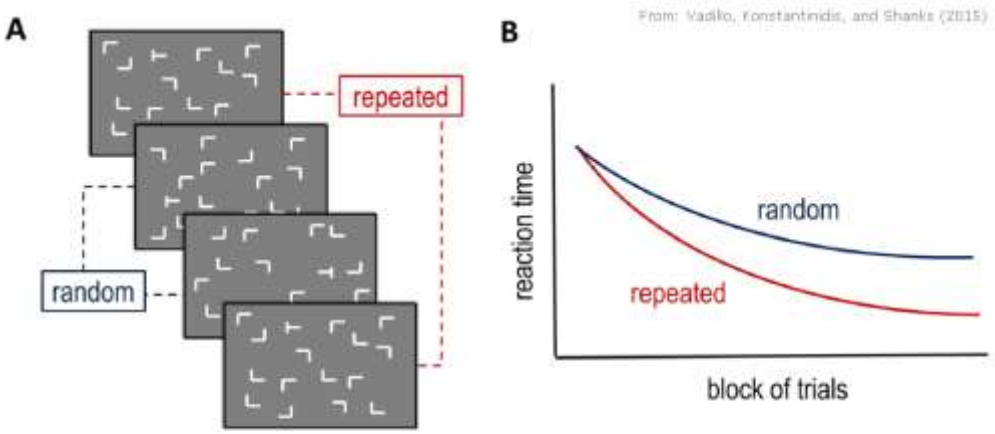


Fig. 1 Panel A shows a sequence of search displays as used in standard contextual cuing experiments. Participants are instructed to search for a T-shaped target among a series of L-shaped distractors. Some search displays are regularly repeated during training, whilst others are new, unrepeated (random) displays. Panel B shows the typical pattern of results: Participants become faster at finding the target among the distractors in repeated displays

Figure 4. Contextual Cueing (Chun, 2000)

Finally, another line of research worth mentioning has been pursued by Ap Dijksterhuis, and it has been formalized in the so called Unconscious Thought Theory (Aarts & Dijksterhuis, 2000; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; Rey, Goldstein, & Perruchet, 2009). Mostly based on the framework of decision making, these experiments are usually composed of two parts: participants first read features related to some items (i.e. cars, apartments, etc.), and then they are asked to decide which one is the best, according to their judgment. The authors compared three groups of participants: one group replied immediately after being informed on the features ('impulsive' group); one group is given some minutes to analyze and ponder on the features of each item ('conscious' group); the last group is engaged in an orthogonal, demanding task before providing an answer ('unconscious' group). They showed that the conscious group was the one

performing the worst, whereas the unconscious participants –explicitly engaged in another task and therefore deciding implicitly- proved to provide the best answers. This effect has been named by the author the 'deliberation-without-attention effect' (Dijksterhuis et al., 2006). This work raised several controversies and criticisms on the ground of several methodological limitations (Dijksterhuis, Knippenberg, Holland, & Veling, 2014; Newell & Shanks, 2014; D. Shanks & Stjohn, 1994), such as small sample sizes and methodological fallacies that led to failure in replications (Abbott, 2013; D. R. Shanks et al., 2013).

All in all, even though many studies have been published on implicit learning, there is no uncontroversial evidence in favor of its existence, according to the most stringent criteria. It is reasonable to assume that unconscious processing in general, and implicit learning more specifically, does exist and affects our behavior. Nevertheless, conclusive evidence in favor of its existence is still lacking. In chapter 2 (i.e. the first experimental study of this work) we will try to provide methodologically sounded evidence in favor of the existence of unconscious learning. In the next paragraph, an overview of the most influential models of consciousness is briefly outlined.

Overview of the current models of consciousness

In recent years, several theories have been proposed to model the mechanisms of consciousness and conscious cognition. In this section of the

work, the main frameworks will be discussed and reviewed from the point of view of implicit learning and unconscious processing.

One of the most popular model of consciousness is named Global Workspace Theory, and it was originally proposed by Baars (Baars, 1997, 2005) (see fig.5).

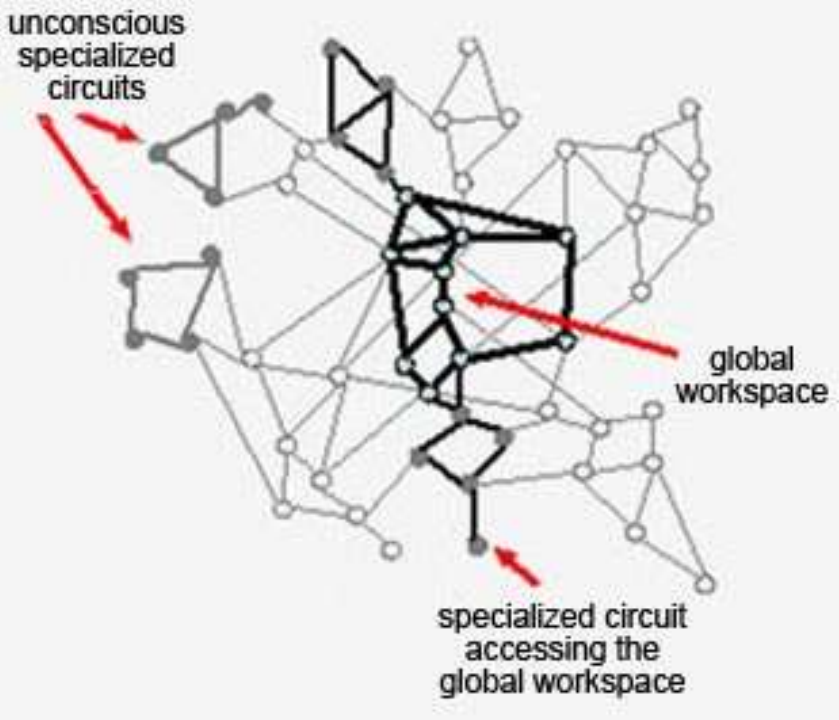


Figure 5 – Representation of the global workspace theory

The idea behind this theory resides on experimental evidence that reveals a globally-distributed brain activity and a long-range intercommunication between cortical regions during conscious mental efforts (Baars, 1997; Massimini, 2005). This observation led Baars to hypothesize the existence of a central workspace which gathers in a single network the activity and the output of smaller, specialized modules (see fig. 5). In this view, each module

is viewed as a specialized processor which encodes for a specific domain: language, motor control, visual processing, etc. Importantly, the activity of each module is unconscious until it reaches the wider, global, central network, at which point it becomes accessible for conscious control. This approach has often been explained in terms of the 'theater metaphor', in which the centralized network would be the stage under the focus of the spot light (sort of an attentional mechanism), whereas the public and the actors behind the curtains represent the unconscious, modular processes. Dehaene and colleagues modeled this scheme in a simulation of an effortful Stroop task, as depicted in figure 6 (Dehaene, Kerszberg, & Changeux, 1998).

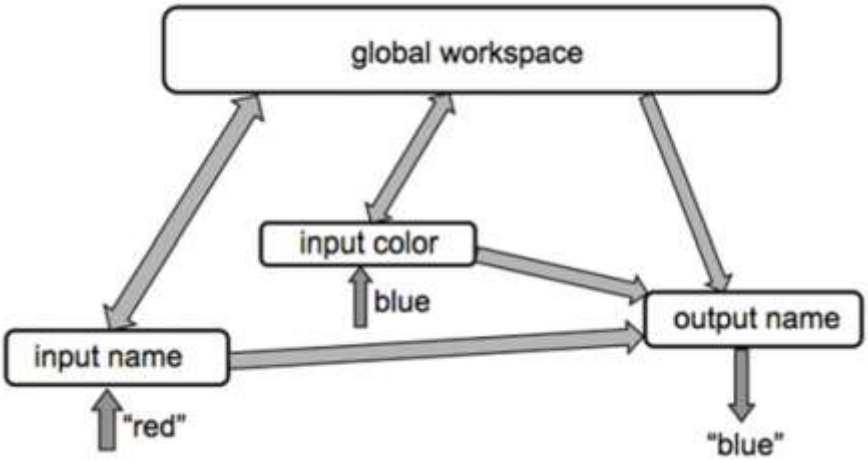


Figure 6 – Simulation of a global workspace model within a stroop task

Remarkably, in this simulation the authors managed to disentangle the difficult condition (when the color and the name were incongruent) and the easy condition (i.e. the congruent case), by showing that the neurons in the global workspace area were activated only during the difficult conditions.

After a certain number of blocks, the activation of these neurons decreased, as if the task was becoming automatic and therefore unconscious. Similarly, we can apply this model to the different paradigms outlined in the previous section. Specifically, it is possible to hypothesize within this framework that the rules of the artificial grammar (or any other rules or association) learnt during the training phase will be stored in a specialized circuit responsible for learning that particular function: for example, statistical transitions would be stored in brain regions involved in processing regularities. Hence, unconscious knowledge would affect behavior while remaining constrained to local brain regions, without entering the main global workspace, thus without reaching awareness.

A completely different modeling approach views consciousness as the result of introspection: an agent is conscious when it has an internal model which represents his own internal states (i.e. interoceptive sensory information, internal representation of its own states, etc.). This idea has roots already in one of the father of artificial intelligence, Marvin Minsky, who theorized this approach already in the 60s (Minsky, 1965). Evolutionarily, the importance of predictability is crucial to increase the chance to survive: consciousness may have arisen when models of the world have become sufficiently complex to incorporate a model of the agent itself (Dawkins, 2006). In this prospect, unconscious learning would be the acquisition of information which escapes representation in the internal model, but it's encoded in other 'lower-level' models. The awareness tests introduced in the previous paragraph, based on the definition of consciousness as the acquisition of meta-knowledge (zero correlation and the guessing criteria (Destrebecqz & Peigneux, 2005)), comply with this theoretical framework: we are conscious of knowledge only if such information is coded in the internal model and we are able to express confidence and judgments about it (i.e. we are able to access this information). In other words, conscious

knowledge is such since it is encoded in this internal model, and is therefore accessible by the agent: all other knowledge which is not modeled in the 'internal model' –but it is coded in other, simpler models- remains unconscious. Regarding the experimental results in the implicit learning literature, it is possible to reason that the information which is not encoded in the internal model of the agent, may still be able to affect its behavior, escaping the scrutiny of consciousness: in other words, some rules and patterns may be encoded and learnt during the experiment, but the agent fails in modeling them in its internal model. In a simplistic, summary view, this framework defines as conscious whatever is encoded in this internal model.

On a similar note lies the idea of different hierarchies (or levels) of representations: as a matter of fact these models are categorized in what has been defined as Higher Order Thought (HOT) theory. In this framework a mental state is defined as conscious when there is another mental state that refers to it (Rosenthal, 1997): the difference with the previous approach consists in the fact that the states are not modeling each other (as in the previous internal model), but they just point at each other hierarchically, the higher level referring to the information encoded in the lower ones. At least two macro-levels are necessary for the simplest model to exist: the lower level is supposedly unconscious and processes information in a simpler way (based, for example, on associations, i.e. Hebb rules driving neuronal plasticity), whereas the higher level is based on more complex computations (i.e. symbolic, or, for example, first order logic) and it is allegedly conscious. Only the higher levels are defined as conscious by virtue of the fact that they monitor and supervise the activity of the lowest levels, without necessarily modeling it (Chella, Frixione, & Gaglio, 2008; Chella & Gaglio, 2007). Interestingly, this idea has been exploited in many frameworks and in several models in which the information is structured in modules: sub-conceptual,

conceptual, linguistic, etc... (Hélie & Street, 2008; Sun & Peterson, 1994). Specifically, one study from Pasquali and Cleeremans (Pasquali, Timmermans, & Cleeremans, 2010) investigated the implementation of a system composed of two neural networks: the first one was trained to perform different kinds of tasks (e.g. discriminating strings in an artificial grammar learning paradigm) whereas the second one was monitoring the performance of the first network (having the input and output units of the first network as its own inputs) and it was learning to wager about the performance of the first network: in other words, the second level was judging how confident it was about the pursuance of the lower level (see fig. 7). From this work it is possible to delineate the operative definition proposed by Cleeremans about consciousness, which culminates in the Radical Plasticity hypothesis. Such hypothesis states that: “Conscious experience occurs if and only if an information processing system has learnt about its own representation of the world. Consciousness is the brain’s theory about itself, gained through experience while interacting with the world, and crucially with itself.” (Cleeremans, 2007). As in the previous frameworks, it is possible to interpret the occurrence of unconscious learning considering that the lower levels (reminiscent of the specific modules in the global workspace theory) learn the contingencies in the environment (e.g. the association between colors and motion directions) affecting the behavior of the participants, whereas the higher levels fail in learning that the lower have learnt such information. The crucial difference with the previous models consists in the fact that a special emphasis is put on the idea that consciousness is depicted as the result of a learning process, which implies that it is a graded rather than one-or-none phenomenon.

Nonetheless, all these models can be grouped together as they all predict the presence of at least two levels (or groups): one set of unconscious, specialized modules which learn the contingencies in the

environment, and a higher level which is conscious and models or embraces all the lower levels. In the Global Workspace theory, for example, the GW can be viewed as the higher level in which all the information is integrated and accessible to all other lower levels; similarly the internal model can be framed as a higher model integrating and framing information from lower modules. Finally, the Higher Order Thought theory are intrinsically based on such idea. A completely different perspective on consciousness, which does not rely on a division in hierarchical layers or internal models, is finally presented in the last paragraph.

The last approach differs from the previous ones in the sense that it is a method for defining and quantifying, rather than modeling, consciousness. It has been initially proposed by Giulio Tononi a decade ago (Tononi, 2004, 2008). The idea is that it is possible to define consciousness as the capacity of the system to integrate information into a unified experience and to differentiate among a large collection of (emotionally) meaningful different states. This approach provides a quantitative measure named φ, which is based on the mathematical formulation of these concepts (differentiation and integration), and which characterizes consciousness as a graded and measurable quantity. One study from Gamez and Morton (Aleksander, Gamez, & Morton, 2009) computed different φ values on several simulated networks revealing that multi-levels architectures such as the Global Workspace Theory or the Higher Order Thought theory, maximize the φ values, coherently with the idea that these architectures underlay a conscious system. On the other hand, a recent work from Seth and colleagues revealed that fully connected networks can have arbitrary levels of φ values, severely undermining its reliability (Seth, Izhikevich, Reeke, & Edelman, 2006).

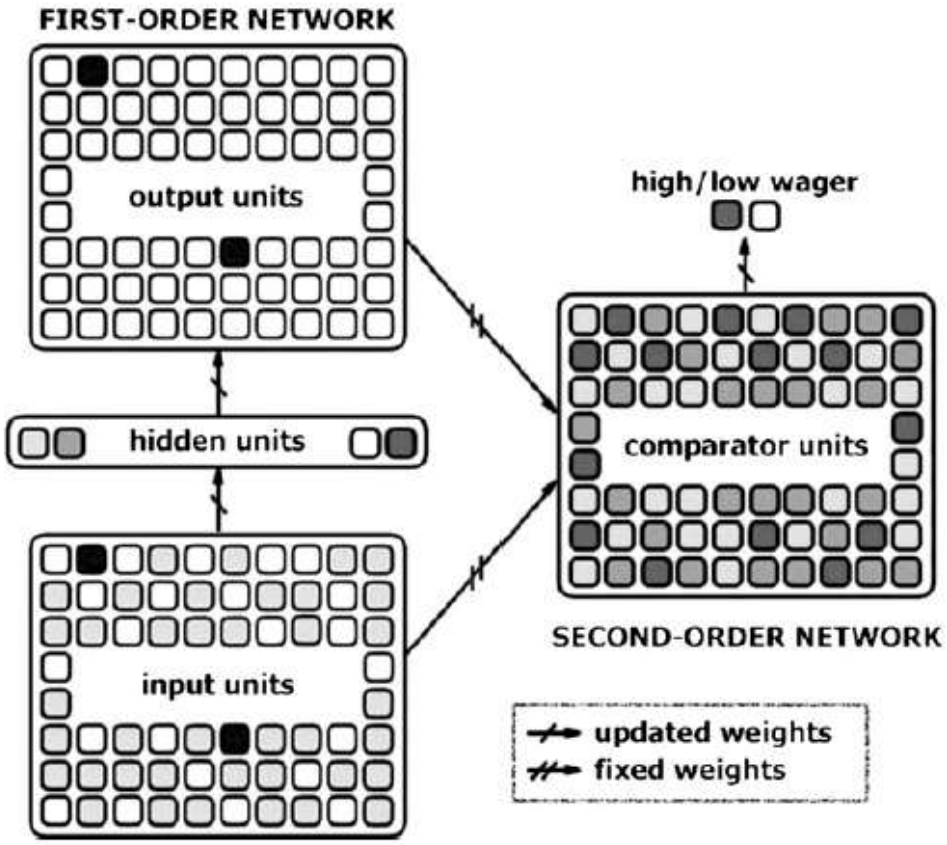


Figure 7 – Representation of the 2-levels network, from Pasquali et al. (2010)

In conclusion, all these frameworks share the idea of having at least two levels: one unconscious and one conscious. Possibly, the optimal way to validate one or the other model is to search for their neuro biological correlates (Block, 2005; Metzinger, 2000). However, any model that aims at explaining the fundamentals of consciousness have to take into consideration the effect observed in unconscious learning paradigms. Unconscious learning therefore is crucial not only as a tool to investigate the

difference between unconscious and conscious behavior in humans, but also as a way of testing and validating any potential theory whose ambition is to unveil the theoretical and biological foundations of consciousness. In the last experimental study (chapter V) the pupil is shown to be a physiological correlate of unconscious processing: in the remaining of the introduction a short overview about the pupil is provided.

The pupil

Pupillary changes have attracted progressively more attention in the scientific community, and particularly in cognitive sciences, for a few reasons: first pupil size has been shown to be a reliable marker of several cognitive processes (see below), and secondly it is a fairly easy and non-invasive recording, which can be performed in several experimental setups. Nevertheless, pupillary changes in response to unconscious processes have been so far poorly investigated, both in the supraliminal and in the subliminal domain. In this section the physiological mechanisms, the cognitive processes involved and the few studies about pupillary responses and unconscious processing are briefly introduced.

The pupil and its physiological mechanisms

It has been long known that pupil size responds to changes in luminance through the pupillary light reflex (Campbell & Gregory, 1960; De Groot & Gebhard, 1952), as well as several cognitive processes which play an important role in the dilation or constriction of the pupil size (see next paragraph). Their effects on the pupil are delivered via the actions of different sub-cortical nuclei in the brain, even though the precise role of each one is not well defined yet. One hypothesis has been put forward by Aston-Jones and Cohen, as reviewed in (Gary Aston-Jones & Cohen, 2005), and it postulates the central role of the **Locus Coeruleus** and the Norepinephrine system (LC-NE) in driving the pupillary responses.

Importantly, the LC plays an important role in several cognitive functions (G. Aston-Jones, Rajkowski, & Kubiak, 1997; Gary Aston-Jones & Cohen, 2005; Jepma, Te Beek, Wagenmakers, van Gerven, & Nieuwenhuis, 2010). Its activity is known to affect arousal and the sleep-wake cycle (Berridge, 2008; Jouvet, 1969), cognitive control (Sara, 2009), stress (Bremner, Krystal, Southwick, & Charney, 1996; Strawn & Geracioti, 2008), depression (Anand & Charney, 2000; Moret & Briley, 2011), decision making (Nieuwenhuis, Aston-Jones, & Cohen, 2005) and attention and behavioral flexibility (Gary Aston-Jones, Rajkowski, & Cohen, 1999; Gabay, Pertzov, & Henik, 2011). In the hypothesis put forward by Aston-Jones and colleagues (Gary Aston-Jones & Cohen, 2005), the Locus Coeruleus plays an important role not only affecting the arousal level of the subject, but also in the gain of cortical circuits. Specifically, they identify a bimodal scheme, also reflected in the pupil activity: a tonic and a phasic mode of operation. Overall, these 2 modalities have been interpreted, in the framework of cognitive behavior, as reflecting the conundrum between exploration (higher tonic activity) and exploitation (higher phasic activity). In line with this interpretation lies the work of Einhauser (Einhäuser, Stout, Koch, & Carter, 2008), which provided experimental evidence in favor of this hypothesis. In their work Einhauser

and colleagues showed that, during a visual perception rivalry task, pupil size was a good predictor regarding the occurrence of the switch of perception, and subsequent stability. On the other hand, an independent group (Hupé, Lamirel, & Lorenceau, 2008) replicated their results but questioned the effect of some confounding factor (i.e. motor response) during the task, thus suggesting the involvement of another mechanisms in driving the effect. However, whether the activity in LC is causal in driving the behavioral strategy or whether such activity reflects the usage of these strategies remains disputed.

Interestingly, a recent work from Joshi and colleagues investigated further the relationship between Locus Coeruleus and pupil size, probing concurrently the superior and inferior colliculi, and the cingulate cortex (Joshi, Li, Kalwani, & Gold, 2016). All these regions are anatomically linked with LC: inferior colliculus receives dense projections from LC, as well as the intermediate layer of the superior colliculus, which has been shown to contribute to the dilation of the pupil based on contrast saliency (C. -a. Wang, Boehnke, White, & Munoz, 2012). The authors of the study assessed the relationship between the neural activity of these areas and pupil size in several ways, and their results reveal that LC is not the only actor in driving the response of the pupil: notably also the neural activity of superior and inferior colliculi was reliably linked with the pupil diameter changes and, to a lesser extent, also cingulate cortex. The authors of the study suggested that all the neural activity of these regions and the dynamic of the pupil diameter is orchestrated and led by a third part. Possible (and not mutually exclusive) candidates are the *paragigantocellularis nucleus* (PGi) of the ventral medulla, which receives and projects to cortical and subcortical regions, such as the Edinger-Westphal nucleus and the LC itself (Breen, Burde, & Loewy, 1983; Vogt, Hof, Friedman, Sikes, & Vogt, 2008), and the mesencephalic cuneiform nucleus (MCN), a circuit centered on the superior colliculus that

has been shown to play a key role in pupil changes specifically in attention and orienting to salience stimuli (C. A. Wang & Munoz, 2015).

Another recent study shed light on the mechanisms that are driving the pupillary dilation in response to cognitive processes. Further than the noradrenergic activity (i.e. the norepinephrine LC-NE system), Jacob Reimer and colleagues (Reimer et al., 2016) investigated the role of the cholinergic levels in cortex responsible for changes in pupil size. They observed a relationship between forebrain cholinergic activity and movement (e.g. walking) (Eggermann, Kremer, Crochet, & Petersen, 2014; Nelson & Mooney, 2016), which is itself linked to arousal, and proposed the hypothesis that rapid dilation of the pupil are associated with noradrenergic phasic activity, whereas cholinergic sustained activity drives long-lasting dilations in the pupil.

In conclusion, different hypotheses have been suggested regarding the mechanisms that drive the pupillary response due to cognitive activity. A complex subcortical network composed of different hubs like Locus Coeruleus and Superior Colliculus is clearly involved in driving pupillary fluctuations, even though further studies will be needed to shed light on the nuclei and neurotransmitters directly or indirectly involved in the pupillary response.

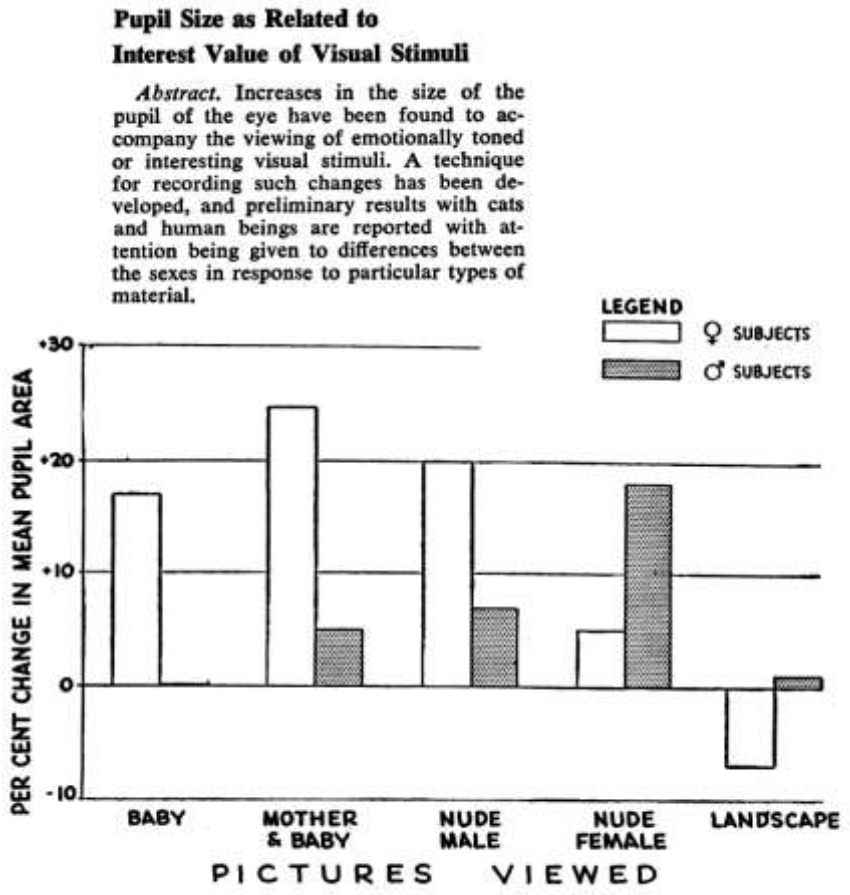
Pupillary response as a marker of cognitive processes

As mentioned in the previous section, beyond changes in luminance, cognitive processes can also modulate pupillary response. In a seminal paper in 1960, Hess and Holt showed for the first time a link between pupillary responses and the emotional value of images (Hess & Polt, 1960). Specifically, they displayed different images having either a pleasing content (i.e. nude of men or women, babies and babies with mothers) or neutral images (i.e. landscapes). As shown in figure 8, they reported a larger dilation of the pupil when visioning images judged as more pleasing (nude pictures, babies and mother with babies for female subjects), whereas they did not report any changes in pupil size when visioning neutral images (i.e. landscapes).

Despite these interesting results, the limitation of the methodological approach (due also to technical issues) was somehow limiting their interpretation: only 5 participants took part in the study and only a handful of pictures was considered during the experiment. Nevertheless, one decade later, another group replicated the results with a larger sample size and dataset of stimuli (but, inevitably, similar technical limitations), suggesting that pupil is modulated not by pleasant stimuli, but rather by arousal content, irrespective of its valence (positive or negative) (Libby, Lacey, & Lacey, 1973).

Later studies have gathered evidence in line with this interpretation (Aboyoun & Dabbs, 1998; Partala & Surakka, 2003), showing that also stimulation in other sensory domains (i.e. auditory) can induce arousal-related pupillary responses (Partala, Jokiniemi, & Surakka, 2000). Pupil

response can therefore be driven by the arousing quality of the stimulus, but also by cognitive processes, such as for example workload (Hampson, Opris, & Deadwyler, 2010; Kahneman & Beatty, 1966).



Pupil Size as Related to Interest Value of Visual Stimuli

Abstract. Increases in the size of the pupil of the eye have been found to accompany the viewing of emotionally toned or interesting visual stimuli. A technique for recording such changes has been developed, and preliminary results with cats and human beings are reported with attention being given to differences between the sexes in response to particular types of material.

Fig. 1. Changes in mean pupil size, in terms of percentage of decrease or increase in area, from size during viewing of control patterns, in response to various pictures.

Figure 8 – Pupil responds to arousal (Hess and Holt, 1960)

Interestingly, Bradley and colleagues, measuring skin conductance and heartbeat concurrently with pupil diameter, were able to link dilation in pupil size due to arousing stimulation to the activity of the sympathetic system, and

dilation due to cognitive load to the depression of the parasympathetic system (Bradley, Miccoli, Escrig, & Lang, 2008).

Further than workload, pupil size can respond to a plethora of different cognitive effects. In 2011, Preuschoff and colleagues investigated the link between uncertainty (modeled as surprising events) and pupillary responses (Preuschoff, 't Hart, & Einhäuser, 2011). In order to reduce any confound due to changes in luminance, they designed a gambling task completely based on auditory stimuli. The hypotheses that motivated their study were first that pupil dilation is driven by changes in noradrenaline (NA) and second that phasic noradrenaline responses have been hypothesized to signal unexpected uncertainty (Dayan & Yu, 2006), in the same way as dopamine affects reward prediction error (W Schultz, 1998; Wolfram Schultz, 1997, 1999): therefore they reasoned that pupil dilation could be a reliable physiological marker for uncertainty. Their results showed that pupil dilation does indeed reflect surprise but not expected reward, providing evidence in favor of the hypothesis that noradrenaline signals uncertainty. More precisely, works from Angela Yu and Peter Dayan (Yu & Dayan, 2005) have framed this model in a Bayesian fashion, hypothesizing that Acetylcholine reflects expected uncertainty (i.e. known unreliability) whereas Norepinephrine signals unexpected uncertainty (i.e. surprising events).

Similar results were obtained by O'Reilly and colleagues (O'Reilly et al., 2013), who designed a saccadic planning experiment in which they were able to disentangle purely surprising events (formally defined by the Shannon information) and events which led to an update of the internal model (modeled, in a predictive framework, by the Kullback-Leibler divergence). Notably, they found significant activation of Lateral Intraparietal area (LIP) following a surprising event, whereas Anterior Cingulate Cortex (ACC) was significantly activated following an update of the internal model. Crucially,

they found different profiles of pupillary response in the two conditions: an increase in pupil size in response to a surprising event, and a later decrease following updates of the internal model (see figure 9).

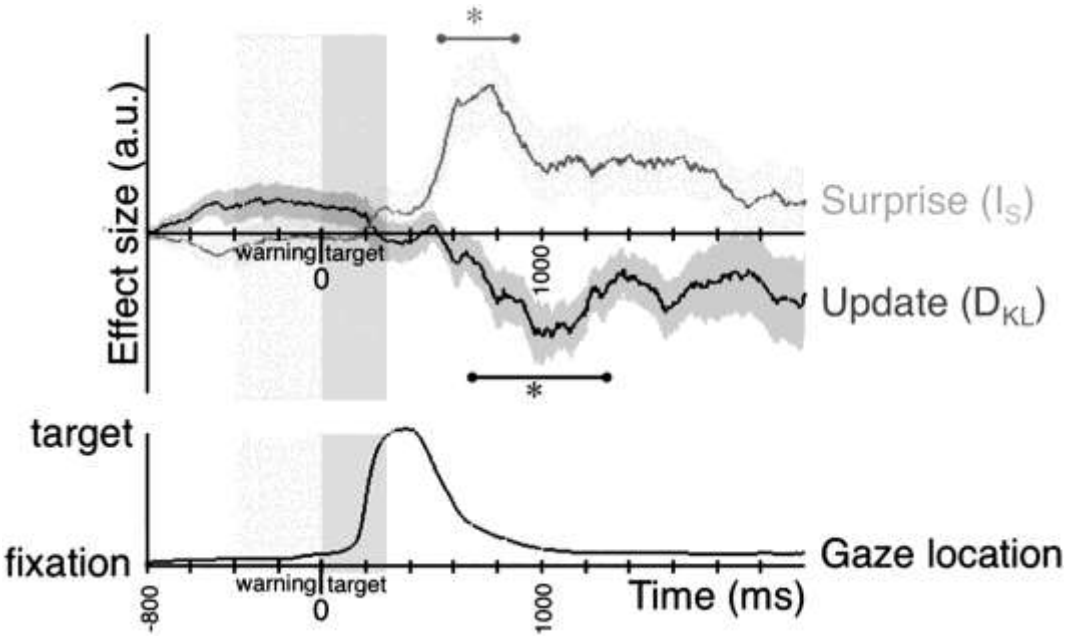


Figure 9 – Pupil response to surprise and model update (O'reilly et al. 2013).

Interestingly, Naber and colleagues (Naber, Frässle, Rutishauser, & Einhäuser, 2013) related pupillary responses to the retrieval of declarative memories, reporting a larger constriction for novel rather than familiar scenes. Their interpretation of these results is that pupil constriction can serve as a marker for novel stimuli. Surprisingly, they neither discussed in depth the possible confounds related with the effort and arousal due to the

instruction to memorize the images, nor did they control for changes in the image luminance. Moreover, the linear model of the pupil response they chose may be replaced by other approaches which could highlight better the dynamic of the pupillary response. Nevertheless, their results may be interpreted in light of the previous studies, in which pupillary responses have been shown to track surprise and uncertainty (i.e. novel stimuli).

A further study from De Gee related pupil dilation to a perceptual decision making task: surprisingly, their results showed that pupil dilation was a reliable predictor of 1) the upcoming choice of the participant and 2) of the individual bias. Their interpretation is that neuro-modulatory systems regulating pupil size are involved in decision making and it is possible to track during the time-course of the decision their influence on the pupil. The authors didn't investigate further the relationship between uncertainty in the decision-making process and the pupillary response: nevertheless the link they reported between pupillary response and individual bias suggests a further mechanisms that drive its response on top of uncertainty.

At the end of our brief survey, an influential study published by Nassar and colleagues (Nassar et al., 2012) established a link between pupil metrics and learning dynamics: they investigated how, in a dynamic environment, new information is regulated in updating internal models, and how this is reflected in changes in pupil size. Specifically, they found first a positive correlation between change-point probability (i.e. the posterior probability of a substantial change in the environment) and changes in pupil size; and second a positive correlation between average pupil size and relative uncertainty (i.e. uncertainty about the true underlying parameters of the environment). Interestingly, they were also able to demonstrate individual learning differences and changes in pupil size. Their interpretation suggests a relationship between arousal state and learning rate coordinated by what

they label a ‘learning-arousal network’, which includes the Locus Coeruleus and Anterior Cingulate Cortex.

The short review of studies revealed that pupil has been related to several cognitive mechanisms. Despite the fact that the several physiological mechanisms that are driving the pupillary responses have been investigating, it is still lacking a validated theory able to explain all these results in the light of one coherent hypothesis. In the last part of this work (i.e. discussion) the thesis that precision-weighted prediction error may explain the pupillary dynamic is briefly introduced and discussed in the predictive coding framework.

Pupillary responses and unconscious processing

Pupillary changes in response to unconscious processes have been so far poorly investigated, specifically in the domain of implicit learning (i.e. using supraliminal stimuli). It is easy though to imagine how pupil would be affected by unconscious processing: one of the feature of implicit learning is the automaticity and the near absence of mental effort, which in turn would act on pupillary responses when compared to explicit learning scenarios. Even though it would be relatively straightforward to test this prediction, so far very few studies have investigated this question (Diede & Bugg, 2017). In a different context, it has been showed that amnesic patients who were unable to judge whether a set of stimuli was novel or already seen, had an increase in pupillary response when visioning novel stimuli, suggesting that

pupil may respond to novelty even when not consciously perceived as such (Bruno Laeng et al., 2007).

Relatively more work has been performed regarding how pupillary responses are affected by subliminal perception (B. Laeng, Sirois, & Gredeback, 2012). A Japanese group from Kanagawa university investigated how pupil was affected in a subliminal mere exposure effect, an effect that supposedly modulates autonomic nervous responses, such as pupil diameter (Yoshimoto, Imai, Kashino, & Takeuchi, 2014). In their study, they showed that perceptual fluency was influenced by subliminal exposure to the stimuli, and this effect was observable in the pupillary responses. Their interpretation calls for a reduction in mental effort, which in turn affects pupil size. A study from Gomes and colleagues (Gomes, Montaldi, & Mayes, 2015) found a priming effects in the pupil that they named the pupil priming effect. Specifically, after a perceptual detection task, participants performed a recognition memory test, in which they revealed a larger pupil dilation for missed target than for correctly rejected objects. Even though the authors never tested directly the actual subliminal (i.e. unconscious) nature of the perception of the items, this study suggests that pupillary response may reflect processes that don't reach a conscious level.

Similarly, a study from the Utrecht university revealed a pupillary response to rewarding subliminal cues, but specifically when the task was demanding (Bijleveld, Custers, & Aarts, 2009). The interpretation of their finding was that subliminal cues were strategically integrated in the process of recruiting resources: the selective response of the pupil only in the difficult condition reflects a sort of strategical processing, which is in line with the interpretation that subliminal cues are not merely causing a reflex in the response, but are strategically integrated in the allocation of resources. In

this study pupil was used as a way to measure how subliminal presented information is processed in the brain.

In the last experimental part of this work (chapter V) it is shown how pupil response can track unconsciously salient event, thus shedding light on the role of conscious and unconscious mechanisms in driving the pupillary response.

References

Aarts, H., & Dijksterhuis, A. (2000). Habits as Knowledge Structures: Automaticity in Goal-Directed Behavior. *Journal of Personality and Social Psychology*, *78*(1), 53–63. http://doi.org/10.1037//0022-3514.78.1.53

Abbott, A. (2013). Disputed results a fresh blow for social psychology. *Nature*, *497*(7447), 16. http://doi.org/10.1038/497016a

Aboyoun, D. C., & Dabbs, J. M. (1998). the Hess Pupil Dilation Findings: Sex or Novelty? *Social Behavior and Personality: An International Journal*, *26*(4), 415–419. http://doi.org/10.2224/sbp.1998.26.4.415

Anand, A., & Charney, D. S. (2000). Norepinephrine dysfunction in depression. *Journal of Clinical Psychiatry*.

Ashworth, P. (1998). Handbook of Implicit Learning. *Physiotherapy*, *84*(5), 240. http://doi.org/10.1016/S0031-9406(05)65559-9

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–50. http://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Foote, S. L., & Segal, M. (1985). Impulse conduction properties of noradrenergic locus coeruleus axons projecting to monkey cerebrocortex. *Neuroscience*, *15*(3). http://doi.org/10.1016/0306-4522(85)90077-6

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus

coeruleus in attention and behavioral flexibility. *Biological Psychiatry*. http://doi.org/10.1016/S0006-3223(99)00140-7
Aston-Jones, G., Rajkowski, J., & Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience*, *80*(3), 697–715. http://doi.org/10.1016/S0306-4522(97)00060-2
Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus Coeruleus Neurons in Monkey Are Selectively Activated by Attended Cues in a Vigilance Task. *The Journal of Neuroscience*, *14*(July), 4467–4480.
Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science*, *3*(1), 73–79. http://doi.org/10.1111/j.1745-6916.2008.00064.x
Berridge, C. W. (2008). Noradrenergic modulation of arousal. *Brain Research Reviews*. http://doi.org/10.1016/j.brainresrev.2007.10.013
Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science*, *20*(11), 1313–1315. http://doi.org/10.1111/j.1467-9280.2009.02443.x
Bouret, S., & Sara, S. J. (2004). Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *European Journal of Neuroscience*, *20*(3), 791–802. http://doi.org/10.1111/j.1460-9568.2004.03526.x
Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. http://doi.org/10.1111/j.1469-8986.2008.00654.x
Breen, L. A., Burde, R. M., & Loewy, A. D. (1983). Brainstem connections to the Edinger-Westphal nucleus of the cat: a retrograde tracer study. *Brain Research*, *261*(2), 303–306. http://doi.org/https://doi.org/10.1016/0006-8993(83)90633-9
Bremner, J. D., Krystal, J. H., Southwick, S. M., & Charney, D. S. (1996). Noradrenergic mechanisms in stress and anxiety: I. Preclinical studies. *Synapse*, *23*(1), 28–38. http://doi.org/10.1002/(SICI)1098-2396(199605)23:1<28::AID-SYN4>3.0.CO;2-J

Campbell, F. W., & Gregory, A. H. (1960). The effect of pupil size on visual acuity. *Nature*, *187*, 1121–1123. http://doi.org/10.1038/1871121c0

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, *40*(12), 2185–2197. http://doi.org/10.1016/S0028-3932(02)00037-4

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*(4), 387–395. http://doi.org/10.3758/BF03202793

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*. http://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. http://doi.org/10.1006/cogp.1998.0681

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning : news from the front, *2*(10), 587–590.

Dayan, P., & Yu, A. (2006). Norepinephrine and neural interrupts. *… in Neural Information Processing Systems*. Retrieved from http://www.cogsci.ucsd.edu/~ajyu/Papers/nips05.pdf%5Cnfile:///Users/alexandershaw/Library/Application Support/Papers2/Articles/2006/Dayan/2006 Dayan.pdf%5Cnpapers2://publication/uuid/C29CFE48-18A1-4599-B0D9-7D575F24826F

De Groot, S. G., & Gebhard, J. W. (1952). Pupil size as determined by adapting luminance. *Journal of the Optical Society of America*, *42*(7), 492–495. http://doi.org/10.1364/JOSA.42.000492

de Haan, E. H., Young, A., & Newcombe, F. (1987). Face recognition without awareness. *Cogn Neuropsychol*, *4*(4), 385–415. http://doi.org/10.1080/02643298708252045

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. http://doi.org/10.1016/j.tics.2006.03.007

Dennett, D. C. (2003). Dan Dennett on our consciousness.

http://doi.org/YouTube http://www.youtube.com/watch?v=kLa2WZPkQls
Destrebecqz, A., & Peigneux, P. (2005). Methods for studying unconscious learning. *Progress in Brain Research*. http://doi.org/10.1016/S0079-6123(05)50006-2
Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*(1135), 67–78. http://doi.org/10.1098/rstb.1985.0010
Diede, N. T., & Bugg, J. M. (2017). Journal of Experimental Psychology : Learning , Memory , and Cognition Cognitive Effort Is Modulated Outside of the Explicit Awareness of Conflict Frequency : Evidence From Pupillometry Cognitive Effort Is Modulated Outside of the Explicit Awareness of Co.
Dienes, Z., & Altmann, G. T. M. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? *How Implicit Is Implicit Learning?*, 107–123.
Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2*(5), 1322–1338. http://doi.org/10.1037/0278-7393.21.5.1322
Dienes, Z., & Berry, D. C. (1997). Implicit learning: Below the subjective threshold? *Psychonomic Bulletin & Review*, *4*(1), 3–23. http://doi.org/10.3758/BF03210769
Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *The Behavioral and Brain Sciences*, *22*(5), 735-755-808. http://doi.org/10.1017/S0140525X99002186
Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: the deliberation-without-attention effect. *Science (New York, N.Y.)*, *311*(5763), 1005–7. http://doi.org/10.1126/science.1121629
Dijksterhuis, A., Knippenberg, A. van, Holland, R. W., & Veling, H. (2014). Newell and Shanks' approach to psychology is a dead end. *Behavioral & Brain Sciences*, *37*(1), 25–26. http://doi.org/10.1017/S0140525X1300068X
Dulany, D. E., Carlson, R. a., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, *113*(4),

541–555. http://doi.org/10.1037/0096-3445.113.4.541
Eggermann, E., Kremer, Y., Crochet, S., & Petersen, C. C. H. (2014). Cholinergic Signals in Mouse Barrel Cortex during Active Whisker Sensing. *Cell Reports*, *9*(5), 1654–1661. http://doi.org/10.1016/j.celrep.2014.11.005
Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1704–9. http://doi.org/10.1073/pnas.0707727105
Erdelyi, M. H. (1986). Experimental indeterminacies in the dissociation paradigm of subliminal perception. *Behavioral and Brain Sciences*, *9*(1), 30–31. http://doi.org/10.1017/S0140525X00021348
Fu, Q., Fu, X., & Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness and Cognition*, *17*(1), 185–202. http://doi.org/10.1016/j.concog.2007.01.007
Gabay, S., Pertzov, Y., & Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception & Psychophysics*, *73*(1), 123–9. http://doi.org/10.3758/s13414-010-0015-4
Gomes, C. A., Montaldi, D., & Mayes, A. (2015). The pupil as an indicator of unconscious memory: Introducing the pupil priming effect. *Psychophysiology*, *52*(6), 754–769. http://doi.org/10.1111/psyp.12412
Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by Marginally Perceptible (“Subliminal”) Stimuli: Dissociation of Unconscious From Conscious Cognition. *Journal of Experimental Psychology: General*, *124*(I), 22–42. http://doi.org/10.1037/0096-3445.124.1.22
Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews. Neuroscience*, *9*(12), 934–46. http://doi.org/10.1038/nrn2497
Hampson, R. E., Opris, I., & Deadwyler, S. A. (2010). Neural correlates of fast pupil dilation in nonhuman primates: Relation to behavioral performance and cognitive workload. *Behavioural Brain Research*, *212*(1), 1–11. http://doi.org/10.1016/j.bbr.2010.03.011
Heilbronner, S. R., & Haber, S. N. (2014). Frontal Cortical and

Subcortical Projections Provide a Basis for Segmenting the Cingulum Bundle: Implications for Neuroimaging and Psychiatric Disorders. *The Journal of Neuroscience*, *34*(30), 10041–10054. http://doi.org/10.1523/JNEUROSCI.5459-13.2014

Hendrickx, H., De Houwer, J., Baeyens, F., Eelen, P., & Van Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 201–220. http://doi.org/10.1037/0278-7393.23.1.201

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews. Neuroscience*, *11*(7), 523–32. http://doi.org/10.1038/nrn2850

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science (New York, N.Y.)*, *132*(3423), 349–50. http://doi.org/10.1126/science.132.3423.349

Hupé, J.-M., Lamirel, C., & Lorenceau, J. (2008). Pupil dilation does not predict subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(28), E43; author reply E44. http://doi.org/10.1073/pnas.0803456105

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. http://doi.org/10.1016/0749-596X(91)90025-F

Jepma, M., Te Beek, E. T., Wagenmakers, E.-J., van Gerven, J. M. A., & Nieuwenhuis, S. (2010). The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Frontiers in Human Neuroscience*, *4*(August), 170. http://doi.org/10.3389/fnhum.2010.00170

Jimenez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc3&NEWS=N&AN=1996-04794-008

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–

234. http://doi.org/10.1016/j.neuron.2015.11.028
Jouvet, M. (1969). Biogenic Amines and the States of Sleep. *Science*, *163*(3862), 32 LP-41. Retrieved from http://science.sciencemag.org/content/163/3862/32.abstract
Juslin, P., Winman, A., & Olsson, H. (2000). Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect overreliance on the. *Psychological Review*, *107*(2), 384–396. http://doi.org/10.1037//0033-295X.107.2.384
Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science (New York, N.Y.)*, *154*(756), 1583–1585. http://doi.org/10.1126/science.154.3756.1583
Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, *10*(3), 294–340. http://doi.org/10.1006/ccog.2000.0494
Kuntz, A. (1934). *The autonomic nervous system*. Lea & Febiger. Retrieved from https://books.google.be/books?id=F8IEAQAAIAAJ
Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. http://doi.org/10.1177/1745691611427305
Laeng, B., Waterloo, K., Johnsen, S. H., Bakke, S. J., Låg, T., Simonsen, S. S., & Høgsaet, J. (2007). The eyes remember it: oculography and pupillometry during recollection in three amnesic patients. *Journal of Cognitive Neuroscience*, *19*(11), 1888–1904. http://doi.org/10.1162/jocn.2007.19.11.1888
Lashley, K. (1951). The problem of serial order in behavior. *1951*, (7), 112–147. http://doi.org/10.1093/rfs/hhq153
Levitt, P., & Moore, R. Y. (1978). Noradrenaline neuron innervation of the neocortex in the rat. *Brain Research*, *139*(2), 219–231. http://doi.org/10.1016/0006-8993(78)90925-3
Libby, W. L., Lacey, B. C., & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*. http://doi.org/10.1111/j.1469-8986.1973.tb00526.x
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*(4), 529–539. http://doi.org/10.1017/S0140525X00044903
Lorenceau, J., Paradis, A.-L., Lamirel, C., Poline, J.-B., Artiges, E.,

Thirion, B., & Caclin, A. (2008). Cortical dynamics of bistable form/motion binding: fMRI and eye movements. *Journal of Vision*, *8*(6), 1105. Retrieved from http://dx.doi.org/10.1167/8.6.1105

Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 224–233. http://doi.org/10.1037/0278-7393.17.2.224

Miller, G. A. (1958). Free Recall of Redundant Strings of Letters. *Journal of Experimental Psychology*, *56*(6), 485–491. http://doi.org/10.1037/h0044933

Moret, C., & Briley, M. (2011). The importance of norepinephrine in depression. *Neuropsychiatric Disease and Treatment*, *7*(SUPPL.), 9–13. http://doi.org/10.2147/NDT.S19619

Mountcastle, V. B., LaMotte, R. H., & Carli, G. (1972). Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. *Journal of Neurophysiology*, *35*(1), 122 LP-136. Retrieved from http://jn.physiology.org/content/35/1/122.abstract

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11. http://doi.org/10.1167/13.2.11.doi

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. http://doi.org/10.1038/nn.3130

Nelson, A., & Mooney, R. (2016). The Basal Forebrain and Motor Cortex Provide Convergent yet Distinct Movement-Related Inputs to the Auditory Cortex. *Neuron*, *90*(3), 635–648. http://doi.org/https://doi.org/10.1016/j.neuron.2016.03.031

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behavioral and Brain Sciences*, *37*, 1–19. http://doi.org/10.1017/S0140525X12003214

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532.

http://doi.org/10.1037/0033-2909.131.4.510
Nieuwenstein, M. R., Blom, T. N., Morey, R. D., & Wicherts, J. M. (2015). On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage. *Judgement and Decision Making*, *10*(1), 1–17.
Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Report on Mental Process. *Psychological Review*, *84*(3), 231–259.
Nissen, M. J., & Bullemer, P. (1987a). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. http://doi.org/10.1016/0010-0285(87)90002-8
Nissen, M. J., & Bullemer, P. (1987b). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. http://doi.org/10.1016/0010-0285(87)90002-8
O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), E3660--9. http://doi.org/10.1073/pnas.1305373110
Ouellette, J. A., & Wood, W. (1998). Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior. *Psychological Bulletin*, *124*(1), 124–54. http://doi.org/10.1037/0033-2909.124.1.54
Overgaard, M. (2001). The role of phenomenological reports in experiments on consciousness. *Psycoloquy*, *12*, 29.
Partala, T., Jokiniemi, M., & Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 123–129. http://doi.org/10.1145/355017.355042
Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human Computer Studies*, *59*(1–2), 185–198. http://doi.org/10.1016/S1071-5819(03)00017-X
Pavlidou, E. V., & Williams, J. M. (2014). Implicit learning and reading: Insights from typical children and children with developmental dyslexia using the artificial grammar learning

(AGL) paradigm. *Research in Developmental Disabilities*, *35*(7), 1457–1472. http://doi.org/10.1016/j.ridd.2014.03.040
Pavlidou, E. V., Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of Dyslexia*, *59*(1), 55–77. http://doi.org/10.1007/s11881-009-0023-z
Perruchet, P., & Amorim, M. a. (1992). Conscious knowledge and changes in performance in sequence learning: evidence against dissociation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(4), 785–800. http://doi.org/10.1037/0278-7393.18.4.785
Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, (SEP), 1–12. http://doi.org/10.3389/fnins.2011.00115
Reber, A. S. (1967a). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. http://doi.org/10.1016/S0022-5371(67)80149-X
Reber, A. S. (1967b). Implicit learning of artificial grammars. Retrieved from http://philpapers.org/rec/REBILO
Reimer, J., Mcginley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., Mccormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Publishing Group*, *7*(May), 1–7. http://doi.org/10.1038/ncomms13289
Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*. http://doi.org/10.3758/BF03207490
Rey, A., Goldstein, R. M., & Perruchet, P. (2009). Does unconscious thought improve complex decision making? *Psychological Research*, *73*(3), 372–379. http://doi.org/10.1007/s00426-008-0156-4
Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences*, *10*(9), 419–423. http://doi.org/10.1016/j.tics.2006.07.011
Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews. Neuroscience*, *10*(3), 211–223. http://doi.org/10.1038/nrn2573
Schultz, W. (1997). Dopamine neurons and their role in reward

mechanisms. *Current Opinion in Neurobiology*, *7*(2), 191–197. http://doi.org/10.1016/S0959-4388(97)80007-4
Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. http://doi.org/10.1007/s00429-010-0262-0
Schultz, W. (1999). The Reward Signal of Midbrain Dopamine Neurons. *News in Physiological Sciences*, *14*(December), 249–255. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11390860
Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163–196. http://doi.org/10.1037/0033-2909.115.2.163
Shallice, T., & Saffran, E. (1986). Lexical processing in the absence of explicit word identification: Evidence from a letter-by-letter reader. *Cognitive Neuropsychology*, *3*(4), 429–458. http://doi.org/10.1080/02643298608252030
Shanks, D. R. (2003). Attention and awareness in “implicit” sequence learning. In *Attention and implicit learning. Advances in Consciousness Research, vol. 48* (pp. 11–42).
Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., … Moore, C. (2013). Priming Intelligent Behavior: An Elusive Phenomenon. *PLoS ONE*, *8*(4). http://doi.org/10.1371/journal.pone.0056515
Shanks, D. R., Rowland, L. a, & Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research*, *69*(5–6), 369–82. http://doi.org/10.1007/s00426-004-0211-8
Shanks, D., & Stjohn, M. (1994). Characteristics of Dissociable Human Learning-Systems, (February 2010). http://doi.org/10.1017/S0140525X00035032
Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403–415. http://doi.org/10.3758/MC.36.2.403
Song, S., Howard Jr., J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learn Mem*, *14*(3), 167–176. http://doi.org/10.1101/lm.437407
Soon, C. S., Brass, M., Heinze, H., & Haynes, J. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545. http://doi.org/10.1038/nn.2112

Squire, L. R., & Dede, A. J. O. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*(3). http://doi.org/10.1101/cshperspect.a021667

Strawn, J. R., & Geracioti, T. D. (2008). Noradrenergic dysfunction and the psychopharmacology of posttraumatic stress disorder. *Depression and Anxiety*. http://doi.org/10.1002/da.20292

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306. http://doi.org/10.1126/science.2296719

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition*, *31*(7), 1060–1071. http://doi.org/10.3758/BF03196127

Vadillo, M. a., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-015-0892-6

Vogt, B. A., Hof, P. R., Friedman, D. P., Sikes, R. W., & Vogt, L. J. (2008). Norepinephrinergic afferents and cytology of the macaque monkey midline, mediodorsal, and intralaminar thalamic nuclei. *Brain Structure and Function*, *212*(6), 465–479. http://doi.org/10.1007/s00429-008-0178-0

Wang, C. -a., Boehnke, S. E., White, B. J., & Munoz, D. P. (2012). Microstimulation of the Monkey Superior Colliculus Induces Pupil Dilation Without Evoking Saccades. *Journal of Neuroscience*, *32*(11), 3629–3636. http://doi.org/10.1523/JNEUROSCI.5512-11.2012

Wang, C. A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*. http://doi.org/10.1016/j.conb.2015.03.018

Warrington, E. K., & Weiskrantz, L. (1968). A study of learning and retention in amnesic patients. *Neuropsychologia*, *6*(3), 283–291. http://doi.org/10.1016/0028-3932(68)90026-2

Warrington, E. K., & Weiskrantz, L. (1970). Amnesic syndrome: consolidation or retrieval? *Nature*, *228*, 628–630. http://doi.org/10.1038/228628a0

Warrington, E. K., & Weiskrantz, L. (1982). Amnesia: A disconnection syndrome? *Neuropsychologia*, *20*(3), 233–248. http://doi.org/10.1016/0028-3932(82)90099-9

Witt, A., & Vinter, A. (2012). Artificial grammar learning in children: Abstraction of rules or sensitivity to perceptual features? *Psychological Research*, *76*(1), 97–110. http://doi.org/10.1007/s00426-011-0328-5

Yoshimoto, S., Imai, H., Kashino, M., & Takeuchi, T. (2014). Pupil response and the subliminal mere exposure effect. *PloS One*, *9*(2), e90670. http://doi.org/10.1371/journal.pone.0090670

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. http://doi.org/10.1016/j.neuron.2005.04.026

Chapter II: Unconscious associative learning with conscious cues

This chapter is a modified version of the following article:

Alamia, A., Orban de Xivry, J.-J., San Anton, E., Olivier, E., Cleeremans, A., & Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of Consciousness*, *2016*(1), 1–10. http://doi.org/10.1093/nc/niw016

Abstract

Despite extensive research, the very existence of unconscious learning in humans remains much debated. Skepticism arises chiefly from the difficulty in assessing the level of awareness of the complex associations learned in classical implicit learning paradigms.

Here, we show that simple associations between colors and motion directions can be learned unconsciously. In each trial, participants had to report the motion direction of a patch of colored dots but unbeknownst to the participants, two out of the three possible colors were always associated with a given direction/response, while one was uninformative.

We confirm the lack of awareness by using several tasks, fulfilling the most stringent criteria. In addition, we show the crucial role of trial-by-trial feedback, and that both the stimulus-response (motor) and stimulus-stimulus (perceptual) associations were learned.

In conclusion, we demonstrate that simple associations between supraliminal stimulus features can be learned unconsciously, providing a novel framework to study unconscious learning.

Introduction

Unconscious learning can be defined as “learning without awareness, regardless of what sort of learning is being acquired” (Shanks & Stjohn, 1994). One can frame the current literature on unconscious learning along two dimensions: the first one determines whether the stimuli used during learning are supraliminal or subliminal, whereas the second dimension characterizes the complexity of the rules or associations to be learnt (see Fig. 1A). More specifically, the term "complex rule" is used here to refer to task structures that are composed of a large number of contingencies, such as in sequence learning tasks (in which numerous transition between successive key presses have to be learned), or artificial grammar tasks (in which a set of probabilistic rules drive the generation of grammatical strings). In opposition, simple rules can be defined as task structures composed of a small number of contingencies. Notably, there is not a precise separation between these two classes of rules, which rather define two extremes of a continuum. The use of supraliminal stimuli to induce learning of abstract, complex rules (top left corner in Fig. 1A) is the hallmark of implicit learning (Reber, 1967). Throughout the years, many experimental paradigms have provided reliable and replicable evidence of implicit learning, from artificial grammar learning (Reber, 1967) to sequence learning (Nissen & Bullemer, 1987), from control of complex systems (Berry & Broadbent, 1984) to statistical learning (Saffran, Johnson, Aslin, & Newport, 1999). In the example of the artificial grammar case, the complexity of the rules ensues from the presence of a large set of probabilistic associations that generate the strings. Moving to the opposite side of the theoretical space described in Figure 1A, subliminal stimuli used to learn simple associations set the

framework of what has been defined as subliminal learning (Clark & Squire, 1998; Clark, Manns, & Squire, 2002; Olsson & Phelps, 2004). In a typical subliminal learning paradigm, a cue, which is non-consciously perceived, predicts an outcome or prompts a response. In this case, the complexity is low since there is only one association driving the behavioral effect, although the stimuli are perceived subliminally. Looking at the top right corner of Figure 1A, we find the learning of subliminal stimuli associated to complex rules. This has been poorly investigated because of both methodological and interpretational issues. On the one hand, it is difficult to perform complicated associations between subliminal, rapidly presented stimuli (Atas, Faivre, Timmermans, Cleeremans, & Kouider, 2014); on the other hand, this kind of learning would not be of particular interest since it would not add any insightful perspective to the existing frameworks (Kido & Makioka, 2015). Finally, and rather surprisingly, there has been only a single attempt, to the extent of our knowledge, at studying directly the unconscious learning of supraliminal cues governed by simple rules (bottom left corner of Figure 1A). In this series of experiments from the late 80s, Lewicki and colleagues found that human participants could learn hidden covariations between the features of different stimuli in the absence of explicit awareness (P Lewicki, Hill, & Czyzewska, 1992; Pawel Lewicki, Czyzewska, & Hoffman, 1987; Pawel Lewicki, Hill, & Czyzewska, 1994). Despite the interest raised by this approach, the interpretation of these studies has been thoroughly criticized (D. Shanks & Stjohn, 1994) and, even more crucially, their findings have failed the test of independent replication (Hendrickx, De Houwer, Baeyens, Eelen, & Van Avermaet, 1997), leaving the question of the actual existence of this type of unconscious learning unaddressed.

Yet experimental support for the existence of unconscious learning of simple associations between conscious cues would provide decisive responses to enduring criticisms that have been formulated against the

existence of unconscious learning. Indeed, Shanks and colleagues have listed 4 criteria to fulfill in order to provide evidence in favor of unconscious learning: the sensitivity criterion regards the sensitivity of the measures of awareness; the information criterion suggests that the measure of awareness should probe the same information as the experimental task; the immediacy criterion imposes that the tests should be concomitant (or follow immediately) the experimental task; and finally the relevance criterion suggests that the measure of awareness should avoid any irrelevant information. The failure of the current literature in meeting these criteria suggested a substantial lack of evidence in favor of the existence of unconscious learning (Lovibond & Shanks, 2002; Newell & Shanks, 2014; D. Shanks & Stjohn, 1994).

To address this issue, we developed a simple motion direction discrimination task in which participants were asked to report the motion direction of a colored patch of dots; unbeknownst to them, there was an association between motion direction and 2 out of the 3 possible colors. In 10% of trials, participants were asked to report also the color of the patch together with the motion. In Experiment 1, we tested whether participants were able to learn this association. In Experiment 2, we specifically investigated the extent to which participants were aware of the relevant contingencies, addressing the 4 criteria suggested by Shanks and colleagues (Newell & Shanks, 2014; D. Shanks & Stjohn, 1994). In Experiment 3, we studied the role of feedback in such learning. Finally, in Experiment 4, we tested whether learning involved either sensory-motor or sensory-sensory associations.

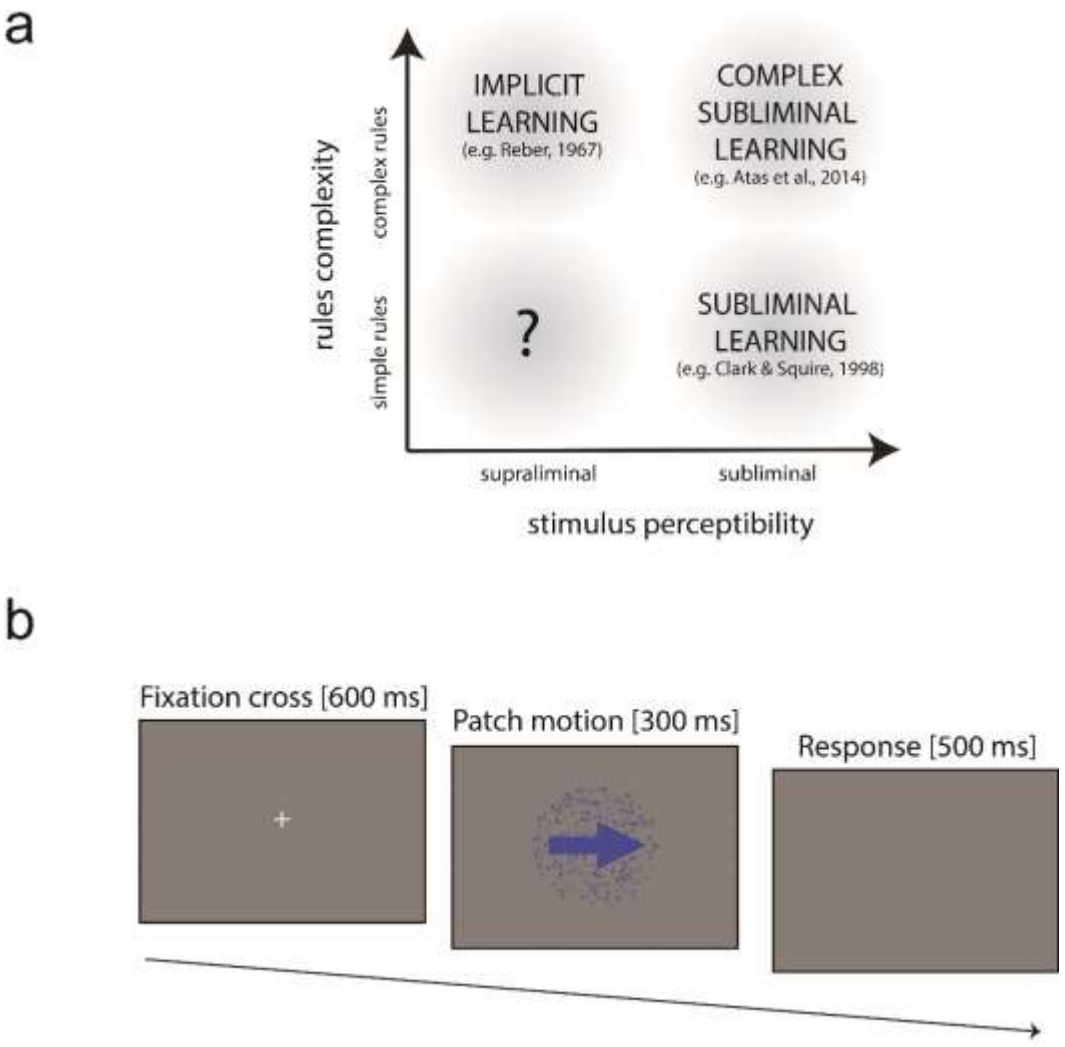


Fig.1. Schematic depiction of the unconscious learning framework (a) and experimental design (b). In the scheme (a), the x-axis represents the stimulus perceptibility (subliminal or supraliminal) and the y-axis represents the rules complexity (simple – complex rules). Within this space, we define 4 possible categories of unconscious learning paradigms. In the lower part (b), the experimental design is shown: following a fixation cross displayed for 600 ms, a patch of moving dots was displayed for 300 ms. The participants had 500 ms to provide a response.

Experiment 1

Participants

Fourteen healthy participants (7 females, mean age=24.2 years, SD=5.63) took part in the first experiment, receiving monetary compensation for their participation. We chose an *a priori* sample size of 14 subjects, since we had no prior data on which to base our initial estimate. In this and all subsequent experiments we stopped acquiring subjects when we reached the sample size planned before starting the experiments. All of them reported normal or corrected-to-normal vision. All experiments were carried out according to the Declaration of Helsinki and were approved by the Ethics Committee of the Université catholique de Louvain. Written informed consents were obtained from all the participants.

Procedure

The experiment took place in a quiet room, with the participants sitting comfortably on a chair in front of a 19" CRT screen with a 100Hz refresh rate. The distance between the screen and the chin support was 58 cm. The task was implemented using version 3.0.9 of the Psychotoolbox (Brainard, 1997) in Matlab 7.5 (The MathWorks, Inc. Natick, Massachusetts, United States).

At the beginning of each trial, a white cross was displayed at the center of the screen for 600 ms on a gray background (gray levels 0.7). A patch of moving dots was then displayed for 300 ms in the center of the screen, followed by a 500 ms blank screen, after which a new trial began (Fig. 1B). The stimulus was a patch of 2400 dots with a radius of 12°. Each

dot was updated every 10 ms. All the dots had a lifetime of 3 frames, during which they followed a straight line and were then displaced to a new random location. In order to minimize the retinal persistence effect and avoid the perception of the dot trajectory as lines on the screen, each dot was displayed every other frame, such that there were two interleaved streams of dots displayed alternately. The coherence of the dot motion, i.e. the percentage of dots moving coherently in the same direction, was kept constant during each block.

Participants were instructed to discriminate the motion direction of the dots. They could respond anytime from the patch onset until the end of the trial, by clicking on the left or right mouse buttons with their right hand to indicate a leftward or rightward motion, respectively. Auditory feedback was provided at the end of each trial, signaling a correct (high pitch) or incorrect (low pitch) response. Failure to provide a response was considered as an error.

The experiment consisted of 20 blocks of 60 trials each, and lasted around 45 minutes altogether (see Table 1). The task started with 7 blocks of “perceptual training” in which the stimulus consisted exclusively of white dots (gray levels 0.1) in order to determine the level of coherence that will be used for each individual for the rest of the experiment.. Indeed, during this training phase, the coherence of the patch was tuned block by block so as to maintain the response accuracy between 70% and 80%. The first block was always performed with 100% coherence patches, but in the following blocks, as soon as accuracy reached 90%, the coherence level was decreased by 15% (or by 20% if accuracy was 100%) in the subsequent block. Conversely, if accuracy fell below 70% in a block, coherence was increased by 10% in the subsequent one. In the following 13 blocks of the “unconscious learning” phase (see Table 1), colored patches were shown (red, green, or blue), and motion coherence was kept constant and equal to the coherence value of the

last block of the “perceptual training” phase. Crucially, and unbeknownst to participants, in the “unconscious learning” phase, colors and motion direction were associated: one color was always presented in association with leftward motion, another color was always associated with rightward motion, and the third color was equally likely to be associated to rightward and leftward patch motion (Table 2).

	Perceptual Training **(white dots)**	**Unconscious learning** **(colored dots)**	**Change in the associative structure** **(colored dots)**	**Unconscious learning** **(colored dots)**	**Explicit tasks** **(familiarity, generative, questionnaire)**	**Explicit Blocks** **(colored dots)**
EXP 1	7 blocks	8 blocks	2 blocks	3 blocks		
EXP 2	7 blocks	8 blocks			2 blocks	2 blocks
EXP 3	7 blocks	8 blocks	2 blocks	3 blocks		
EXP 4	8 blocks	8 blocks	2 blocks	2 blocks		

Table 1. Experimental designs of the four experiments.

COLOR	**MOTION**	**RESPONSE**
Color 1	Right	Right
Color 2	Left	Left
Color 3	R / L	R / L

Table 2. Associations between color, motion direction and response in experiments 1, 2 and 3.

The color-motion associations were pseudo-randomized across subjects; colors were pseudo-randomly interleaved every 3 trials (e.g. red-blue-green, green-red-blue, etc.), such that no color appeared more often than twice in a row. Importantly, the association between color and motion was discontinued during the 16th and 17th blocks and restored in the last three blocks (Table 1). Furthermore, to ensure that color information was actually processed by participants, they were asked to report the color of the patch in 10% of the trials, selected randomly. Participants responded by clicking with the mouse on one of three colored circles displayed on the screen. None of the participants reported explicit awareness of the association between colors and motion direction when questioned about it during debriefing at the end of the experiment.

Data analysis

The behavioral data were analyzed with Generalized Linear Mixed Models (GLMM) implemented in the SAS 9.3 Software (SAS Institute, Cary NC). We tested 2 GLMMs, considering accuracy (binary) and reaction times (normal) as dependent variables. In both models, the following independent factors were considered: BLOCK (from 8 to 20), ASSOCIATION (set to 1 if the color was associated with a motion direction, 0 otherwise), COLORS (three levels; included to account for possible effects of the color of the patch on motion discrimination irrespective of its association to a motion direction), and SUBJECT (to account for inter-subject differences). Both models were fitted using all the trials, avoiding any pre-processing.

Results

We found that participants' accuracy was higher in the trials in which the color of the dots provided information on the motion direction, as revealed by the factor ASSOCIATION in the GLMM analysis ($F(1,10698)=13.13$, $p<0.001$) (Fig. 2). The lack of significant ASSOCIATION x BLOCK interaction prevents us from concluding anything about the dynamics of the learning, but suggests, in accordance with previous studies (Turk-browne, Scholl, Chun, & Johnson, 2010), that such learning already occurs in the early phase of the experiment. The behavioral advantage induced by the predictive colors disappeared in blocks 16 and 17, during which the contingencies were disrupted, and quickly recovered in the subsequent blocks, after the association had been restored, as confirmed by a GLMM performed on data gathered from blocks 15^{th} to 18^{th}. This analysis revealed a significant effect of the factor ASSOCIATION ($F(1,4098)=11.68$, $p<0.001$), and of the BLOCK x ASSOCIATION interaction ($F(4,4098)=2.75$, $p=0.026$). A Tukey-corrected post-hoc analysis of this interaction revealed better accuracy for predictive colors in block 15^{th} ($t=-3.39$, $p<0.001$) and 18^{th} ($t=-2.59$, $p=0.009$), but not in blocks 16^{th} and 17^{th} ($t=-0.48$, $p>0.250$ and $t=-0.51$, $p>0.250$). Overall, these analyses confirmed that participants learned the associations between color and motion direction.

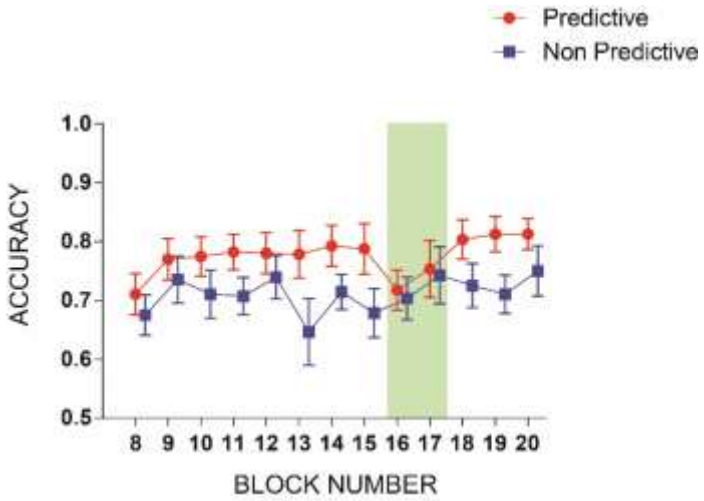


Fig.2. Performance data from Experiment 1. Error bars represent standard errors. The colors represent the two conditions (red: informative color condition, blue: non-predictive color condition). The violation of the association in block 16th and 17th is marked by the light green rectangle. The performance during the "perceptual training" phase with white patches (blocks 1 to 7) is not displayed.

Regarding the reaction times (RT), only the factor ASSOCIATION was close to be significant ($F(1,10533)=3.20$, $p=0.0736$), whereas all other factors and interactions were far from being significant (all $F<1.8$, $p>0.10$) (Fig. S1). This was expected, given that the motion signal was displayed for only 300 ms and that the participants were provided a very short period of time to respond, in order to emphasize the effect of the color-motion association on accuracy rather than on RT.

Discussion

Overall, this first experiment provides evidence in favor of robust learning of the color-motion association. In Experiment 2, we aimed at testing directly participants' awareness of the color-motion direction association by means of two tasks and one questionnaire.

Experiment 2

Participants and procedure

Twenty-three healthy participants (12 females, mean age=22.73 years, SD=2.59) participated in the second experiment for monetary compensation. Because we expected negative results in the awareness tasks, we increased the sample size to 24 in order to improve our statistical power. Since one of the subjects did not come on the day of the experiment, we finally acquired data on 23 subjects. All of them reported normal or corrected-to-normal vision. The first part of the experiment (i.e. "perceptual training" and "unconscious learning" phases) was exactly the same as in Experiment 1 (table 1). The second part was executed immediately after the first one and consisted of three tasks, performed in a pseudo-randomly order by the different subjects. The first task was a generation task consisting of two interleaved types of trials. In the first type of trials (generative color trials), a patch of white dots moved leftward or rightward for 300 ms, while in the second type of trials (generative motion trials), a static but colored patch was displayed for the same amount of time. Participants were asked to associate either a color to the white moving patch, or a motion direction to the colored but static dots; the response was provided by clicking with the mouse on the

selected color or motion direction. The entire block was composed of 60 trials (30 for each type). The other task was a familiarity test in which participants were asked to rate, from 1 to 10, the familiarity of displayed patches. All six possible combinations of colors and motion directions were included in this task, which thus also included color-motion combinations to which participants had never been exposed during the experiment. Overall, in this task, 42 trials were performed. The generative and familiarity tasks took about 4 minutes each. The third test was a short questionnaire with three questions: the first question inquired about any perceived difference between the rightward and leftward motion directions, the second one concerned colors, and the last question asked participants to indicate explicitly whether they had noticed any association between color and motion direction. Finally, participants were told about the association between colors and motion directions, and were asked to perform 2 additional blocks of the discrimination task while being now explicitly aware of the association.

The same analyses as in Experiment 1 were performed twice on two distinct datasets: once on all the subjects (n=23), and then only on those who did not provide a correct response in the questionnaire about the color/motion association ("implicit" group, n=18). Given that the results of these analyses unveiled a lack of effect, we computed the Bayes Factor to estimate the likelihood of the null hypothesis being true (Smith, 2001). The Bayes Factor (BF) can be used as an alternative way to test statistical hypothesis. It relies on the estimation of the probability of a statistical model (or hypothesis) given the observed data. One major advantage of this approach is that it allows researchers to estimate the validity of the null hypothesis, in comparison to alternative hypotheses. Indicatively, a BF between 0.3 and 3 suggests a lack of sensitivity. A BF below 0.3 or above 3 provides strong evidence in support for the alternative hypothesis, or for the null hypothesis, respectively. In order to compute the Bayes Factor (BF), we

compared the Bayesian Information Criterion (BIC) estimated from each model with and without the explanatory variable (Masson, 2011; Smith, 2001).

Results

We confirmed that the discrimination accuracy was higher in trials with predictive colors than with the control color (Fig.3) (ASSOCIATION: $F(1,10664)=4.58$, $p<0.001$). We also found a progressive increase in accuracy across blocks (BLOCK: $F(7,154)=4.58$, $p<0.001$) but no significant ASSOCIATION x BLOCK interaction ($F(7,10664)=0.69$, $p>0.250$). Similar results were obtained when restricting the analysis to the "implicit group" only (n=18; ASSOCIATION: $F(1,8344)=17.44$, $p<0.001$; BLOCK: $F(7,119)=3.23$, $p=0.003$, interaction $F(7,8344)=0.84$, $p>0.250$). Regarding the RT (Fig. S1), only the factor ASSOCIATION was significant, with faster responses being associated with predictive colors (ASSOCIATION: $F(1, 10499)=25.93$, $p<0.001$; BLOCK: $F(7,154)=1.23$, $p>0.250$; interaction $F(7, 10499)=1.00$, $p>0.250$; implicit group, n=18: ASSOCIATION: $F(1,8200)=10.77$, $p=0.001$; BLOCK: $F(7,119)=1.21$, $p>0.250$; interaction $F(7, 8200)=0.89$, $p>0.250$).

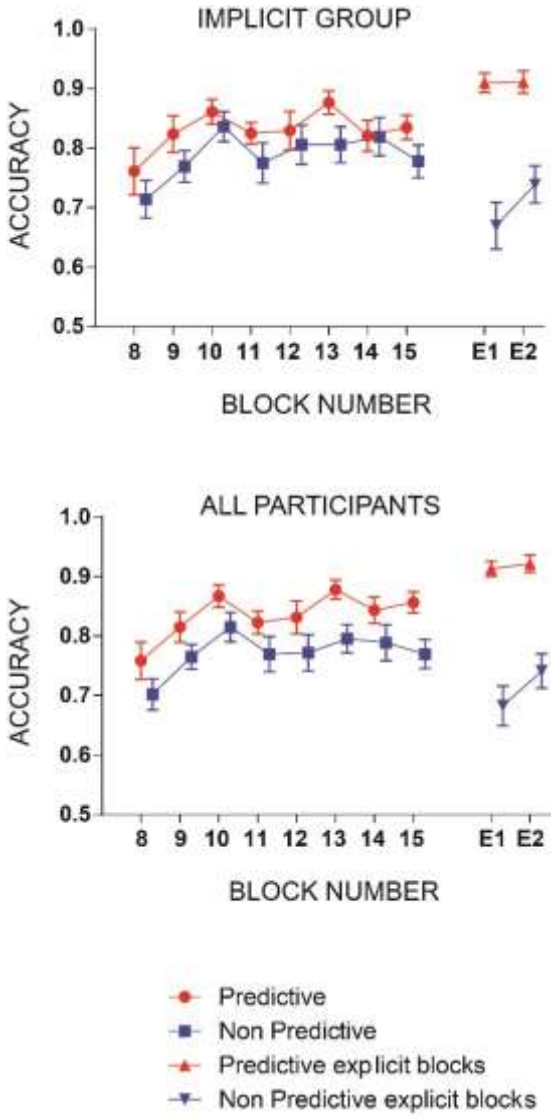


Fig.3. Accuracy data from Experiment 2. In the upper part, the data from the group of subjects who did not provide the correct associations in the questionnaire is shown (implicit group, n=18), whereas all the participants are included in the data shown in the lower panel (n=23). Error bars represent standard errors. E1 and E2 are two blocks in which participants were instructed about the associations. The data from the "perceptual training" phase with white patches (from 1 to 7) is not displayed in the figure.

We compared the accuracy and RT of all the participants in the last 2 blocks of the "unconscious learning" phase (blocks 14th and 15th) with the data obtained in the "explicit blocks" performed at the end of the experiment (E1 and E2, Fig.3), after the color-motion association had been explicitly revealed to the subjects (see Table 1). As expected, we found a significant effect of the factor ASSOCIATION on accuracy ($F(1,5426)=145.95$, $p<0.001$), but not of the factor accounting for the awareness of the subjects (EXPLICITNESS: $F(1,22)=1.59$, $p=0.220$); however, the interaction between these two factors was highly significant ($F(1,5426)=45.86$, $p<0.001$). Similarly, the analysis on RT revealed an effect of all the factors and their interaction (ASSOCIATION: $F(1, 5326)=168.94$, $p<0.001$; EXPLICITNESS: $F(1, 22)=17.41$, $p<0.001$; interaction : $F(1, 5326)=89.31$, $p<0.001$). These changes in both RT and accuracy following awareness of the association suggests a drastic change in strategy, in which participants started presumably to base their decision primarily on the color information.

Concerning the questionnaire, only 5 subjects out of 23 explicitly reported the color-motion association. Another subject reported only one correct color-motion association, and two other subjects reported incorrect color-motion associations. All other subjects reported having noticed no systematic association whatsoever.

Regarding the generative tasks, in trials in which participants were asked to associate a motion direction to a color (Fig. 4A), we considered as dependent variables the motion direction chosen by the participants, and, as factors, the COLORS and TASK-ORDER (categorical variable, accounting for whether the task was performed before or after the questionnaire). We found no significant main effects for either group (all subjects, n=23: COLORS: $F(2,659)=0.29$, $p>0.250$; TASK-ORDER: $F(1,659)=0.29$, $p>0.250$; "implicit group", n=18: COLORS: $F(2,48)=0.12$, $p>0.250$; TASK-ORDER: $F(1,465)=0.79$,

p=0.3752), but found a significant interaction for the group including all the subjects (all subjects: F(2,659)=3.07, p=0.0474), suggesting that, after the questionnaire, subjects were more likely to associate the informative colors to the correct motion direction, as revealed by a significant post hoc analysis (difference between left-motion associated color and right-motion associated color: t=4.03, Tukey-Kramer corrected p<0.001; between second informative color and non-informative color t=6.06, Tukey-Kramer corrected p<0.001). The significant interaction was, however, not observed in the "implicit group" (F(2,465)=0.53, p>0.250). To confirm these negative findings in the implicit group, we computed a BF by comparing the BIC obtained from each model with and without the explanatory variable COLORS (Masson, 2011; Smith, 2001). The results confirm a lack of effect of the COLORS factor for the implicit group (implicit group: BF= 33.11 p<0.03). The second generative task (Fig. 4B), in which participants associated a color to a given motion direction, revealed no effects in any factors or interaction for both groups (all the subjects: MOTION-DIRECTION: F(1,42)=0.58, p>0.250; TASK-ORDER: F(1,643)=0.23, p>0.250); interaction: F(1,643)=0.91, p>0.250); implicit group: MOTION-DIRECTION: F(1,32)=1.32, p=0.2599; TASK-ORDER: F(1, 503)=0.01, p>0.250); interaction: F(1, 503)=2.10, p=0.1480). The BF confirmed the lack of result for the factor MOTION-DIRECTION and its interaction for both groups (all subjects: BF= 20.08 p<0.05, implicit group: BF= 90.01 p<0.02).

Considering the familiarity task (Fig. 4C), we did not find any significant effect for the factors ASSOCIATION (correct association, incorrect association or control color; F(2,48)=0.14, p>0.250), TASK-ORDER (F(1,688)=2.09, p=0.1491), and their interaction (F(2,688)=0.46, p>0.250) for the implicit group (n=18). The Bayes factor confirmed the lack of effect for the ASSOCIATION factor in the model (BF= 38.11, p<0.03). When considering

all subjects, we found a significant effect of the interaction between the factors (F(2,881)=3.31, p=0.0370), but nor a main effect of the task order, neither of the association factor (ASSOCIATION: F(2,63)=1.14, p>0.250; TASK-ORDER: F(1,881)=0.84, p>0.250). This interaction shows a significant difference between the first informative color and the non-informative color, as revealed by a post hoc a

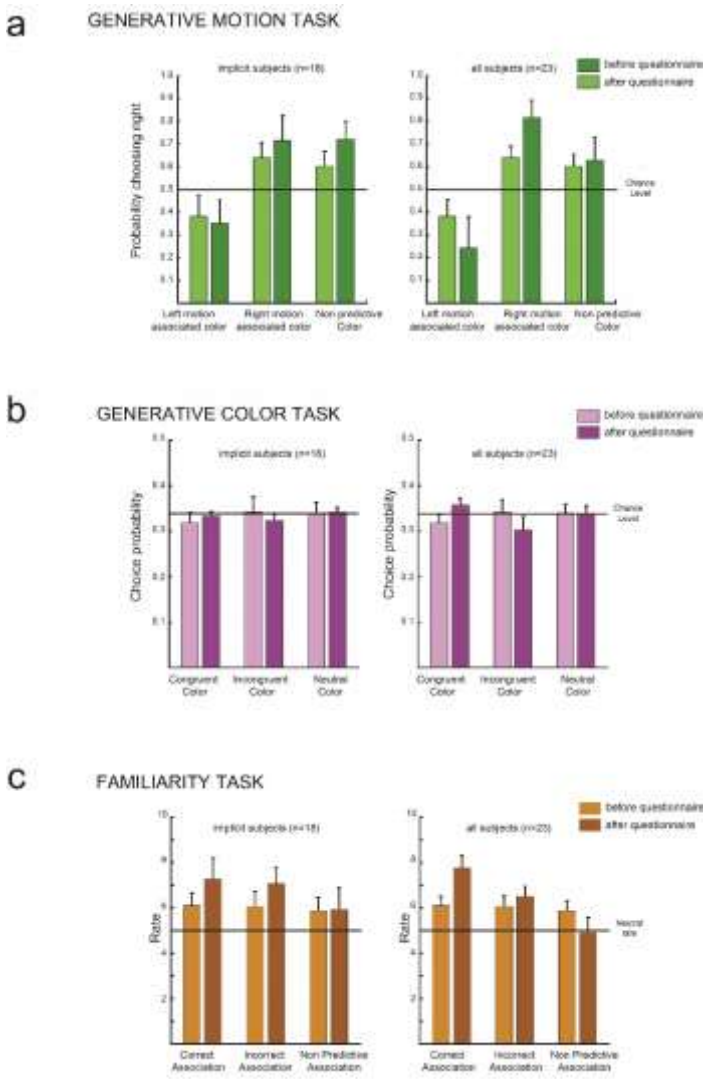


Fig.4. Results from the awareness testing tasks. The data from the participants who did not respond correctly to the questionnaire (implicit group, n=18) is shown on the left, whereas all participants are included in the dataset shown on the right (n=23). Participants were split based on whether they performed the awareness testing tasks before (light colors) or after (darker colors) the questionnaire. Results from the generative motion task (a): on the y-axis, the probability of choosing rightward motion is shown. The results from the generative color task are displayed in the middle part (b): Values on the y-axis represent the probability of choosing either of the 3 color conditions displayed along the x-axis (congruent, incongruent or neutral color). In the last part (c) the results from the familiarity task are shown: on the y-axis, the average familiarity ratings associated to the three different types of patches are shown (correct predictive association, incorrect association or non-predictive association).

Moreover, to confirm that all the tests measured the same variable (i.e. awareness), we correlated the results of the generative and familiarity tasks. In order to do so, we computed the Euclidean distance from the rates (or proportions of answers in the generative tasks) provided by each subject to the optimal 'explicit' behavior, such that small distances would reflect more awareness of the associations. The results confirmed a correlation between the two generative tasks (Pearson correlation, $r=0.53$, $p<0.01$) and between the familiarity task and the second generative task (assigning color to motion, $r=0.5823$, $p<0.01$), but not between familiarity and the first generative task (assigning motion to color, $r=0.1648$, $p>0.25$). Altogether, these correlational analyses confirmed that all the tests measured the awareness of the learnt associations.

Discussion

In Experiment 2, we found that when restricting the analyses to the group of subjects that failed to report the association between color and

motion, the learning of this association was still robust in spite of the fact that none of the awareness tasks showed significant results. This demonstrates that even though some participants gained explicit awareness of the association, learning took place in the absence of awareness for most of them.

The difference between the last two blocks of the implicit phase and the two explicit blocks highlighted a change in the strategy adopted by the participants: when the association became fully explicit, participants seem to focus mostly on the color feature rather than on the motion. Such difference strongly suggests that if the participants had had explicit knowledge of the association during the main experiment, their behavior should have been similar to that exhibited in the explicit blocks. These results thus provide strong evidence for the unconscious nature of the learning.

In Experiment 3, we attempted to gain insight into the basic learning mechanisms at play in our task. We wondered whether participants learned the association through reinforcement or Hebbian learning. To address this question, we investigated the role of feedback in the learning of the color-motion association.

Experiment 3

Participants and procedure

Fourteen healthy participants (9 females, mean age=23.42 years, SD=1.74) participated in the third experiment for monetary compensation.

All of them reported normal or corrected-to-normal vision. The experimental design was the same as in Experiment 1, except that no auditory feedback was provided during the "unconscious learning" phase of the experiment (see Table 1). Therefore, we used the same sample size as in the first experiment. Auditory feedback was still provided during the first 7 blocks of training with the white dots patches in order to obtain a level of response accuracy between 70 and 80% as in Experiment 1.

Results

As in the previous experiment, we analyzed accuracy and RT in 2 GLMMs. Regarding accuracy (Fig.5), we did not find any significant effect: ASSOCIATION ($F(1,10697)=2.41$, $p=0.1204$), BLOCK: $F(12,156)=0.87$, $p>0.250$, COLORS: $F(2,25)=1.29$, $p>0.250$, BLOCK x ASSOCIATION $F(12,10697)=1.15$, $p>0.250$). In order to test specifically the lack of effect of the factor ASSOCIATION, we computed the BF on the basis of the BIC obtained from the models with and without this factor. This analysis confirmed the lack of effect of the factor ASSOCIATION on the response accuracy ($BF=2.8478 \times 10^{11}$, $p<0.0001$). The very large value of the BF, along with the highly significant estimated p-value, confirmed that the sample size provided enough statistical power to properly test the hypothesis. Concerning the RT (Fig.S1), we found a significant effect of COLORS ($F(2,25)=4.04$, $p=0.0301$), while the other effects were not significant (ASSOCIATION $F(1,10254)=0.05$, $p>0.250$; BLOCK: $F(12,156)=0.85$, $p>0.250$; BLOCK x ASSOCIATION $F(12,10254)=1.35$, $p=0.1850$)). The color effect revealed that overall the subjects were slower in detecting the blue color, irrespective of its association with the motion direction.

We then compared the response accuracy in Experiment 1 and 3, by means of a GLMM with COLORS, ASSOCIATION, BLOCK, SUBJECT and EXPERIMENT as factors. Interestingly, the results revealed a significant effect of the factor ASSOCIATION ($F(1,21615)=33.68$, $p<0.001$), BLOCK ($F(12,156)=1.91$, $p=0.0375$), EXPERIMENT ($F(1,21615)=42.44$, $p<0.001$) and the interaction EXPERIMENT x ASSOCIATION ($F(1,21615)=17.55$, $p<0.001$), thus confirming the importance of the auditory feedback not only in performing the task, but also in the unconscious learning of the association.

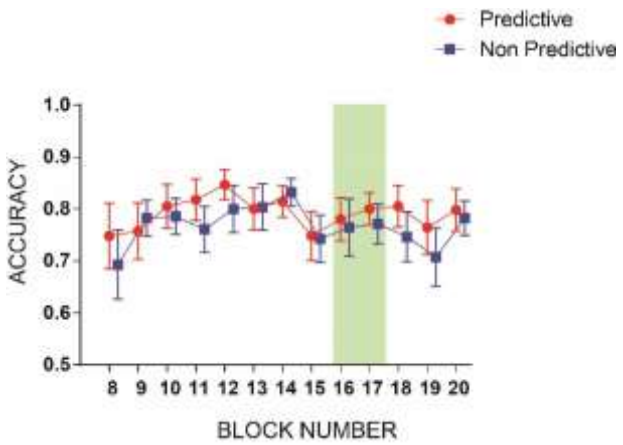


Fig.5. Performance data from Experiment 3. The informative/non-informative colors are represented in red/blue. Error bars represent standard errors. The violation of the association in block 16th and 17th is marked by the light green rectangle. The data from the "perceptual training" phase with white patches (from 1 to 7) is not displayed.

Discussion

Experiment 3 indicates that when no feedback on response accuracy is provided, learning fails to take place. This suggests that this type of unconscious learning relies on reinforcement learning mechanisms (Dayan & Balleine, 2002; Sutton & Barto, 1998), rather than on Hebbian-like associative learning between concurrent features of the stimuli (Munakata & Pfaffly, 2004). Since reinforcement learning is usually studied in the context of stimulus-response associations (Dayan & Balleine, 2002), these findings suggested that participants may have learned to associate the color of the stimulus with the response dictated by the corresponding motion direction, as opposed to associating directly the color with the motion. We tested this hypothesis in Experiment 4.

Experiment 4

Participants and procedure

Twenty-four healthy participants (14 females, mean age=27.04, SD=4.78) joined the fourth experiment for monetary compensation. All of them reported normal or corrected-to-normal vision. The sample size was chosen from an expected effect size of 0.4676% (estimated from experiment 1) and aiming for a power of 0.80 and an alpha of 0.05. The experimental design was similar to that of Experiment 1, except as detailed below.

The “perceptual training” phase with the white dots lasted for 8 blocks, followed by 12 blocks of “unconscious learning” phase with colored patches and fixed coherence levels (see Table 1). Auditory feedback was provided

trial by trial in every block. The motion of the dots was directed upward or downward, and the participants were instructed to respond left or right to indicate the motion direction. The rule linking the response to the motion direction changed in every block, so that if the upward motion was associated with the left response in one block, the left response was associated with the downward motion in the subsequent block. This alternation of the rules was reminded to the subjects both verbally by the experimenter and visually at the beginning of each block by displaying a message for 5 seconds on the computer screen. Importantly, the subjects were randomly separated in two groups: in the first group, the association between color and motion was kept constant during the whole experiment, while in the second group, it was the association between color and response which was kept constant (see Table 3). In this way, only one association (either color-motion direction or color-response side) was maintained during the whole experiment. Similarly to Experiment 1, these associations were discontinued in blocks 17 and 18 (see Table 1).

Color-Motion group:

Block i			Block i+1		
COLOR	COLOR	MOTION	ANSWER	MOTION	MOTION
Color 1	**Color 1**	**Upward**	Right	**Upward**	**Upward**
Color 2	**Color 2**	**Downward**	Left	**Downward**	**Downward**
Color 3	Color 3	U / D	R / L	U / D	U / D

Color-Response group:

Block i			Block i + 1		
COLOR	MOTION	ANSWER	COLOR	MOTION	ANSWER
Color 1	Upward	**Right**	**Color 1**	Downward	**Right**
Color 2	Downward	**Left**	**Color 2**	Upward	**Left**
Color 3	U / D	R / L	Color 3	U / D	R / L

Table 3. Associations between color, motion direction and response in Experiment 4.

We performed 2 GLMMs considering as dependent variables either the accuracy or the RT. The only main difference with the previous analyses was the independent factor named GROUP, which indicated to which group the subject was assigned (either color-motion association or color-response association).

Results

The first GLMM performed on accuracy (see Fig.6) revealed a significant effect of the factors BLOCK ($F(11,198)=4.07$, $p<0.001$), ASSOCIATION ($F(1,14098)=8.16$, $p=0.0043$), and their interaction ($F(11,14098)=3.12$, $p<0.001$). A post-hoc analysis of the interaction revealed a significant effect of ASSOCIATION for blocks 9, 12, 13, 15 and 16 (all $t<-2.10$, Tukey-Kramer corrected $p<0.03$). Surprisingly, the post-hoc analysis revealed also a significant difference in block 18, the second block in which the rules were violated, but highlighting a higher accuracy for the non-

informative color (t=2.11, p=0.0347). In the last two blocks, when the rules were restored, no difference emerged from the post-hoc analysis (block 19: t=-0.88, p=0.3787, block 20: t=0.26, p>0.250). Crucially, no difference between the two groups emerged (GROUP: F(1,14098)=0.00, p>0.250; GROUP x ASSOCIATION F(1,14098)=1.98, p=0.1592; GROUP x ASSOCIATION x BLOCK F(11,14098)=0.66, p=0.7751), indicating that both groups learnt equally well. The BF comparing the model with and without the factor GROUP confirmed this lack of significant effect (BF= 2.1138×10^{18}, p<0.0001). Regarding the analysis of the RT (Fig.S1), only the BLOCK factor and the BLOCK x GROUP interaction revealed a significant effect (BLOCK: F(11,198)=10.45, p<0.001, BLOCK x GROUP F(11,13832)=1.83, p=0.0438); no effect of the factor ASSOCIATION (F(1,13832)=1.83, p=0.1758) or GROUP (F(1,13832)=2.71, p=0.0996) was revealed. Regarding the BLOCK x GROUP interaction, none of the pairwise post-hoc comparisons showed a significant difference between the groups, suggesting that the effect was driven by different trends in the two groups, with the first group reducing its reaction time abruptly in the 5th block (Fig.S1). Finally, we compared the results from Experiments 1 and 4, adding to the GROUP variable a level accounting for the data from Experiment 1. GROUP was thus composed of three levels: two for each group of Experiment 4 and one for Experiment 1. No difference between the groups was revealed by the GLMM (GROUP x ASSOCIATION: F(2,28043), p>0.25), as confirmed by the BF analysis performed comparing the BIC of the model with and without the interaction GROUP x ASSOCIATION (BF=1.406×10^{7} p<0.001).

Discussion

In contrast to our expectation, we found that both the sensory-sensory and the sensory-motor associations were learnt equally well, with no difference between the groups. This suggests that the mechanisms involved in this type of associative learning are general rather than modality-specific, and that the events whose association gets unconsciously learned do not have to be perfectly concurrent in time, since the response followed the color by a few hundreds of milliseconds.

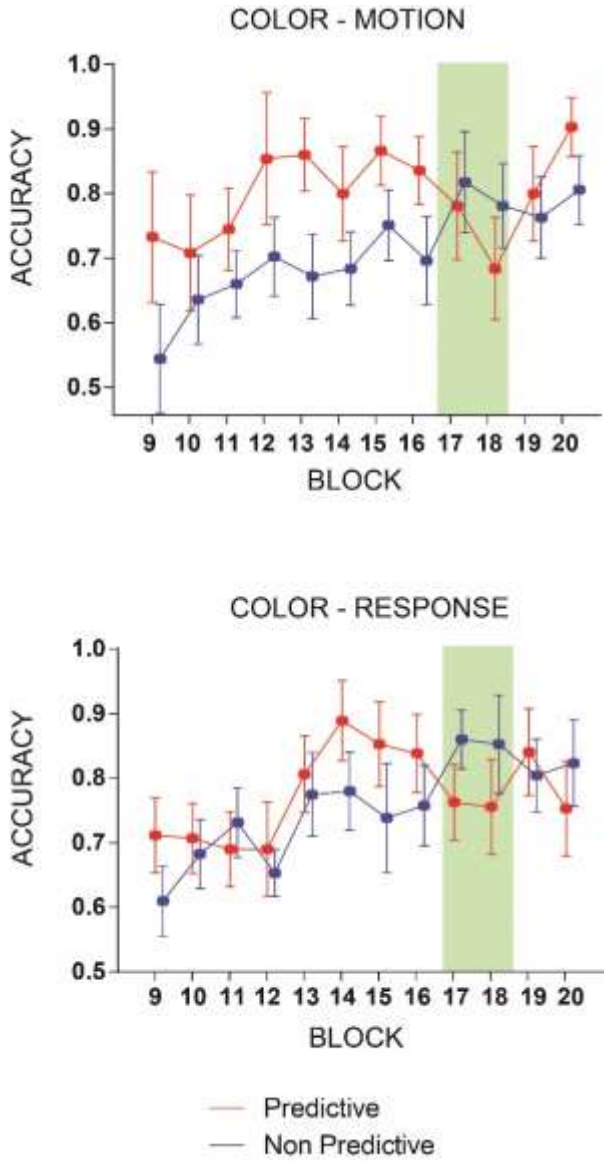


Fig.6. Performance in Experiment 4. In the upper panel, the data from the group in which the *color-motion* association was kept constant (n=12) is shown, and in the lower panel the data from the group in which the *color-response* association was kept constant is represented (n=12). The red and

blue dots correspond to the informative and non-informative colors, respectively. Error bars represent standard errors, and violation of the association in block 16th and 17th is marked by the light green rectangle. The data from the “perceptual training” phase with white patches (from 1 to 8) is not displayed in the figure.

General Discussion

In the present study we investigated whether participants can learn unconsciously a direct association between supraliminal features of task-relevant stimuli. The results showed robust and reliable learning of the association. Importantly, in Experiment 2 we directly tested the degree of awareness of such learning in light of the 4 criteria suggested by Shanks and Newell in 1994 (Dawson & Reardon, 1973; Newell & Shanks, 2014; D. Shanks & Stjohn, 1994). Briefly, these criteria require that the assessment of awareness should be devoid of biasing factors not relevant for the behavioral measure (reliability), that they should be performed immediately after the experiment (immediacy), and under optimal retrieval conditions (sensitivity; Newell & Shanks, 2014). The fourth crucial criterion is the relevance criterion (D. Shanks & Stjohn, 1994), which stipulates that the test of awareness should target the very same information that drives changes in behavior. In our design, given the simple nature of the learned association, it is easy to fulfill both the reliability and relevance criteria. Whereas in standard implicit learning paradigms, participants may potentially achieve success by exploiting information that the test of awareness fails to probe, in our case, the simplicity of the contingency excludes this possibility. Indeed, only the

learning of the association between the color of the dots and the direction of their motion can lead to the improvement in accuracy that we observed. Further, we extensively probed participants' conscious knowledge of this association directly by means of generation and familiarity tasks, and through a questionnaire. Regarding the sensitivity criterion, our Bayesian approach made it possible to convincingly conclude (Bayes factors > 30 (Dienes, 2011)) that our null findings can be interpreted as offering support for the absence of differences rather than as resulting from a lack of sensitivity (Vadillo, Konstantinidis, & Shanks, 2015). Finally, regarding the immediacy criterion, we administered the tests as soon as the learning phase ended, so reducing the effect of interference or forgetting as much as possible. It is noteworthy that, during the generative task, the group who did not report the correct association in the questionnaire performed at random when coupling the colors with the motion directions. This is a substantive finding because it suggests that the color-motion association is used only in the narrow context of the task in which it was learned and cannot be transferred to different task-sets (Graf & Schacter, 1985; D R Shanks, Johnstone, & Staggs, 1997).

We believe that our study provides the first demonstration of unconscious learning of simple associations, thereby filling the gap in the theoretical framework illustrated in Figure 1. Previous studies in the framework of delay conditioning (Clark & Squire, 1998; Clark et al., 2002), fear conditioning (Maren, 2001) or in the context of the relationship between unconscious processing and perceptual load (Bahrami, Carmel, Walsh, Rees, & Lavie, 2008; Carmel, Saker, Rees, & Lavie, 2007) could be viewed as providing already indirect evidence for unconscious learning of supraliminal stimuli. However, these studies did not test systematically the awareness of the learned association, thus failing to address the criteria discussed above (D. Shanks & Stjohn, 1994). One of the possible reasons for this lack of previous demonstration of the unconscious learning of simple

associations could be that designing an experiment in which the association between perceivable stimulus features is unconsciously learnt was quite challenging. On the one hand, the predictive cue should be perceived and actively processed by the participants in order to affect their behavior (Jiang & Chun, 2001; Jimenez & Mendez, 1999); on the other hand, the contingency with the other feature should remain implicit. The balance between these two extremes was difficult to obtain. One important aspect of our task is that the color had to be processed actively by the participants because of the secondary task, which consisted in reporting the color of the patch in 10% of the trials. Nevertheless, participants failed to perceive explicitly the association, possibly because of what can be defined as 'a change of narrative': the secondary task related to the predictive cue is a sufficiently convincing justification for the presence of the colors in the task, such that participants do not have to search for an explanation, which would eventually lead them to figure out the association (David R Shanks, 2003). Future experiments should further investigate whether this interpretation holds true. Furthermore, the presence of the secondary color task allows us to exclude inattentional amnesia as an alternative interpretation (Wolfe, 1999), since it compelled the participants to direct actively their attention to the colors during the task.

Whereas in Experiment 1 and 2, we provide evidence in favor of the learning without awareness of the color-motion association, in Experiment 3 and 4 we investigate the possible mechanisms involved in this type of learning. Specifically, Experiment 3 showed that auditory feedback was necessary for learning to occur, thus evoking reinforcement learning (Niv, 2009), while Experiment 4 showed that both the color-motion and the color-response associations were learnt. Feedback can, in some cases, be processed as a reward signal and induce a phasic dopaminergic response (Hyman, Malenka, & Nestler, 2006) which would then reinforce the circuits

that link the color with the motion features or the color with the response representation. This interpretation concurs with a recent review on the pharmacology of implicit learning suggesting a similar link with dopaminergic systems (Uddén, Folia, & Petersson, 2010).

Conclusion

In conclusion, we provide a novel and robust experimental design that can be used to investigate unconscious associative learning, and we have begun to decipher its basic mechanistic features. However, many important questions remain unanswered. For instance, we found a large inter individual variability in both the magnitude of the learning and the level of awareness of the association. Exploring the extent to which attention, working memory capacity or cognitive control (Stillman, Feldman, Wambach, Howard, & Howard, 2014) are involved in this variability is an important goal for further research.

References

Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., & Kouider, S. (2014). Nonconscious learning from crowded sequences. *Psychological Science*, *25*(1), 113–9. doi:10.1177/0956797613499591

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Unconscious orientation processing depends on perceptual load. *Journal of Vision*, *8*(3), 12.1–10. doi:10.1167/8.3.12

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge, (June 2015), 37–41. doi:10.1080/14640748408402156

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4),

433–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952

Carmel, D., Saker, P., Rees, G., & Lavie, N. (2007). Perceptual load modulates conscious flicker perception. *Journal of Vision*, *7*, 14.1–13. doi:10.1167/7.14.14

Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences*. doi:10.1016/S1364-6613(02)02041-7

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: the role of awareness. *Science*, *280*(5360), 77–81. doi:10.1126/science.280.5360.77

Dawson, M. E., & Reardon, P. (1973). Construct validity of recall and recognition postconditioning measures of awareness. *Journal of Experimental Psychology*, *98*(2), 308–315. doi:10.1037/h0034372

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*. doi:10.1016/S0896-6273(02)00963-7

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. doi:10.1177/1745691611406920

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*(3), 501–518. doi:10.1037/0278-7393.11.3.501

Hendrickx, H., De Houwer, J., Baeyens, F., Eelen, P., & Van Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 201–220. doi:10.1037/0278-7393.23.1.201

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. *Annual Review of Neuroscience*, *29*, 565–98. doi:10.1146/annurev.neuro.29.051605.113009

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning, (4), 1105–1124. doi:10.1080/0272498004200051

Jimenez, L., & Mendez, C. (1999). Which Attention Is Needed for Implicit Sequence Learning ?, *25*(1), 236–259.

Kido, K., & Makioka, S. (2015). Serial order learning of subliminal visual stimuli: evidence of multistage learning. *Frontiers in Psychology*, *6*(February), 1–13. doi:10.3389/fpsyg.2015.00076

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 523–530. doi:10.1037/0278-7393.13.4.523

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *The American Psychologist*, *47*(6), 796–801. doi:10.1037/0003-066X.47.6.796

Lewicki, P., Hill, T., & Czyzewska, M. (1994). Nonconscious indirect inferences in encoding. *Journal of Experimental Psychology. General*, *123*(3), 257–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7931091

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology. Animal Behavior Processes*, *28*(1), 3–26. doi:10.1037/0097-7403.28.1.3

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, *24*, 897–931. doi:10.1146/annurev.neuro.24.1.897

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing, 679–690. doi:10.3758/s13428-010-0049-5

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science*. doi:10.1111/j.1467-7687.2004.00331.x

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behavioral and Brain Sciences*, *37*, 1–19. doi:10.1017/S0140525X12003214

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. doi:10.1016/0010-0285(87)90002-8

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. doi:10.1016/j.jmp.2008.12.005

Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828. doi:10.1111/j.0956-7976.2004.00762.x

Reber, A. S. (1967a). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. doi:10.1016/S0022-5371(67)80149-X

Reber, A. S. (1967b). Implicit learning of artificial grammars. Retrieved

from http://philpapers.org/rec/REBILO

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Shanks, D. R. (2003). Attention and awareness in “implicit” sequence learning. In *Attention and implicit learning. Advances in Consciousness Research, vol. 48* (pp. 11–42).

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *50*(1), 216–252. doi:10.1080/713755680

Shanks, D., & Stjohn, M. (1994). Characteristics of Dissociable Human Learning-Systems, (February 2010). doi:10.1017/S0140525X00035032

Smith, J. M. B. and A. F. M. (2001). Bayesian Theory. *Measurement Science and Technology*. doi:10.1088/0957-0233/12/2/702

Stillman, C. M., Feldman, H., Wambach, C. G., Howard, J. H., & Howard, D. V. (2014). Dispositional mindfulness is associated with reduced implicit learning. *Consciousness and Cognition*, *28*(1), 141–150. doi:10.1016/j.concog.2014.07.002

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *9*, 1054. doi:10.1109/TNN.1998.712192

Turk-browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning, *134*(4), 552–564. doi:10.1037/0096-3445.134.4.552

Turk-browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2010). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness, *21*(10), 1934–1945. doi:10.1162/jocn.2009.21131.Neural

Uddén, J., Folia, V., & Petersson, K. M. (2010). The neuropharmacology of implicit learning. *Current Neuropharmacology*, *8*, 367–381. doi:10.2174/157015910793358178

Vadillo, M. a., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-015-0892-6

Wolfe, J. M. (1999). Inattentional Amnesia Jeremy M. Wolfe. *Fleeting Memories: Cognition of Brief Visual Stimuli*.

Supplementary figure

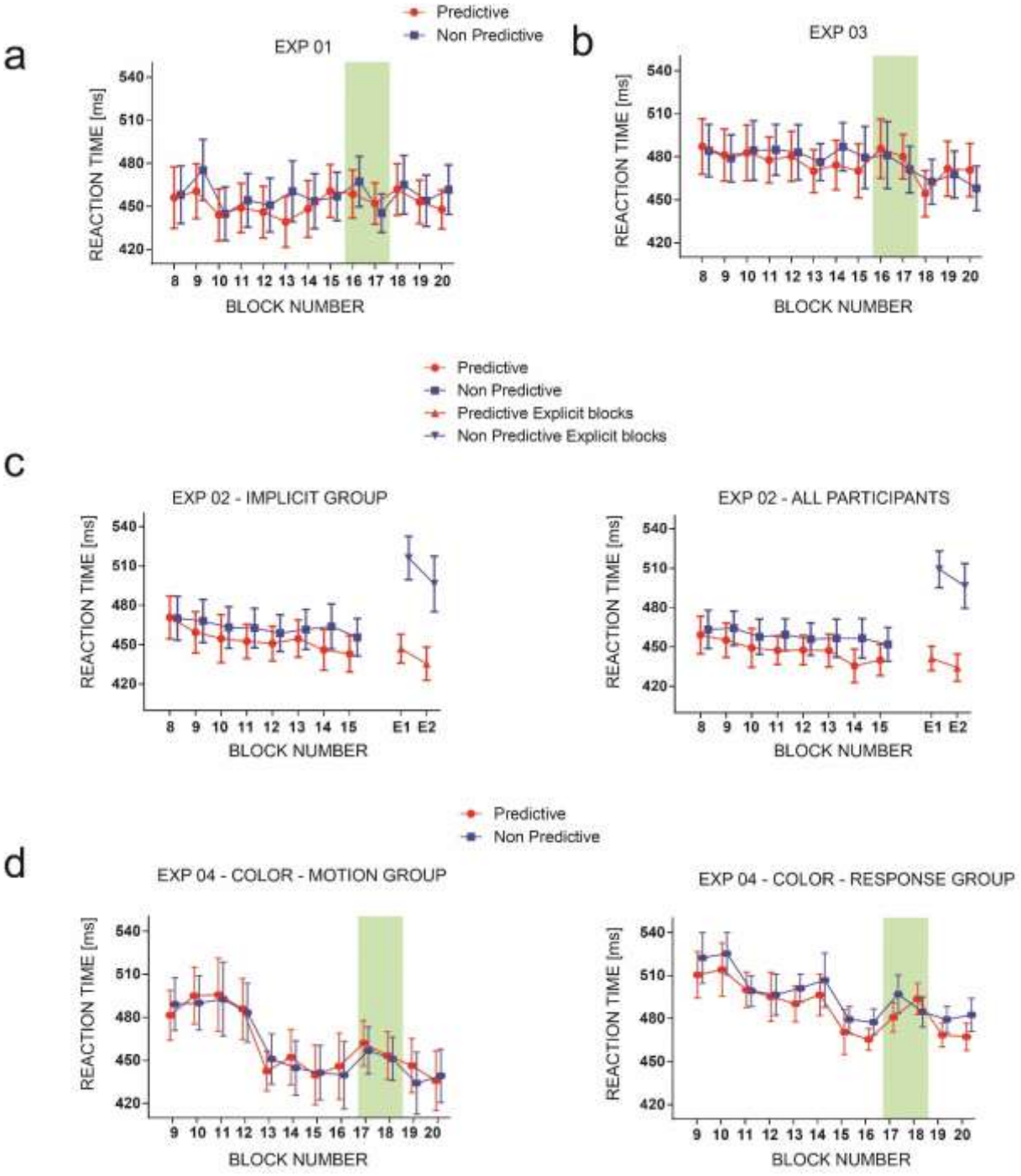


Fig.S1. Reaction time data from Experiments 1 to 4. Bars represent standard errors. The data from the "perceptual training" phase with white patches is not displayed in the figure. Panels (a) and (b) illustrate the RT data in Experiment 1 and 3, respectively. Reaction time data from Experiment 2 is shown in panel (c), with, on the left, data from the group of subjects who did not provide the correct associations in the questionnaire, and on the right, the data from all the participants. E1 and E2 correspond to the blocks in which participants were instructed about the associations. Panel (d) shows the data from Experiment 4, with, on the left, the group in which the *color-motion* association was kept constant (n=12), and on the right the data from the group in which the *color-response* association was maintained throughout the experiment.

Chapter III: Unconscious learning affects eye movements in a perceptual decision making task

This chapter is a modified version of an article in preparation:

Alamia, A., Solopchuk,O. & Zenon, A. “Unconscious learning affects eye movements in a perceptual decision making task”.

Abstract

Visual attention allows relevant information to be selected for further processing. It is most often studied in contexts in which the visual features that are relevant for the task are explicit. Visual features can also bias decision-making unconsciously, but the influence of these unconscious biases on the allocation of visual attention remains unclear. In this study, we explored attentional allocation during a motion discrimination task with unconscious cues. Participants were instructed to report the motion direction of 2 colored patches of dots. Unbeknown to participants, dot colors were sometimes informative of the correct response. We found that subjects learnt the associations between colors and motion direction but failed to report this association during the questionnaire filled at the end of the experiment, confirming that learning remained unconscious. The eye movement analyses revealed that attention was preferentially allocated to the most informative stimuli, indicating that visual attention can be drawn to unconscious sources of information.

.

Introduction

Attention is a mechanism for allocating cognitive resources to relevant stimuli (Desimone & Duncan, 1995). The highest priority for allocating attention is thought to be associated to the stimulus that maximizes the expected information gain (i.e. provide the most information given the stimuli and its context; Manohar & Husain, 2013; Summerfield & Egner, 2009; Vossel, Geng, & Friston, 2014). Attentional selection can be driven by stimulus- and context-specific features that make a given visual object stand out from its surrounding (Itti, 2005). This type of attention has been coined stimulus-driven, or bottom-up attention. On the other hand, selective attention can be driven to stimuli, features, or spatial location which are especially relevant with respect to the task that we are performing, an instance called goal-directed or top-down attention (Baluch & Itti, 2011). Most studies of goal-directed attention have focused on attentional allocation to conscious cues (e.g. blue stimuli when looking for a blue flower). However, unconscious processing is omnipresent in human perception and action and we can hypothesize that attentional exploration could be influenced by unconscious sources of information too. Despite persistent skepticism regarding the existence of unconscious processes (Shanks & Stjohn, 1994; Vadillo, Konstantinidis, & Shanks, 2015), many studies have reported evidence of unconscious processing in perceptual decision making (Alamia et al., 2016; Greenwald, Klinger, & Schuh, 1995; Peremen & Lamy, 2014), motor learning (Axel Cleeremans, 1993; Clerget, Poncin, Fadiga, & Olivier, 2012; Destrebecqz, Peigneux, Laureys, Degueldre, & Fiore, 2005), and dynamic system control (Berry & Broadbent, 1988). A study from Zhao and

colleagues (Zhao et al., 2013) provided evidence in support of preferential attentional allocation to unconscious cues by investigating the relationship between statistical learning and visual attention (i.e. spatial and feature-selection). In a series of experiments, one of four locations displayed a statistically structured sequence of abstract shapes, whereas the order was random in the other locations. They showed that reaction times (RT) of an orthogonal task were decreased in the location exhibiting statistical regularities, thus suggesting that attention was affected by regularities even when these were not relevant for performing the task. In the study presented in the next chapter, we also showed that temporal statistical regularities affect overt attention (i.e. eye movements): participants were attending more frequently the predictable onset of a novel target. Interestingly, we found this effect exclusively when the predictability was task-relevant (Alamia & Zénon, 2016). In contextual cueing (M M Chun & Jiang, 1998; Marvin M. Chun, 2000), participants perform a visual search between two sets of stimuli having two different colors, only one of which includes the target. Unbeknown to participants, target location is paired with the spatial configurations of either one or the other set of stimuli. Even though the truly unconscious nature in this paradigm has been called into question (Smyth & Shanks, 2008), the results show that participants implicitly learn the rule, and bias their eye movements toward the portion of space where the target is more likely to appear (Jiang & Chun, 2001). Thus, in the visual domain, evidence seems to suggest that attention is affected by predictability, even though previous studies have not tested subjects' awareness thoroughly (Newell & Shanks, 2014; Shanks & Stjohn, 1994). Conversely, a recent work on predictability in the auditory domain (Southwell et al., 2017) found that auditory stimuli which were more predictable did not increase attentional capture neither when task-relevant nor when task-irrelevant, despite the reported evidence of an electrophysiological response reminiscent of

attentional mechanisms (i.e. increase in neural gain). A possible interpretation, as suggested by the authors, is that auditory and visual domains rely on different mechanisms.

In summary, the way unconscious sources of information drive attentional allocation remains poorly understood. No studies so far have investigated whether, and under which conditions, unconscious cues attract overt attention. In the present study, we aimed at exploring this question by exploiting an experimental design leading to robust unconscious learning, in which color information biases decisions in a motion discrimination task (Alamia et al., 2016).

Methods

Participants

22 healthy participants (15 females, mean age=23.17, std=1.68) took part in this experiment, receiving monetary compensation for their participation. All of them reported normal or corrected-to-normal vision. Two participants were discarded from the analyses because, during the debriefing at the end of the experiment, they could explicitly verbalize the association between color and motion. All the participants signed a participation consent before the experiment. The experiment was approved by the local Ethics committee of the Université Catholique de Louvain and was carried out in accordance with the Declaration of Helsinki.

Experimental design

Participants were comfortably sit, with the head placed on a chin rest, at a distance of 58 cm from the screen. Eye movements and blinks were recorded with an Eyelink© 1000+ eye tracker (SR Research Ltd., Kanata, Ontario, Canada; sample rate of 500Hz). The experiment was implemented in Matlab 7.5 (The MathWorks, Natick, MA, USA), using the version 3.0.9 of the Psychotoolbox (Brainard, 1997). The experiment lasted around 45 minutes, and it was composed of 14 blocks, each lasting 56 trials. Each trial consisted of three parts (fig. 1): at first a fixation cross was displayed until the participants maintained fixation continuously for 600ms; then two motion patches (see below) were presented until participants provided an answer, but no longer than 2000ms; finally, if the participant failed to provide an answer during the response period, an additional 1000ms blank display was presented to allow more time to respond.

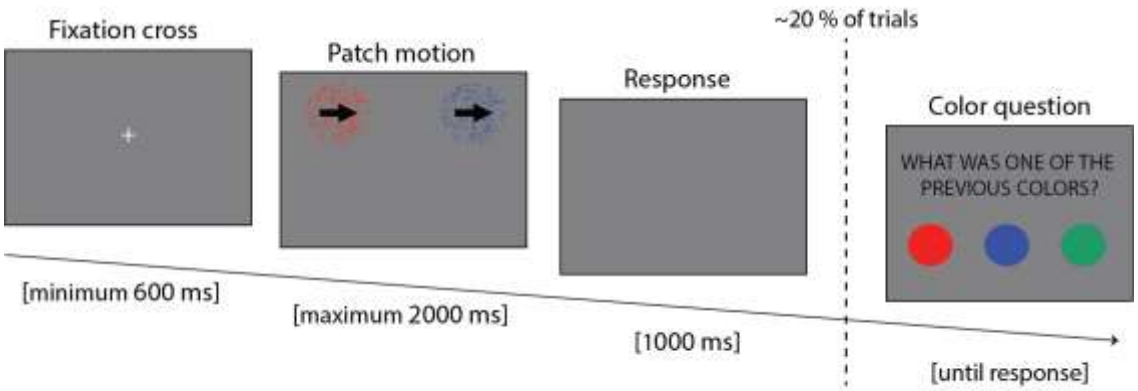


FIGURE 1

Figure 1 – Trial example showing the three parts: fixation cross, stimulus presentation and response time.

An auditory feedback was provided to inform the participants of the accuracy of their response. The stimuli consisted of two patches of dots, each 6° wide and located at 2 of the 4 pseudo-randomly selected corners of the screen, 20° from the center. The motion of the two patches of dots was either rightward or leftward, and both patches always had the same motion direction. Participants were instructed to fixate the cross at the beginning, and then visually explore the two patches to report their motion direction. The patches could have two different coherence levels: 25% and 60%. The coherence level of a patch reflects the percentage of dots moving towards the main direction of motion (i.e. left/rightward). The motion direction of the remaining dots was selected at random. The lower the coherence the more difficult it is to discriminate the motion direction (Gold, Law, Connolly, & Bennur, 2008). Thus, there were three types of trials: easy (both patches at 60% coherence), difficult (both patches having 25% coherence) and mixed (one patch 60% and the other 25%). The overall difficulty level of the task was tuned subject by subject in the first 6 blocks by changing the lifetime of the dots of both patches. In addition to coherence, we also manipulated dot lifetime, which is the number of frames each dot is displayed before disappearing: the longer the lifetime, the easier it is to perceive the motion. On the first training block, the lifetime of both patches was 15, and it was decreased in subsequent blocks provided the participants' performance was above 70%. The levels of lifetime were 15, 6, 4, 3 and 2. All except one participant reached the lowest level before the 7th block. Starting from the 7th block, the lifetime remained unchanged throughout the whole experiment. All the dots of each patch were of the same color, with three possible colors (i.e. red, blue and green). Unbeknownst to the participants, starting from the 7th block, one color was always associated to the rightward direction, another color to the leftward direction and the last color was uninformative of the motion direction of the dots. This association was pseudorandomized

between participants. Briefly, two types of trials were possible: both patches shared the same color, or they had different colors. The color associated with leftward motion and the one associated with rightward motion were never presented together (the color-motion association was 100% valid). The frequency of occurrence of the colors was balanced during the whole experiment. Moreover, in 20% of the trials (11 out of 56 in a block) participants were asked to report one of the patch colors, forcing them to pay attention to the colors and providing us with an additional measure of the attended color. All the possible types of trials are summarized in table 1 (see in Data Analysis). At the end of the experiment, participants responded to a de-briefing questionnaire composed of 4 questions: first, whether one motion direction was easier to discriminate than the other; second, whether one of the four positions was attended more than the others; third, whether the motion was easier to perceive with one of the three colors; fourth, whether they had remarked an association between colors and motion. In case of positive answer to this last question, they were asked to report the nature of the association.

Data Analysis

During the whole experiment, we recorded eye position, participants' responses and reaction times (RT). Two participants out of 22 reported the correct color-motion association at the final questionnaire, and thus were excluded from further analyses.

Statistical analyses were performed in JASP (Love et al., 2015), applying Bayesian ANOVA: all the Bayes Factors refer to the alternative hypothesis, and are reported as BF_{10}. All other analyses, including eye movement preprocessing and feature extraction were performed in Matlab 7.5 (The MathWorks, Natick, MA, USA).

Accuracy and RT analysis

We performed behavioral and eye-movements analyses starting from the 7th block, i.e. the block in which the color-motion association was introduced in the experiment. Regarding the behavioral part, we tested two Bayesian ANOVAs considering either the accuracy (model I) or RT (model II) as dependent variables: the accuracy was modeled as a binary variable, whereas the RT was modeled as Gaussian. The independent variables considered were: PREDICTABILITY (a categorical variable modeling whether the color was informative or not), BLOCKS (categorical variable from 7 to 14), DIFFICULTY (categorical variable modeling whether the trial was difficult or easy) and all their interactions. Initially, we analyzed only trials in which the patches had the same color and the same coherence level, in order to remove all confounds induced by eye movements during patch selection (condition 7 in table 1). In a second analysis, to study how selective attention affected accuracy and RT, we performed 2 additional Bayesian ANOVAs, one on accuracy (model III) and one on RT (model IV), considering the patch on which attention was allocated. We determined attention allocation as the distance between the eye position and the patches at the time of the participant's response (see eye movement analysis below). Trials in which participants moved back and forth between the patches were excluded from this analysis (~15% of the trials). As in the previous analysis, we considered PREDICTABILITY and DIFFICULTY of the attended patch as factors. In this analysis and in the subsequent analyses regarding eye movements, we focused on 2 sub-categories of trials in order to simplify the interpretation of the findings (see table 1): same coherence level in both patches (i.e. 60/25% coherence) but one patch having a predictive color whereas the other having a non-predictive one. This approach allowed us to investigate specifically the effects of the color-motion association, independently from the effect of

difficulty (i.e. coherence levels). We did not investigate further the remaining 4 conditions: conditions 3 and 4 lead to trivial results (participants pay more attention to easier than more difficult patches), whereas in conditions 5 and 6 it would be challenging to properly disentangle the effects of predictability from the effect of patch coherence.

	PATCH 1		PATCH 2	
	COHERENCE	COLOR	COHERENCE	COLOR
Condition 1	60 % (easy)	predictive	60 % (easy)	non-predictive
Condition 2	25 % (difficult)	predictive	25 % (difficult)	non-predictive
Condition 3	60 % (easy)	predictive	25 % (difficult)	predictive
Condition 4	60 % (easy)	non-predictive	25 % (difficult)	non-predictive
Condition 5	60 % (easy)	predictive	25 % (difficult)	non-predictive
Condition 6	25 % (difficult)	predictive	60 % (easy)	non-predictive
Condition 7	As coherence 2	As color 2	As coherence 1	As color 1

Table 1: all possible conditions. Only 4 cases have been considered for the analyses.

Eye movement analysis

Concerning eye movements, we removed the blinks (automatically detected by the Eyelink© algorithm) by means of linear interpolation.

Afterward, we determined attentional allocation trial by trial, on the basis of which patch was attended by the participants when they provided the answer: first we computed the distance between the eye position and each patch, then we normalized these distances by their sum, and finally we attributed a positive value in the attentional allocation binary variable to the patch with a normalized distance lower than 0.4. We then compared the percentage of trials in which the predictive or non-predictive patches were attended (model V). Similarly to the behavioral analysis, we included only the conditions 1 and 2 in table 1. In addition, we included only trials in which participants attended a single patch (around 85% of all the trials). In a second analysis, we focused on the remaining portion of trials in which participants switched their attention from one patch to the other (around 15% of the trials): we compared the percentage of time in which they switched from predictive to non-predictive color and vice versa (model VI). Both models V and VI were implemented by means of Bayesian ANOVA. Finally, as an indirect measure of attention, we tested whether participants reported more often the predictive or the non-predictive color, when asked, at the end of each trial, which colors had been presented (Bayesian paired t-test).

Results

Behavioral analysis

In the first model (model I), we tested whether accuracy was affected by the PREDICTABILITY of the color, the BLOCK and the DIFFICULTY factors. This analysis was restricted to trials in which both patches had the same color and coherence level. As expected, we found a significant effect of the factor DIFFICULTY (BF_{10}>>100 very strong evidence), indicating that

participants were better at discriminating the motion direction of the patches when both had a coherence of 60%, than when both had 25% coherence (figure 2A). Interestingly, we found also an effect of the factor PREDICTABILITY (BF_{10}=17.86, strong evidence), with better performance for predictive than non-predictive colors, confirming that participants learned the implicit association between color and motion. No interaction between PREDICTABILITY and DIFFICULTY was found (BF_{10}<1). We found a significant negative results of the factor BLOCK (BF_{01}=42.08) but all related interactions were not significant (all BF_{10}<1).

Regarding the RT (model II) we found a similar significant effect of DIFFICULTY (BF10>>100 very strong evidence), indicating faster responses for easier patches (i.e. patches with 60% coherence), and an effect of PREDICTABILITY (BF10>>10 strong evidence). All the other factors or interactions, were far from reaching significance (all BF10<1).

The second behavioral analysis aimed at investigating how accuracy and RT changed as a function of attentional allocation (figure 2B). Here we included only trials in which coherence was identical in both stimuli, such that the patches differed only along one dimension. We found that the difficulty level of the attended patch affected performance (Model III; BF10>>100, very strong evidence), but we found no main effect of the PREDICTABILITY factor (BF10=1.2). Nevertheless, we found positive evidence of an interaction between DIFFICULTY and PREDICTABILITY (BF10=3.425), suggesting that participants were exploiting the predictability specifically when both patches were difficult (post hoc comparison between predictive and non-predictive patches in the difficult condition: BF10=3.824, in the easy condition BF10=0.410). Regarding the RT (model IV), we also found a significant impact of the factor DIFFICULTY (BF10>>100, very strong evidence) but no other effects (all BF10<1).

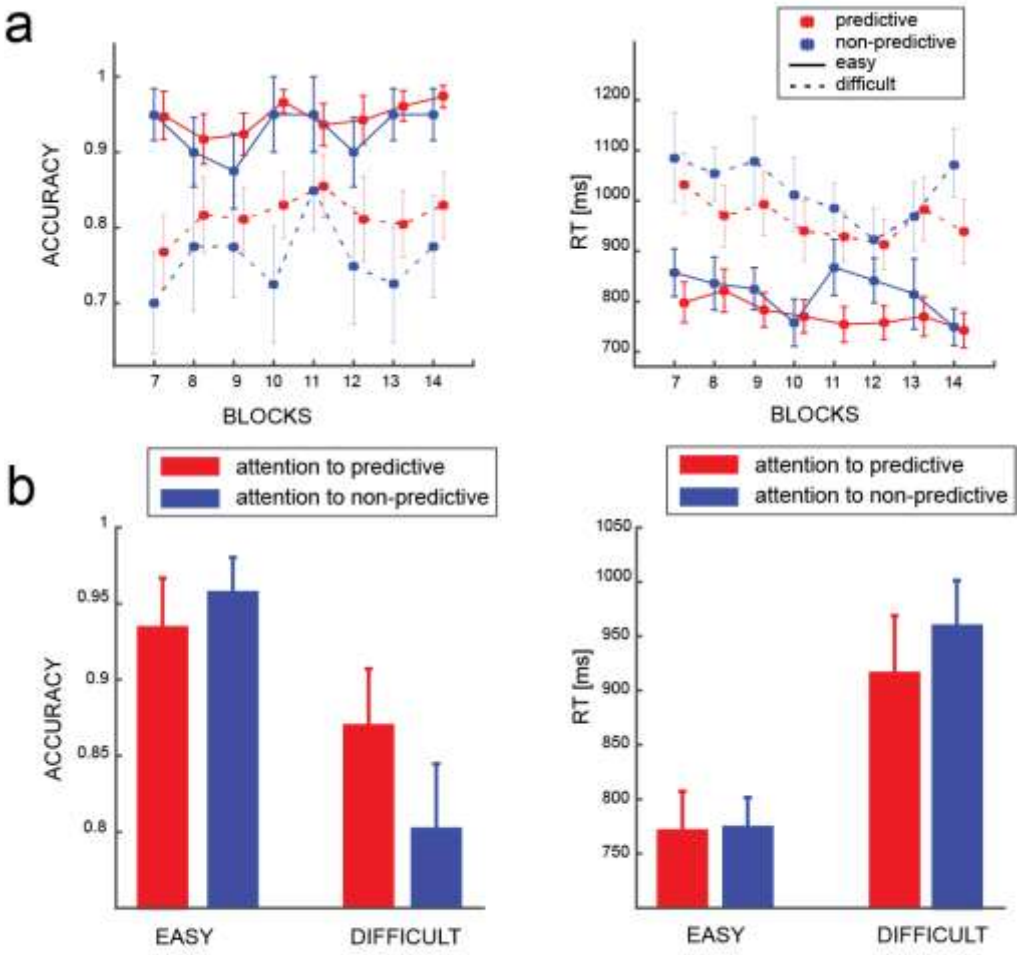


FIGURE 2

Figure 2: a) Averaged accuracy and RT results for the easy (solid lines) and difficult (dashed lines) trials, and for predictive (red) and non-predictive (blue) trials. b) Averaged accuracy and RT results according to which patch was attended by participants (same color code as in 'a'). In all the panels, error bars are standard errors.

All in all, these analyses show that participants' accuracy was higher when they attended the patch whose color conveyed information about motion direction, specifically when the motion of both patches was hard to perceive.

Eye movement analysis

For the eye movement analysis, we considered as binary dependent variable (DV) the percentage of trials in which participants attended the predictive or the non-predictive patch (model IV; factor PREDICTABILITY), and whether both patches were easy or difficult (factor DIFFICULTY). As shown in figure 3A, we found a significant result for both factors (DIFFICULTY BF_{10}= 11.48, PREDICTABILITY BF_{10}= 11.48, positive evidence), and a strong interaction between the two factors (BF_{10}= 53.61, strong evidence). A post-hoc analysis revealed a significant difference between predictable and non-predictable colors in the difficult (BF>1000) but not in the easy condition (BF=1.433), suggesting that participants were looking more at the informative patch when the coherence of both patches was lower (i.e. 25%), in agreement with the results of the previous analysis (model III).

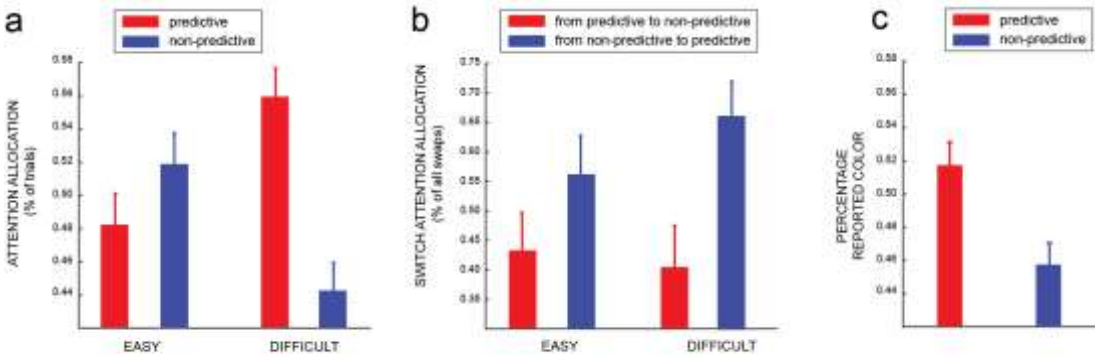


FIGURE 3

Figure 3: a) Average percentage of trials in which participants looked at the predictive (in red) or non-predictive patches (in blue) when both are easy (solid lines) or difficult patches (dashed lines). b) Average percentage of trials in which participants switched attention from one patch to the other (same color code as 'a'). c) Average percentage of times participants reported the predictive (red) or non-predictive (blue) color. In all the panels, error bars are standard errors

Figure 3B shows the percentage of trials in which participants made a saccade from one patch to the other (around 15% of trials, with 3 participants having a percentage closer to 40% and all the other participants < 10%). We found a positive effect of PREDICTABILITY (BF_{10}= 5.11, positive evidence), irrespective of the difficulty level of both patches. It is important to note that the number of trials included in the analysis varied between participants and was sometimes as low as 5 trials per condition per subject.

Finally, we investigated which patch color was reported more frequently when one patch had a predictive color and the other patch a non-predictive one. We found that participants reported more frequently the predictive than the non-predictive patch (bayesian paired t-test: BF>100,

very strong evidence), as shown in figure 3C, in agreement with the previous analyses that suggested a bias of attention toward the informative colors.

Discussion

In this study, we investigated the effect of unconscious learning on visual attention by means of eye movements. Participants were instructed to report the motion direction of two patches which could have different difficulty levels, and different colors: unbeknownst to them, two out of three colors were associated with the correct response (i.e. one motion direction). Participants failed to notice this association consciously, but nevertheless exploited the color information to perform the task, replicating our previous findings (Alamia et al., 2016). The question we addressed in the current study was whether unconscious knowledge of the color-motion association affects eye movements. As expected, we found that learning affected behavioral measures (RT and accuracy), and importantly, that participants were attending more frequently the patches exhibiting the predictive colors, despite not being consciously aware of the associations. Also, the indirect measure of attention, in which subjects had to report the color shown in the previous trial, was affected by the implicitly learnt association.

Previous studies have investigated the relationship between visual attention and implicit learning (Alamia & Zénon, 2016; Chun & Jiang, 1998; Zhao, Al-Aidroos, & Turk-Browne, 2013). Beside visual attention, the allocation of attentional resources has also been studied in the context of implicit sequence learning,that is in serial reaction time task (SRTT) paradigms. In these tasks, participants learn a sequence of responses implicitly. These sequences can be either deterministic, probabilistic or random. In the first two cases (i.e. when the sequence is predictable) participants perform better than in the random case, even when they are not

aware of the predictability (Destrebecqz & Cleeremans, 2001; Destrebecqz, Peigneux, Laureys, Degueldre, & Fiore, 2005). In a seminal study, Nissen and Bullemer (Nissen & Bullemer, 1987) showed that the addition of a secondary orthogonal task, which pulls participant's attentional resources, impaired learning, thus advocating the crucial role of attention in implicit learning. Some authors suggested a rather different interpretation, framing the results in terms of interference between the first and the secondary task (Stadler, 1995), and not in terms of attentional resources. Moreover, contrasting with the Nissen and Bullemer study, Cohen provided results which showed no effect of attention on participants' performance (Cohen, Ivry, & Keele, 1990), even though further studies failed to replicate Cohen results: (Frensch, Lin, & Buchner, 1998), and Cleeremans and colleagues found intermediate results, with lesser, but significant implicit learning in the presence of an orthogonal task (A Cleeremans, Destrebecqz, & Boyer, 1998). In conclusion, the role of attention in implicit learning in the context of SRTT remains disputed.

Beyond implicit supraliminal learning, Kanai and colleagues have investigated the role of attention in subliminal visual stimulation (Kanai, Tsuchiya, & Verstraten, 2006). In their task, a tilt after effect (TAE, i.e. the perceived orientation of a grating pattern can appear altered due to the spatial surrounding and the context) was induced by a subliminal stimulus (due to continuous flash suppression) at two different locations, one of which was attended by participants. Their results showed a TAE with subliminal adaptors both when participants were attending the location and when they were not, suggesting that spatial attention does not influence the perception of subliminal stimuli. However, in a subsequent experiment, they showed that feature-based attention, rather than spatial attention, can modulate visual stimuli unconsciously (Kanai et al., 2006) and other studies have showed that

spatial attention can be also affected by subliminal stimuli (Chou & Yeh, 2011; Mastropasqua & Turatto, 2015; Mulckhuyse, Talsma, & Theeuwes, 2010). Overall, the majority of studies seems to suggest that visual attention can be affected by subliminal information and the other way around, despite further study are required to explain the few negative results.

With regards to the current literature, the element of novelty introduced in our study lays in the nature of the unconscious learning involved in the paradigm: the learning of simple, supraliminal associations makes it straightforward to test the participants' explicit knowledge in performing the task, thus addressing the main criticisms about the truly unconscious nature of learning (see (Alamia et al., 2016) for an extensive investigation of awareness in this task). Our results show that the task-relevant information in the predictive patches biases the decision process about which patch to attend, even when unconscious, in agreement with the theoretical framework presented by Gottlieb and colleagues (Gottlieb, 2012; Gottlieb & Balan, 2010)

Target selection for eye movements and covert visual attention can be seen as a decision making process (Gottlieb, 2012; Gottlieb & Balan, 2010), in which participant decide where to sample the visual space in order to reduce their internal model's uncertainty (Feldman & Friston, 2010). Framed along the exploration-exploitation conundrum (March, 1991; Payzan-LeNestour & Bossaerts, 2012), participants could decide to reduce uncertainty about the correct behavior (i.e. response) by sampling preferentially the more reliable predictors (i.e. exploitation), or to reduce uncertainty about the environment by sampling items that are more unpredictable or surprising (i.e. exploration). The first approach is in line with the Mackintosh model (Mackintosh, 1975), in which informative stimuli are more salient, whereas the second approach is the one advocated by the

Pearce and Hall model (Pearce & Hall, 1980), in which more uncertain stimuli capture attention. Our results are in line with the first interpretation, in which attention is biased toward more predictable stimuli, even when the added information is non-conscious. Nevertheless, it would be possible to test the prediction regarding the second model by changing the color-motion associations at each block, such that exploration of a potentially novel association would be encouraged as much as its exploitation. All in all, attending regularities (i.e. Mackintosh model) could be the best strategy in a relatively stable environment, whereas exploring more uncertain predictors, in accordance with the Pearce Hall model, may be more efficient in a volatile environment. According to physiological studies (G. Aston-Jones, Rajkowski, & Kubiak, 1997; Gary Aston-Jones & Cohen, 2005; Clayton, Rajkowski, Cohen, & Aston-Jones, 2004), which strategy is put in place may depend on the interplay between phasic and tonic activation in the neurons of the norepinephrine system. Further studies on this topic will shed light on the exploration-exploitation tradeoff in unconscious processing.

One limitation of our task resides in the inability to disentangle feature-based attention (Hayden & Gallant, 2005; Kanai et al., 2006) and spatial-based attention on the one hand, and overt and covert attention on the other hand (Filali-Sadouk, Castet, Olivier, & Zenon, 2010; Hunt & Kingstone, 2003; Zénon, Corneil, Alamia, Filali-Sadouk, & Olivier, 2014). However, regarding the dissociation between covert and overt attention, the agreement between the results of the direct (eye movement) and indirect (reporting one of the two colors) measure of attention suggests that covert attention allocation followed the same guiding principles as overt attention. A second limitation regards the early occurrence of implicit learning since the first block (Alamia et al., 2016; Turk-browne, Junge, & Scholl, 2005), which prevented us to explore the interaction between the different time courses of

learning and attention: we can only speculate that, similarly to other tasks (Marvin M. Chun, 2000), unconscious learning precedes attention and drives its effect on eye movements.

Moreover, our results confirmed the dissociations between attention and consciousness (Koch & Tsuchiya, 2007, 2012; Tsuchiya & Koch, 2009). As suggested in other studies (Beilock, Carr, MacMahon, & Starkes, 2002; Olivers & Nieuwenhuis, 2005), attention and consciousness can have different and dissociable effects on behavior, thus hinting that these two processes potentially rely on different neuronal mechanisms. Further studies are required to investigate the relationship between visual attention and consciousness.

References

Alamia, A., Orban de Xivry, J.-J., San Anton, E., Olivier, E., Cleeremans, A., & Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of Consciousness*, *2016*(1), 1–10. http://doi.org/10.1093/nc/niw016

Alamia, A., & Zénon, A. (2016). Statistical Regularities Attract Attention when Task-Relevant. *Frontiers in Human Neuroscience*, *10*(February), 1–10. http://doi.org/10.3389/fnhum.2016.00042

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–50. http://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience*, *80*(3), 697–715. http://doi.org/10.1016/S0306-4522(97)00060-2

Attneave, F. (1954). Some informational aspects of visual perception.

Psychological Review, *61*(3), 183–193. http://doi.org/10.1037/h0054663
Berry, D. C., & Broadbent, D. E. (1988). On the relationship between task performance and verbal knowledge. *Quarterly Journal of Experimental Psychology, 36A, 209-231 .*
Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952
Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, *3*(9), 1–24. http://doi.org/10.1167/9.3.5
Chou, W. L., & Yeh, S. L. (2011). Subliminal spatial cues capture attention and strengthen between-object link. *Consciousness and Cognition*, *20*(4), 1265–1271. http://doi.org/10.1016/j.concog.2011.03.007
Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*. http://doi.org/10.1016/S1364-6613(00)01476-5
Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. http://doi.org/10.1006/cogp.1998.0681
Clayton, E. C., Rajkowski, J., Cohen, J. D., & Aston-Jones, G. (2004). Phasic activation of monkey locus ceruleus neurons by simple decisions in a forced-choice task. *J Neurosci*, *24*(44), 9914–9920. http://doi.org/24/44/9914 [pii] 10.1523/JNEUROSCI.2446-04.2004
Cleeremans, A. (1993). Attention and awareness in sequence learning. *Proceedings of the Fifteenth Annual Conference*. Retrieved from http://books.google.com/books?hl=en&lr=&id=sjVemQTs60IC&oi=fnd&pg=PA330&dq=Attention+and+Awareness+in+Sequence+Learning&ots=20zfwHGciq&sig=gdsOTBEeNJgipzyjqUL6W8H-Y2Q
Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–16.
Clerget, E., Poncin, W., Fadiga, L., & Olivier, E. (2012). Role of Broca's Area in Implicit Motor Skill Learning: Evidence from

Continuous Theta-burst Magnetic Stimulation. *Journal of Cognitive Neuroscience*. http://doi.org/10.1162/jocn_a_00108
Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17–30. http://doi.org/10.1037/0278-7393.16.1.17
Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3 Suppl*(november), 1218–23. http://doi.org/10.1038/81504
Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. http://doi.org/10.1146/annurev.ne.18.030195.001205
Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, *178*(2–3), 363–380. http://doi.org/10.1016/0006-8993(79)90699-1
Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., & Fiore, G. Del. (2005). The neural correlates of implicit and explicit sequence learning : Interacting networks revealed by the process dissociation procedure The neural correlates of implicit and explicit sequence learning : Interacting networks revealed by the process dissocia. http://doi.org/10.1101/lm.95605
Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*(December), 215. http://doi.org/10.3389/fnhum.2010.00215
Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research*, *61*(2), 83–98. http://doi.org/10.1007/s004260050015
Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11). http://doi.org/10.1371/journal.pcbi.1000211
Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. http://doi.org/10.1038/nrn2787
Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*(MAY). http://doi.org/10.3389/fpsyg.2012.00151
Gottlieb, J. (2012). Attention, learning, and the value of information.

Neuron, *76*(2), 281–95. http://doi.org/10.1016/j.neuron.2012.09.034
Gottlieb, J., & Balan, P. (2010). Attention as a decision in information space. *Trends in Cognitive Sciences*, *14*(6), 240–248. http://doi.org/10.1016/j.tics.2010.03.001
Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by Marginally Perceptible ("Subliminal") Stimuli: Dissociation of Unconscious From Conscious Cognition. *Journal of Experimental Psychology: General*, *124*(I), 22–42. http://doi.org/10.1037/0096-3445.124.1.22
Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.
Hayden, B. Y., & Gallant, J. L. (2005). Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron*, *47*(5), 637–643. http://doi.org/10.1016/j.neuron.2005.07.020
Hogarth, L., Dickinson, A., Austin, A., Brown, C., & Duka, T. (2008). Attention and expectation in human predictive learning: the role of uncertainty. *Quarterly Journal of Experimental Psychology (2006)*, *61*(11), 1658–68. http://doi.org/10.1080/17470210701643439
Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. http://doi.org/10.1016/S1053-5357(00)00088-3
Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning, (4), 1105–1124. http://doi.org/10.1080/0272498004200051
Kanai, R., Tsuchiya, N., & Verstraten, F. A. J. (2006). The Scope and Limits of Top-Down Attention in Unconscious Visual Processing. *Current Biology*, *16*(23), 2332–2336. http://doi.org/10.1016/j.cub.2006.10.001
Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. http://doi.org/10.1016/j.tics.2006.10.012
Koch, C., & Tsuchiya, N. (2012). Attention and consciousness: Related yet different. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2011.11.012

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*, 193–243. http://doi.org/10.1080/02724990344000141

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology. Animal Behavior Processes*, *37*, 220–229. http://doi.org/10.1037/a0021384

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*. http://doi.org/10.1037/h0076778

Mastropasqua, T., & Turatto, M. (2015). Attention is necessary for subliminal instrumental conditioning. *Nature Publishing Group*, 1–7. http://doi.org/10.1038/srep12920

Mulckhuyse, M., Talsma, D., & Theeuwes, J. (2010). Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Journal of Vision*, *7*(9), 1081–1081. http://doi.org/10.1167/7.9.1081

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. http://doi.org/10.1016/0010-0285(87)90002-8

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

Pearce, J. M., & Mackintosh, N. J. (2010). Attention and Associative Learning: From Brain to Behaviour - Chris J. Mitchell, Mike E. Le Pelley - Google Books. *Attention and Associative Learning: …*. Retrieved from http://books.google.com/books?hl=en&lr=&id=wn-hHPt_ftgC&oi=fnd&pg=PA11&dq=two+theories+of+attention+review+possible+integration&ots=jwGwLbBIE1&sig=I3XV-yJDFUqafjUpyW9nEdjpn2g%5Cnpapers2://publication/uuid/4C7F3C4A-04D5-4F5B-A0CD-A64947FC3872

Peremen, Z., & Lamy, D. (2014). Do conscious perception and unconscious processing rely on independent mechanisms? A meta-contrast study. *Consciousness and Cognition*, *24*(1), 22–

32. http://doi.org/10.1016/j.concog.2013.12.006
Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, *16*(16), 1843–8. http://doi.org/10.1097/01.wnr.0000183900.92901.fc
Shanks, D., & Stjohn, M. (1994). Characteristics of Dissociable Human Learning-Systems, (February 2010). http://doi.org/10.1017/S0140525X00035032
Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403–415. http://doi.org/10.3758/MC.36.2.403
Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 674–685. http://doi.org/10.1037/0278-7393.21.3.674
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. Retrieved from http://www.erc.caltech.edu/Industry/Conferences/2004/AIC/pdf/Laurent-Itti.pdf
Tsuchiya, N., & Koch, C. (2009). The relationship between consciousness and attention. In *The Neurology of Consciousness* (pp. 63–77). http://doi.org/10.1016/B978-0-12-374168-4.00006-X
Vadillo, M. a., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-015-0892-6
Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–77. http://doi.org/10.1177/0956797612460407

Chapter IV: Statistical regularities attract attention when task-relevant

This chapter is a modified version of the following article:

Alamia, A., & Zénon, A. (2016). Statistical Regularities Attract Attention when Task-Relevant. *Frontiers in Human Neuroscience*, *10*(February), 1–10. http://doi.org/10.3389/fnhum.2016.00042

Abstract

Visual attention seems essential for learning the statistical regularities in our environment, a process known as statistical learning. However, how attention is allocated when exploring a novel visual scene whose statistical structure is unknown remains unclear. In order to address this question, we investigated visual attention allocation during a task in which we manipulated the conditional probability of occurrence of colored stimuli, unbeknown to the subjects. Participants were instructed to detect a target colored dot among two dots moving along separate circular paths. We evaluated implicit statistical learning, i.e. the effect of color predictability on reaction times (RT), and recorded eye position concurrently. Attention allocation was indexed by comparing the Mahalanobis distance between the position, velocity and acceleration of the eyes and the 2 colored dots.

We found that learning the conditional probabilities occurred very early during the course of the experiment as shown by the fact that, starting already from the first block, predictable stimuli were detected with shorter RT than unpredictable ones. In terms of attentional allocation, we found that the predictive stimulus attracted gaze only when it was informative about the occurrence of the target but not when it predicted the occurrence of a task-irrelevant stimulus. This suggests that attention allocation was influenced by regularities only when they were instrumental in performing the task. Moreover, we found that the attentional bias towards task-relevant predictive stimuli occurred at a very early stage of learning, concomitantly with the first effects of learning on RT.

In conclusion, these results show that statistical regularities capture visual attention only after a few occurrences, provided these regularities are instrumental to perform the task.

Introduction

One of the central functions of the human brain is the ability to predict the surrounding dynamics and to optimize interactions with the environment (Clark, 2013; Little & Sommer, 2013). Learning contingencies and regularities is a multi-faceted and elaborated mechanism that allows the brain to perform predictions and optimization (Dayan, Kakade, & Montague, 2000; Kruschke, 2003; O'Brien & Raymond, 2012). Attention is regarded as an important mechanism involved in reducing perceptual uncertainty but its role in learning remains controversial (Gottlieb, 2012). On the one hand, a model proposed by Pearce and Hall (Pearce & Hall, 1980) suggests that unpredictable and surprising cues capture attention more than predictable ones, supposedly because they provide new information about the environmental contingencies. This view is supported by experimental studies that have showed that attention gets preferentially allocated to conditioned stimuli with uncertain outcomes (Hogarth, Dickinson, Austin, Brown, & Duka, 2008). On the other hand, an alternative model (Mackintosh, 1975), which has also received recent experimental support (Kruschke, 2001; Le Pelley, Beesley, & Griffiths, 2011), suggests the opposite view, arguing that predictability attracts attention and that this early attentional capture would be instrumental in learning. Lately, an hybrid model integrating the two theories has been proposed in order to conciliate these controversial experimental findings, postulating the co-existence of two distinct attentional systems, namely a controlled system processing the most unpredictable stimuli in order to learn the dynamics of the environment and an automatic one, exploiting information already acquired and focusing on the stimuli essential to perform the task at hand (Le Pelley, 2004; Pearce & Mackintosh, 2010). A similar reading of the role of attention in learning has been provided by Dayan and colleagues (Dayan et al., 2000), who considered the Pearce–

Hall model (Pearce & Hall, 1980) as appropriate to explore and learn regularities in a new environment whereas the Mackintosh model (Mackintosh, 1975) would drive behavior during the routine execution of a task.

It is noteworthy that the above models were framed in the context of associative learning, in which the predictive structure of the stimuli is directly relevant to the task, being typically associated to rewards or punishments. However, learning can also occur in situations in which it is not useful to the task. In the context of perceptual learning (i.e. the enhancement of perceptual performance consecutive to repeated stimulation; Lu, Hua, Huang, Zhou, & Dosher, 2011), the importance of attention remains controversial. On the one hand, Watanabe and colleagues found that subthreshold task-irrelevant stimuli affected the discrimination of supra-threshold stimuli, thus suggesting that attention is not needed for perceptual learning to occur (Watanabe, Náñez, & Sasaki, 2001). On the other hand, other studies showed opposite results: only task-relevant and actively attended information was learnt (Ahissar & Hochstein, 1993; Shiu & Pashler, 1992). In the context of "statistical learning", a type of learning first coined by Saffran in the framework of language acquisition in infants (Saffran, Aslin, & Newport, 1996), statistical environmental contingencies can also be learned in situations in which they are not useful to the task (Zhao, Al-Aidroos, & Turk-Browne, 2013), and may sometimes remain entirely implicit (Fiser & Aslin, 2001; Perruchet & Pacton, 2006; Saffran, Johnson, Aslin, & Newport, 1999). Statistical learning occurs also irrespective of the perceptual modality; indeed, it has been reported in the visual (Turk-Browne, Isola, Scholl, & Treat, 2008; Turk-Browne, Junge, & Scholl, 2005), auditory (Saffran et al., 1999) and tactile domains (Conway & Christiansen, 2005). Despite a wealth of studies investigating the general mechanisms of statistical learning, its relation to attention remains poorly understood. Previous studies have

suggested that attentional allocation is necessary for the learning to occur (Toro, Sinnett, & Soto-Faraco, 2005; Turk-browne et al., 2005). Conversely, only one study has addressed the effect of statistical learning on attentional allocation (Zhao et al., 2013), arguing in favor of the hypothesis that statistical regularities attract attention, even when not task-relevant. However, in that study, only covert attention allocation was assessed, during a task in which statistical regularities were always irrelevant to the task being performed.

In the current study, we addressed the issue of the relationship between visual attention and statistical learning when the statistical structure of the stimulus sequence is either relevant to the task or not. We performed an experiment in which statistical regularities were manipulated during a simple color detection task, while recording eye movements. Specifically, we controlled the conditional probability of occurrence of the different colors, such that some colors allowed predicting the target occurrence whereas the others did not. We evaluated statistical learning by measuring reaction times (RT) as a function of color predictability (Abla & Okanoya, 2009; Barakat, Seitz, & Shams, 2013), while visual attention allocation was estimated by comparing the position, velocity and acceleration of the eyes with respect to those of the stimuli. The aim of the current study was to provide evidence in favor of one or the other aforementioned model, by comparing the attentional allocation to both the predictive and predicted stimuli in trials where a target was present or not.

Methods

Participants

Nineteen healthy participants (mean age=24.4, SD=2.98, 12 females) took part in the experiment. All of them reported normal or corrected-to-

normal vision. The experiments were carried out according to the Declaration of Helsinki and were approved by the Ethics Committee of the Université Catholique de Louvain. Written informed consents were obtained from all the participants.

Experimental design and equipment

The experiment took place in a dim and quiet room, and lasted for around 40 minutes. The participants were seated comfortably on a chair in front of a 19" CRT screen, with a 75Hz refresh rate, with their head resting on a chinrest 58 cm from the screen to ensure stability during the eye-tracking recordings. An Eyelink© 1000+ eye tracker (SR Research Ltd., Kanata, Ontario, Canada) monitored eye movements and blinks at a sampling frequency of 500 Hz. The task was implemented using the version 3.0.9 of the Psychotoolbox (Brainard, 1997) with Matlab 7.5 (The MathWorks, Natick, Massachusetts, USA).

The experiment consisted of a color detection task: the participant had to click on a computer mouse with the right index finger whenever one of the two dots (1° wide) moving on the screen featured the target color. The experiment was composed of 8 blocks lasting 4 minutes each. Between the 4^{th} and 5^{th} block, the participants were allowed to have a few minutes' break, during which the light was turned on. Blocks were composed of 108 trials, and a different target color (n=6) was designated every 18 trials, by displaying a large dot (4° wide) of that color at the center of the screen for 1500 ms. During each block, each of the six possible dot colors was used as target color, in randomized order. Each correct detection was signaled by a positive auditory feedback and associated to a monetary reward of 2 cents, while each wrong answer was associated with a negative auditory feedback and a negative reward of -2 cents.

In each trial, the two colored moving dots were displayed over a gray background (70% of the maximum luminance of the screen) and moved along two circular paths (16° wide), starting from the center of the screen and heading upwards, such that the right dot moved clockwise toward the right part of the screen, and the left one anticlockwise to the opposite side (fig.1A). The two dots always had the same velocity (180°/s), which was kept constant across trials; the total duration of the circular displacement of the colored dots was then 2000 ms. When the two moving dots were about midway, between 700 to 1500 ms after trial onset (randomized across trials), they both changed color simultaneously. Trials were separated by a 300 ms interval during which the screen remained gray. This particular design, i.e. the circular motion of the stimuli, was chosen to elicit spontaneous eye movements while driving the subject to fixate the center of the screen between each trial, even though no fixation cross was displayed. The 6 possible colors of the dots were red, blue, orange, brown, green and purple, and the two dots never had the same color. Importantly, the probability distribution of the colors appearance on the screen was not random. In the first part of the trial, i.e. before the dot changed color, all the colors (including the targets) appeared with the same probability, but in the second part of the trial, the colors were conditional on the colors displayed in the first part. In particular, the colors were randomly split into two groups: two colors, selected pseudo-randomly between participants, predicted each other with a probability of 100%, while the remaining four colors predicted each other with a probability of 33% (fig.1B). Moreover, the colors had the same overall frequency of occurrence, both in the first and second part of each trial. At the end of the experiment, participants were verbally asked if they had noticed any difference between the first and second parts of the trial, the left and right spots, and between colors. None of them reported any explicit bias in the experiment.

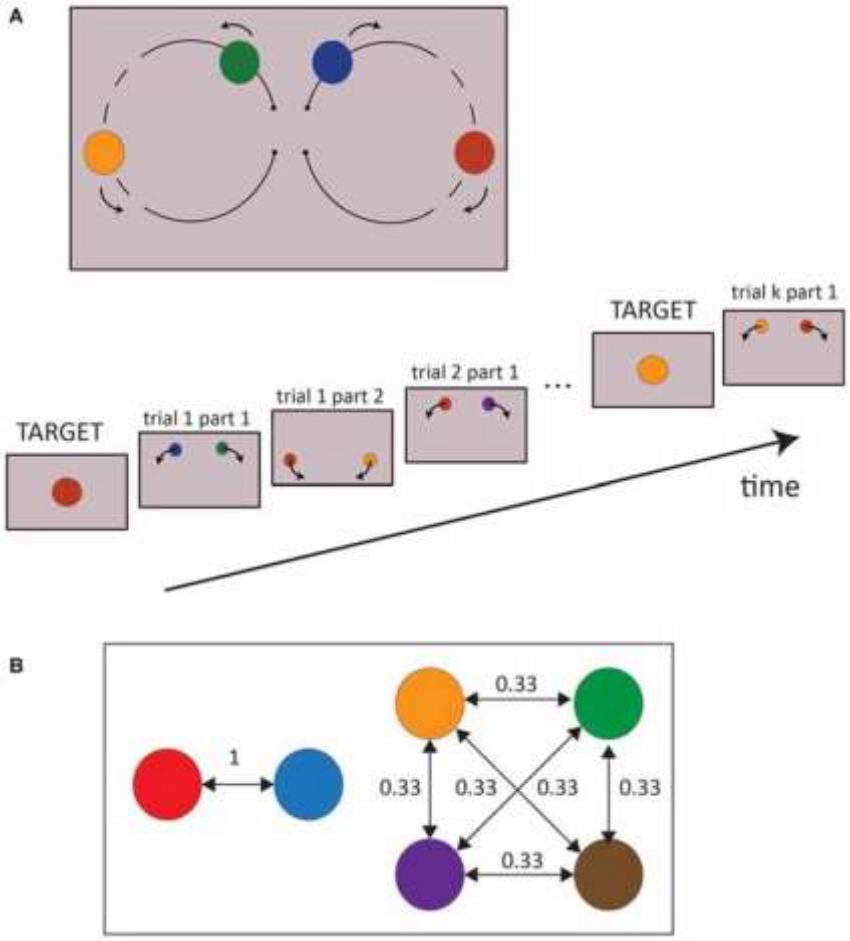


Figure 1 A – Experimental Design. The upper part is a schematic representation of a whole trial, while the lower part of the picture represents the successive stages of a block. The dashed line represents the range of dot positions in which a change of color may occur. *B* – An example of transition probabilities between colors is represented. Two colors were always associated to each other (conditional probability=1, predictable colors) while the remaining colors all shared a conditional probability equal to 0.33 (unpredictable colors). The transition probabilities of the colors were pseudo randomized between subjects.

Data analysis

Reaction Time Analysis

First, we log-transformed the RTs in order to make their distribution closer to normal and removed the outliers (over ±3 SD, around 1% of the data). Second, since we found that the timing (700-1500 ms) of the color change affected the RTs for color detection in the second part of the trial (negative correlation between RTs and color change delay: $R = -0.35$ $p < 0.0001$), we removed this effect by considering the residuals of this regression. So the GLMM analyses described below included as dependent variable the difference between the RT data and the predicted values obtained from the regression, effectively removing this influence from the data. There was no such effect in the first half of the trials.

Eye movement analysis

By inspecting the data we determined that taking into account only the distance between the eye and stimulus positions was not an effective way to determine attention allocation. As shown in figure 2A, in trials in which the participants did not make a saccade toward one of the stimuli, the mere distance between the eyes and the two dots was poorly informative in terms of the actual attentional allocation. In contrast, comparing also the velocity and acceleration of the eyes with those of the targets (fig. 2B) revealed more accurately on which dot attention was allocated. This is consistent with the observation that subjects can track moving stimuli while maintaining their gaze confined in a narrow area and away from the stimuli (Hafed, Goffart, & Krauzlis, 2008). Therefore, we combined all these measures in a single value, namely the Mahalanobis distance (De Maesschalck, Jouan-Rimbaud, & Massart, 2000; Mahalanobis, 1936), to determine which dot was being tracked by attention. Specifically, we computed the Mahalanobis distance

between the position, velocity and acceleration of the eyes and those of each of the two dots.

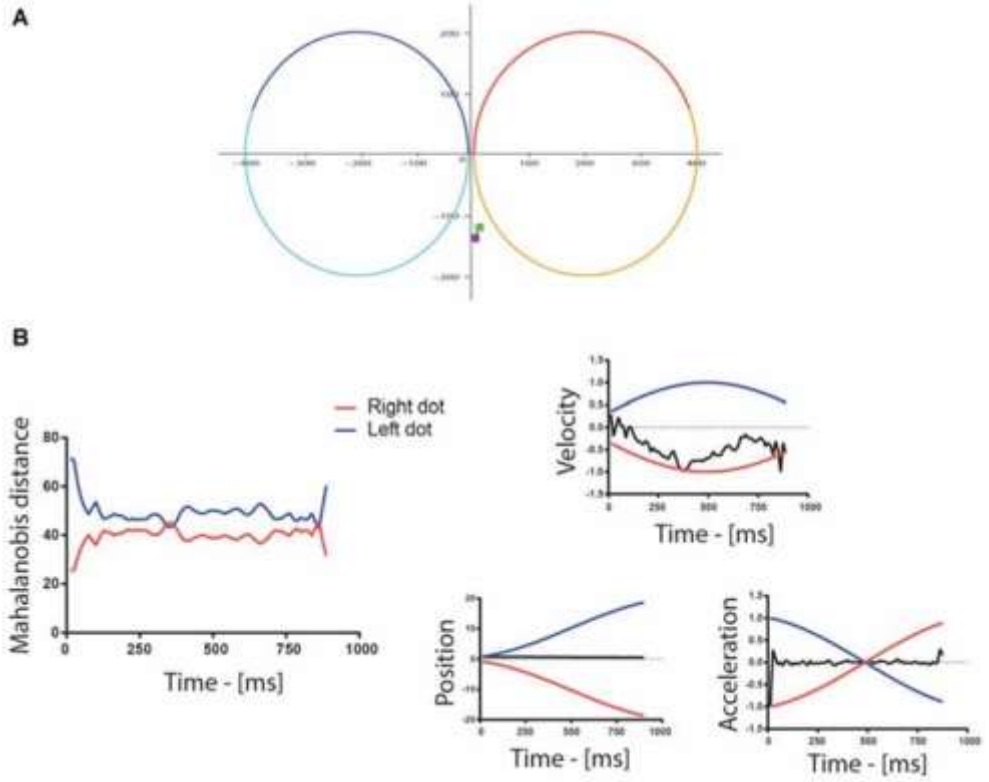


Figure 2 A – Example of eye movements during a trial: the colored spots represent the beginning (green) and end (purple) of the eye trace. In this particular example, the position of the eyes changed little throughout the trial. The red/orange and the blue/cyan traces are the paths of the two dots for the first/second part of the trial. B- The graph to the left shows the Mahalanobis distance between the eyes and the dots following the left (blue) and right (red) trajectories/paths in the first half of the trial. The Mahalanobis distance with the 2 dots was relatively large in this particular trial (around 42) because the position and acceleration profiles of the eyes differed strongly from the ones of the dots, as shown on the graphs on the right of the figure, which illustrate the normalized velocity, the position and the normalized acceleration of the dots (blue-left, red-right) and gaze (black). The position is reported in pixel, whereas the trial-wise normalization of velocity and acceleration was performed by dividing all values by the maximum of their absolute value. The velocity of the gaze, however, followed more closely the

velocity of the right dot, which indicated attention allocation to the right in this example trial.

A binary variable representing attentional allocation to the right/left was computed, assigning a value of one when the distance to the right/left dot was smaller than the distance to the left/right dot. This variable was then aligned either on the trial onset, or on the color change occurring around the middle of the trial and down sampled to 10 Hz in order to limit the number of time bins to analyze. For each 100-ms time bin, the total number of trials in which attention was allocated to the left/right dot was computed and used as a binomial variable in generalized linear mixed models (GLMM).

Statistical methods

All the analyses were performed with the SAS 9.3 Software (SAS Institute, Cary NC), by means of GLMM. We ran GLMM on log-transformed RT data, modeled as a normally distributed variable. In this case the variables included were PREDICTABILITY, differentiating the predicted from the non-predicted colors, PART, which considered if the target was in the first or in the second part of the trial, and BLOCK, a continuous variable from 1 to 8. The SUBJECT factor was considered in the random models, along with all the other factors.

We ran several GLMM on the eye position data as well. In all of them, eye position was modeled as a binomial variable (see above). In the first GLMM we analyzed whether attention was more likely to be allocated on the target dot in trials where it was present (see Fig. 4A). This analysis was performed mostly to confirm the validity of the Mahalanobis distance used in the present study, since we expected a strong attentional bias toward the target dot. The explanatory variables were TARGET-SIDE and TIME-BINS, the first one being representative of the position of the target (right or left side), the second one indicating the bin order (from 1 to 5, i.e. from 0 ms to

500 ms following stimulus display for the first half, and following color change for the second half). The residual covariance structure of the successive time bins was modeled with different variance parameters for each bin (variance component), in order to account for the correlations between successive bins. Because of limitations due to lack of convergence of the fitting algorithm used to optimize the GLMM, these analyses were performed separately for the first and second PART of the trials. In the second GLMM, we evaluated the allocation of attention to the predictive color when displayed together with a non-predictive color, while none of the stimuli were targets (see Fig. 4B). The dependent variable in this case was the attentional allocation to the predictive color, and the independent variables were SIDE and TIME-BINS (same convention as before). Finally, in the last GLMM, we compared the trials in which a target shown in the second part of the trial was either predicted or non-predicted, while non-predictive colors were displayed on the other side. In this case the dependent variable indicated attentional allocation to the target, and the explanatory variables were PREDICTABILITY and TIME-BINS (from 1000 ms before the color change to 500 ms after the color change; see Fig. 4C).

Finally, in order to investigate the timing of the effect of the color predictability, we also performed analyses on RT and attentional allocation restricted to the first block of the experiment, in which we split the first block in 3 sub-blocks of 36 trials each. Here the dependent variable was either the log-transformed and subject-wise normalized RT or the average of the attentional allocation variable over the whole duration of the first half of the trial. The explanatory variables were PREDICTABILITY and SUB-BLOCK.

Results

Reaction Time Analysis

The results of the GLMM analysis on RTs revealed a significant main effect of the factors PREDICTABILITY ($F(1,18)=5.46$, $p=0.0313$), and BLOCK ($F(7,126)=6.90$, $p<0.0001$). Specifically, the BLOCK effect consisted of a decrease in the RTs across blocks, as showed in figure 3 (fig. 3A shows RT from the first part of the trial, fig. 3B from the second part), while the PREDICTABILITY effect consisted of faster responses for predicted targets. Although the PART factor revealed no significant difference between the first and second part of the trial ($F(1,18)=1.30$, $p=0.2683$), the interaction between the factors PREDICTABILITY and PART was significant ($F(1,18)=6.20$, $p=0.0228$), revealing, as confirmed by pairwise comparisons, a difference between the predicted and non-predicted colors only in the second part of the trials (fig. 3B) ($t(1,18)=3.14$, Tukey-Kramer adjusted $p=0.0266$).

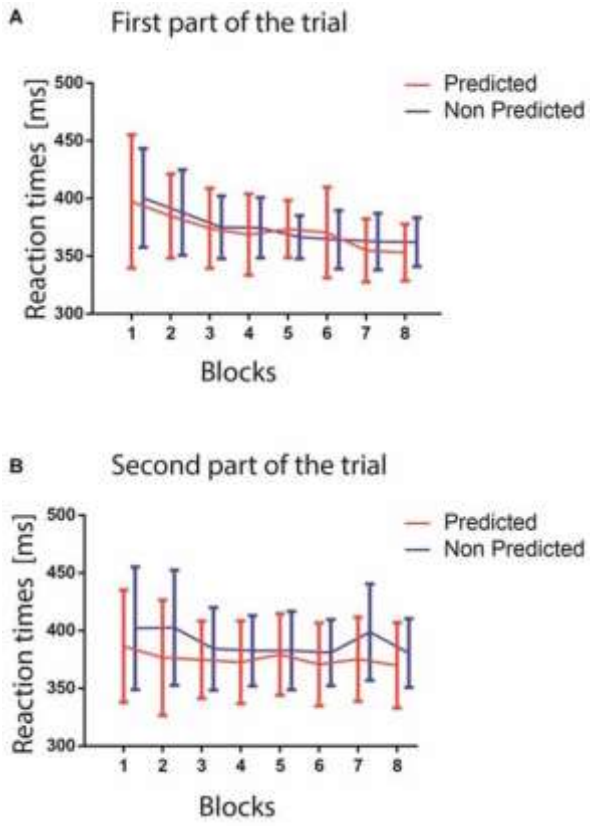


Figure 3 A – The upper figure shows the reaction times in ms for the two predictability conditions in the first part of the trials (red predicted colors, blue unpredicted colors). The blocks are represented on the x-axis. B -The lower figure shows reaction times as a function of the blocks for the second part of the trials (red predictable colors, blue unpredictable colors). In both panels, error bars represent standard errors of the mean computed within each block.

This last result confirmed that the participants learnt implicitly the color association between the stimuli displayed in the first and second parts of the trials, and that learning this association helped them to react faster to the occurrence of the target color when appearing in the second part. The lack of significant interaction between BLOCK and PREDICTABILITY ($p>.1$) suggested that statistical learning occurred very early during the experiment, as already reported in a previous study (Turk-Browne, Scholl, Chun, & Johnson, 2010).

Attention allocation analysis

As expected, when analyzing target trials, we found a progressive allocation of attention to the target (see Fig. 4A; main effect of TIME-BINS: first PART: $F(4,72)=27.31$, $p<0.0001$; second PART: $F(4,72)=41.38$, $p<0.0001$), irrespective of whether it was displayed on the left or right side of the screen (main effect of TARGET-SIDE: first PART: $F(1,18)=2.53$, $p=0.13$; second PART: $F(1,18)=0.87$, $p=0.36$; interaction: first PART: $F(4,72)=0.55$, $p=0.70$; second PART: $F(4,72)=0.62$, $p=0.65$). Tukey-corrected pairwise comparisons for the first part of the trial showed that the preferential allocation of attention to the target was significant only for the last bin (all comparisons between 5th bin and other bins: $p<0.001$), whereas for the second part, all pairwise comparisons were significant ($p<0.05$) except between the 1st and 2nd, 3rd and 4th and 4th and 5th bins. Overall, these results highlight the attentional capture by the target both in the first and second part of the trial, confirming the validity of the attention allocation measure used in the current study.

Moreover, they suggest that participants were slower in allocating attention to the target in the first part of the trial (fifth bin, around 450ms) than in the second part (from the third bin on, starting around 250ms). However, we cannot exclude that this unexpected discrepancy could also be explained by the fact that, early in the first part of the trial, targets were closer from each other than during the second part of the trial. This larger proximity between the stimuli could have hampered the sensitivity of our measure to detect differential attentional allocation to the target side. We then looked at the allocation of attention to the predictive stimuli when no target was present on the screen. We failed to find any significant change in attentional allocation to the predictive stimulus over time (see Fig. 4B; main effect of TIME-BINS: first PART: $F(4,72)=0.07$, $p=0.99$; second PART: $F(1,18)=0.13$, $p=0.7196$), irrespective of the stimulus side (main effects and interactions with the SIDE factor: all $p > 0.15$). The intercept of the model, representing the overall allocation of attention on the predicted stimulus irrespective of both TIME-BINS and SIDE was not significantly different from zero ($p>0.7$ in both PARTS). These results suggest that regularities, when they were not instrumental to the task at hand, did not capture attention. Finally, we compared trials in which a target appearing in the second part was either predicted or not by the color shown in the first part of the trial. Here the TIME-BINS factor included 15 bins, ranging from 1000 ms before to 500 ms after target onset. Since the target could appear as early as 750 ms after the trial onset, the bins ranging between -1000 ms and -700 ms contained slightly fewer trials than the later bins. We found that attention was preferentially allocated to the target when it was predicted (main effect of PREDICTABILITY: $F(1,18)=4.81$, $p=0.042$; see fig. 4C). There was also a significant effect of TIME-BINS ($F(14,252)=14.56$, $p<0.0001$), merely indicating the progressive allocation of attention to the target location, but no significant interaction ($F(14,252)=0.21$, $p=0.99$).

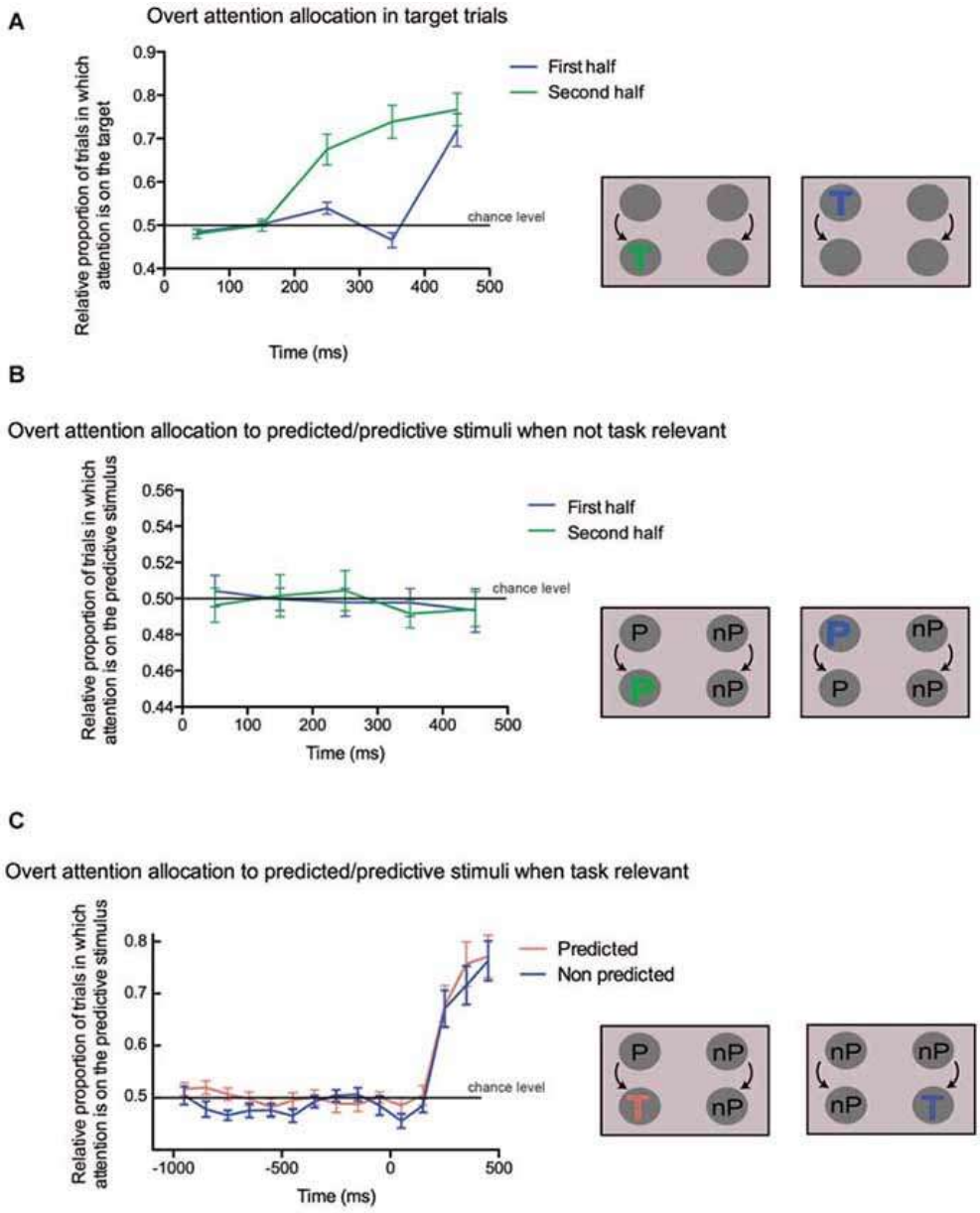


Figure 4 A – Overt attention allocation for the trials in which a target occurred either in the first half (blue line) or in the second half (green line) of the trial. Time is shown along the x-axis, the proportion of trials in which targets attracted attention is on the y-axis. The 0 value on the x axis corresponds

either to the beginning of the trial (first half) or to the change of color (second half). A schematic depiction of the analyzed conditions is shown on the right. The target could appear either in the first part (blue 'T') or in the second part (green 'T'). *B* - Overt attention allocation to predicted/predictive stimuli when not task-relevant. On the x-axis time is in ms, on the y-axis the proportion of trials in which predictive colors attracted attention is shown, indicated as the difference between the predicted and the non-predicted dots. A value of 0 on x-axis corresponds either to the beginning of the trial (first half) or to the change of color (second half). On the right, the predicted ('P') and non predicted ('nP') conditions are illustrated. *C* - Overt attention allocation to the predicted stimulus when task relevant. On the x-axis time is in ms, while on the y-axis there is the relative proportion of trials in which attention is captured by the predictive stimulus when the target is either predictive or not. A value of 0 on the x-axis corresponds to the time of the change of color. The two conditions are illustrated on the right of the panel: in the first case the target is preceded by a predictive color (red 'T'), whereas in the second case the target is preceded by a non-predictive color (blue 'T').

This lack of interaction suggested that attention might have been driven towards the predictive stimulus already in the first part of the trial. In order to confirm this hypothesis, we tested the same model but restricted to the first part of the trials (i.e. the first 8 bins) and still found a significant PREDICTABILITY effect ($F(1,18)=5.71$, $p=0.028$), with no significant interaction ($F(7,126)=1.43$, $p=0.20$). This confirms that attention was biased during the first part of the trial toward the colored dot, which predicted target occurrence in the second part.

Temporal dynamics of learning and selective attention in the first block

In order to determine when statistical learning effects started to influence the RT, we performed a GLMM restricted to the data from the first block only and considering the SUB-BLOCK (n=3, 36 trails each) and PREDICTABILITY as factors. We found no main effect of SUB-BLOCK ($F(2,29)=0.67$, $p=0.5193$) or PREDICTABILITY ($F(1,17)=3.80$, $p=0.0678$), but a significant interaction between these two factors ($F(2,251)=5.31$,

p=0.0055). The difference between the two PREDICTABILITY conditions became significant from the second sub-block (see fig. 5A, Tukey-corrected pairwise comparisons, all p≤0.05; t(251)=3.10, p=0.0260). In conclusion, around 72 trials seem to be enough to learn the color association between predictive and predicted colors.

Finally, we analyzed the effect of PREDICTABILITY on the allocation of attention to the target, restricted to the first block (and to time bins preceding the target onset; fig. 5B). We found again only a significant effect of PREDICTABILITY (F(1,18)=7.15, p=0.015), showing that the preferential allocation of attention to the predicted location of the target occurred very early.

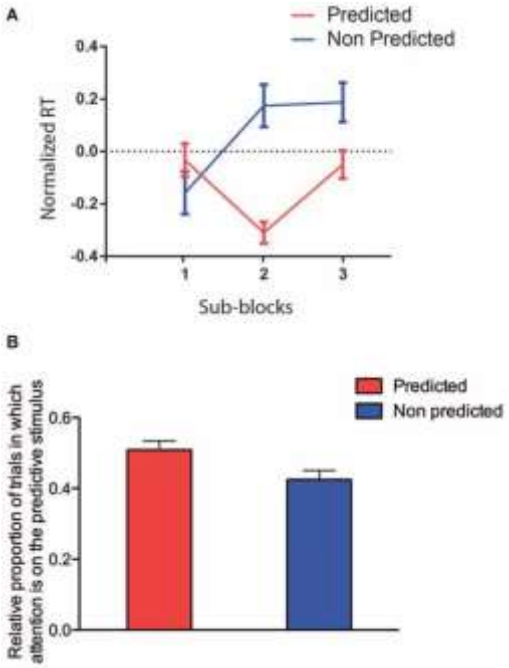


Figure 5 A – Reaction times restricted to the first block (red predictable colors, blue unpredictable colors). On the x-axis, the 3 sub-blocks are, on the y-axis the log-transformed and normalized reaction times are represented.

B – Attention allocation restricted to the first block (red predictable colors, blue unpredictable colors).

Discussion

In the current study we investigated the influence of statistical regularities on the allocation of visual attention in a color detection task. To do so, we manipulated the conditional occurrence of a sequence of colored stimuli, while recording the RT and eye position. None of the participants reported any awareness of the conditional occurrence manipulation, suggesting that it remained implicit during the whole experiment. As previously reported in the literature (Dale, Duran, & Morehead, 2012; Turk-browne et al., 2005), despite being implicit, the temporal predictability of the targets affected markedly the behavioral results, as revealed by shorter RTs in the detection task. Strikingly though, we found that this behavioral advantage was already measurable after a few dozen of trials, proving its remarkable efficacy.

Regarding visual attentional allocation, we report that predictability biases attention only when regularities are instrumental to the execution of the task. In other words, attention was biased by the regularities only when the target was predicted, whereas when the predicted stimulus was not the target, and therefore the statistical structure of the sequence was not helpful to perform the task, both predictive (first part of the trial) and predicted stimuli (second part of the trial) failed to either attract or divert attention. These findings are in agreement with the model proposed by Mackintosh (Mackintosh, 1975), in which predictable stimuli, when either rewarding or related to the behavior of the agent, attract attention.

In the context of the associative learning field, attention is considered as having different possible functions (Gottlieb, 2012): attention for learning,

attention for action and attention for liking. Attention for learning drives attention towards uncertain stimuli, in order to learn new regularities in the environment. Attention for action refers to the allocation of attention in order to optimize the achievement of a certain goal, while attention for liking is attracted to pleasurable and rewarding stimuli. Our finding that regularities attract attention, when they predict the target appearance, is in accordance with the concept of attention for action. On the other hand, predictive stimuli did not attract attention when they were predicting the occurrence of a distractor, even though they allowed inferring that the target would not appear on that side. This may be explained by the implicitness of the learning, which could have prevented participants to make such inference (Custers & Aarts, 2011). The attentional allocation to stimuli predicting the target appearance is also in accordance with attention for liking, since finding the target was always associated to a reward.

Regarding attention for learning, we found that, when focusing on the first block, only 72 trials were necessary for the RT effect to emerge. Similarly, the bias in visual attention toward the predictive stimuli reached significance in the first block as well. These findings seem to indicate that both effects appeared concurrently, in apparent contradiction with the hypothesis that attention would be biased towards potential sources of relevant information such as statistical regularities before any behavioral advantage emerges (Hogarth et al., 2008; Holland & Maddux, 2010), thus being causal in the development of this advantage. This suggests instead that both the faster detection of predicted targets and the preferential allocation of attention to the predicted targets are part of the same underlying statistical learning phenomenon. Obviously, it could also be argued that no temporal dissociation between the attentional allocation and the learning processes was found in the present study because of a lack of sensitivity either of our design or of our measurement method. Possibly, a more

complex statistical design, requiring a longer learning time, could allow us to dissociate the time course of the attentional allocation from the behavioral signature of learning.

Nevertheless, how this statistical learning process develops during the early stage of the task, either before or after attention becomes preferentially allocated to the statistical structure, remains unanswered. It could be proposed that statistical learning occurs pre-attentively, i.e. in the absence of attentional allocation (Li, 2000; Zénon, Hamed, Duhamel, & Olivier, 2009a). But since attention is thought to be necessary for the learning of the statistical regularities to occur (Turk-browne et al., 2005), it is likely that during the early phase of the task, learning takes place when, thanks to the random exploration of the scene by visual attention, attention is allocated by chance to the relevant location.

In the context outlined by our design, it seems that attention allocation is mainly biased by the optimization of the task's performance. This is in accordance with past computational studies proposing that visual selection is a mechanism involved in solving the inference problem of predicting the evolution of the environment to optimize task performance (Dayan et al., 2000; Kruschke, 2003). However, this conclusion is at odds with the one from a recent study showing that, even when irrelevant to the task, regularities attracted covert attention (Zhao et al., 2013). It is noteworthy that in Zhao's study, the difference between attentional allocation on informative and non-informative stimuli was small, and, most importantly, the regularities were present in a stimulus feature that was irrelevant to the task. In contrast, in our experiment, regularities involved the main feature used during the task, even when no target was present, i.e. the color of the dots. In addition, Zhao et al. investigated specifically covert attention by looking at discrimination performance while participants maintained their gaze on a central fixation point. In contrast, we evaluated attentional allocation by looking at eye

position, speed and velocity with respect to stimulus position, speed and velocity. This novel measure of attention allocation does not allow to dissociate overt from covert attention because, even tough, it relies on eye movements, it measures also attentional allocation to peripheral stimuli. In some instances, our measure indexes the fixation of gaze on one of the targets (i.e. overt attention), whereas in other cases we highlight attentional allocation to one specific peripheral location while the gaze remains fixed on a central position, corresponding to covert attention (Filali-Sadouk, Castet, Olivier, & Zenon, 2010; Zénon, Corneil, Alamia, Filali-Sadouk, & Olivier, 2014). The choice of using the eye velocity and acceleration signals, in addition to their position, to track the allocation of attention, was driven by the observation that observers can track moving stimuli while maintaining their gaze away from them. As reported in a study by Hafed and colleagues (Hafed et al., 2008), non-human primates are able to track an imaginary point located between two moving stimuli while maintaining the moving stimuli in peripheral vision. In our experiment, as illustrated in figure 2, even when the position of the eyes was kept at the center of the screen, the velocity profile revealed clearly that the subject was in fact actively tracking one of the two dots, thus revealing preferential attentional allocation. We confirmed the validity of our approach based on velocity and acceleration in the present study by showing, as expected, a strong bias to the target stimulus when displayed on the screen. More experiments will be needed to determine whether the discrepancy between our findings and the ones reported in the study of Zhao and colleagues is caused by the different attentional measures (covert versus overt) or by the task-relevance of the visual features guiding attention.

To summarize, this study shows how implicit learning, as many other cognitive processes (Brady, Konkle, & Alvarez, 2009; Estes, Evans, Alibali, & Saffran, 2007; Umemoto, Scolari, Vogel, & Awh, 2010; Zénon, Hamed,

Duhamel, & Olivier, 2009b), affects visual attention in a target detection task. When stimuli become predictable through the manipulation of conditional occurrences, these statistical regularities are learned very rapidly, and visual attention gets attracted to the informative stimuli when they are instrumental to the task at hand.

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences : An ERP study, *64*, 185–190. doi:10.1016/j.neures.2009.02.013

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5718–22. doi:10.1073/pnas.90.12.5718

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations : Predictable items are enhanced even when not predicted. *Cognition*, *129*(2), 205–211. doi:10.1016/j.cognition.2013.07.003

Brady, T. F., Konkle, T., & Alvarez, G. a. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology. General*, *138*(4), 487–502. doi:10.1167/8.6.199

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6.

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

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. doi:10.1037/0278-7393.31.1.24

Custers, R., & Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and Cognition*, *20*, 368–378. doi:10.1016/j.concog.2010.05.011

Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology / University of Finance and Management in Warsaw*, *8*(2), 196–209. doi:10.2478/v10053-008-0115-z

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3 Suppl*(november), 1218–23. doi:10.1038/81504

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*, 1–18. doi:10.1016/S0169-7439(99)00047-7

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*, 254–260. doi:10.1111/j.1467-9280.2007.01885.x

Filali-Sadouk, N., Castet, E., Olivier, E., & Zenon, A. (2010). Similar effect of cueing conditions on attentional and saccadic temporal dynamics. *Journal of Vision*, *10*(4), 21.1–13. doi:10.1167/10.4.21

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. doi:10.1111/1467-9280.00392

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, *76*(2), 281–95. doi:10.1016/j.neuron.2012.09.034

Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2008). Superior colliculus inactivation causes stable offsets in eye position during tracking. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(32), 8124–37. doi:10.1523/JNEUROSCI.1317-08.2008

Hogarth, L., Dickinson, A., Austin, A., Brown, C., & Duka, T. (2008). Attention and expectation in human predictive learning: the role of uncertainty. *Quarterly Journal of Experimental Psychology (2006)*, *61*(11), 1658–68. doi:10.1080/17470210701643439

Holland, P., & Maddux, J. (2010). Brain systems of attention in associative learning. *Attention and Associative Learning: From Brain to Behaviour*, 305–350.

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

Kruschke, J. K. (2003). Attention in Learning, *Current Directions in Psychological Science*, **12,** 171–175.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*, 193–243. doi:10.1080/02724990344000141

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology. Animal Behavior Processes*, *37*, 220–229. doi:10.1037/a0021384

Li, Z. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, *13*, 25–50. doi:10.1163/156856800741009

Little, D. Y., & Sommer, F. T. (2013). Learning and exploration in action-perception loops. *Frontiers in Neural Circuits*, *7*(March), 37. doi:10.3389/fncir.2013.00037

Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y., & Dosher, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory*, *95*(2), 145–51. doi:10.1016/j.nlm.2010.09.010

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*. doi:10.1037/h0076778

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55.

O'Brien, J. L., & Raymond, J. E. (2012). Learned predictiveness speeds visual processing. *Psychological Science*, *23*(4), 359–63. doi:10.1177/0956797611429800

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

Pearce, J. M., & Mackintosh, N. J. (2010). Attention and Associative Learning: From Brain to Behaviour - Chris J. Mitchell, Mike E. Le Pelley

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning : one phenomenon , two approaches, *Trends in Cognitive Sciences 10*(5). doi:10.1016/j.tics.2006.03.006

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

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, *52*(5), 582–588. doi:10.3758/BF03206720

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2). doi:10.1016/j.cognition.2005.01.006

Turk-browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional Visual Statistical Learning, *Journal of Experimental Psychology: Learning, Memory and Cognition. 34*(2), 399–407. doi:10.1037/0278-7393.34.2.399

Turk-browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning, *Journal of Experimental Psychology: General. 134*(4), 552–564. doi:10.1037/0096-3445.134.4.552

Turk-browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2010). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness, *Journal of Cognitive Neuroscience: 21*(10), 1934–1945. doi:10.1162/jocn.2009.21131.

Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 1419–1429. doi:10.1037/a0019324

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848. doi:10.1167/1.3.467

Zénon, A., Corneil, B. D., Alamia, A., Filali-Sadouk, N., & Olivier, E. (2014). Counterproductive Effect of Saccadic Suppression during Attention Shifts. *PloS One*, *9*(1), e86633. doi:10.1371/journal.pone.0086633

Zénon, A., Hamed, S. Ben, Duhamel, J., & Olivier, E. (2009a). Visual search without attentional displacement. *Journal of Vision*, *9*(11), 9.1–15. doi:10.1167/9.11.9

Zénon, A., Hamed, S. Ben, Duhamel, J. R., & Olivier, E. (2009b). Attentional guidance relies on a winner-take-all mechanism. *Vision Research*, *49*(12), 1522–1531. doi:10.1016/j.visres.2009.03.010

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–77. doi:10.1177/0956797612460407

Chapter V: The eye pupil responds to unconscious prediction error

This chapter is a modified version of an article in preparation:

Alamia, A., Pasqualotto E., Mouraux, A. & Zenon, A. . “The eye pupil responds to unconscious prediction error”.

Abstract

Pupillary responses to cognitive processes have been investigated since the early 60s, when the first seminal experiments linking pupil size and arousal were conducted by Hess and Polt. Nowadays it is well known that multiple cognitive events, such as emotional responses, individual biases, or memory retrieval, trigger changes in pupil size. Across the varieties of these cognitive determinants, surprise, or equivalently prediction errors, appear to be potential common sources to pupil responses. Whether this response is elicited by the processing of explicit, surprising events, or whether this effect can be observed irrespective of conscious perception, is still unknown. In the present study we performed 4 experiments in which we tested whether pupillary dilation can be observed in response to unconscious surprising events. In 4 experiments, participants watched a stream of letters modeled as a probabilistic Markov chain in which each letter could be followed either by a likely (47.5%) or rare (5%) transition. Applying a method based on an autoregressive exogenous model (ARX), we found that pupil size increased following rare events, even though participants had no explicit awareness of the transition frequencies. Withdrawing attention from the letter stream suppressed the pupil response, indicating that the pupil response did not depend on awareness but required active attention. Surprisingly, rare events were rated as more familiar than likely ones, and this increased familiarity correlated with the magnitude of the pupil response, suggesting that the increase in arousal following rare events favored memory encoding. Finally, electrophysiological recordings shed some light on the neuronal mechanisms involved in such responses, showing that pupillary responses correlated with 1) right-frontal electroencephalographic event-related

potentials, in accordance with the possible role of the frontal regions classically involved in statistical learning in the detection of the surprising transitions and 2) with changes in gamma activity in the same region, suggesting that this signal corresponds to bottom-up prediction error. Overall, these findings indicate that pupil size reacts to unconscious prediction errors, and significantly contribute in delineating how the brain processes unconscious information.

Introduction

Other than responding to variations in luminance (De Groot & Gebhard, 1952), pupil size has been shown to be affected by spatial attention (Binda, Pereverzeva, & Murray, 2014; Gabay, Pertzov, & Henik, 2011; Mathôt, van der Linden, Grainger, & Vitu, 2013, 2014) emotions (Bradley, Miccoli, Escrig, & Lang, 2008; Partala & Surakka, 2003), individual biases (de Gee, Knapen, & Donner, 2014), physical effort perception (Zénon, Sidibé, & Olivier, 2014), and working memory loads (Hess & Polt, 1964; Van Gerven, Paas, Van Merrienboer, & Schmidt, 2004), to cite but a few of the factors that appear to drive pupil responses. This variety of responses could reflect the involvement of a common fundamental mechanism. One candidate is surprise or prediction error, i.e. the perception of a sensory event whose occurrence was considered unlikely given the internal model. Indeed, several experiments have shown that stimuli that have low probability under the current model or assumptions of the individual tend to trigger pupil responses (Nassar et al., 2012; Preuschoff, 't Hart, & Einhäuser, 2011; Raisig, Welke, Hagendorf, & van der Meer, 2010). However, these results were obtained in the context of a behavioral task, making it impossible to disentangle: 1) whether pupil responds to the need of adapting behavior to new contingencies, or 2) whether pupil responds to surprise in a more fundamental way, irrespective of task-relevance and behavior. This distinction is important because it implies very different levels of generality of the involvement of the pupil response.

One way to address this question is to investigate the pupil response to unconscious violations of predictions. Such unconscious predictions occur for instance in the context of statistical learning, the process through which

we learn the statistical regularities of the environment (Kim, Seitz, Feenstra, & Shams, 2009; Turk-browne, Scholl, Chun, & Johnson, 2010). This process has been reported for the first time in the auditory domain, showing that 8-months babies were able to learn differences in statistical relationship between non-meaningful syllables (J R Saffran, Aslin, & Newport, 1996); since then statistical learning has been evidenced in many other studies, for example in the context of language acquisition (Arciuli & Torkildsen, 2012; Jenny R. Saffran & Saffran, 2003), but also in visual processing (Alamia & Zénon, 2016; Turk-browne, Isola, Scholl, & Treat, 2008; Turk-browne, Junge, & Scholl, 2005).

In the present study, we tested whether pupil size reacts to the violation of statistical regularities when these are not consciously perceived as such by participants. In a series of four experiments we tested whether the occurrence of a rare transition would lead to pupil dilation. After confirming this prediction in experiment 1, we investigated the role of attention in experiment 2, and then tested the degree of awareness of the statistical rules in experiment 3. Finally, we investigated the neural correlates that drive this effect by means of electrophysiological measures (EEG) in experiment 4.

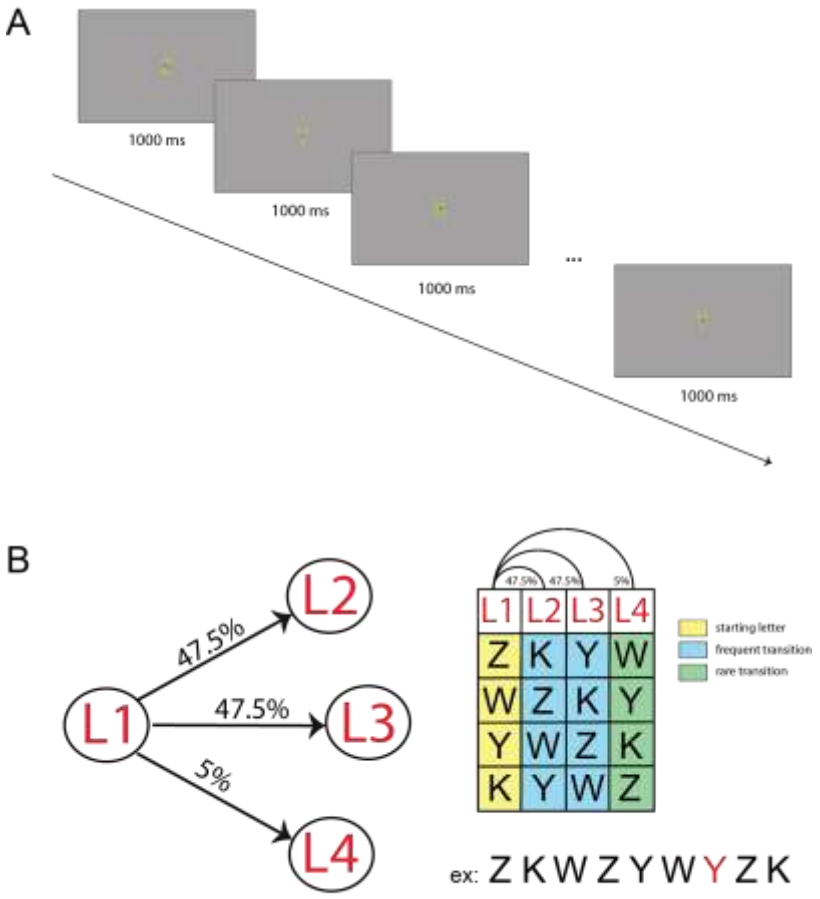


FIGURE 1

Fig.1: Experimental design. A) Each trial was composed of a series of displays containing a letter over-imposed on a fixation point. The letters changed every second. B) The Markovian process driving the letter sequence. Each letter could transition to 3 other letters, two of which being frequent (47.5% chance) and one being rare (5% chance). Below the table, an example of sequence is provided, in which the rare transition is shown in red.

Results

Pupil dilates in response to rare events (experiment I)

In the first experiment, 14 participants (7 females, mean age=29.53 years, SD=3.72) were asked to passively stare a stream of letters for 3 blocks, each 5 minute-long (figure 1A, see Methods). The letters changed every second, according to a probabilistic Markovian process (figure 1B, see Methods): according to these rules, a rare transition occurred 5% of the time. Notably, none of the participants were able to report explicitly the rules (or any subset of rules) in all the experiments we performed. The pupil traces were analyzed by means of an autoregressive model, in which the letter onset and the rare transitions were modeled as exogenous inputs (ARX model, see Methods and (Zénon, 2017) for details). The pupil dilated in response to the onset of the rare events (GLMM analysis: intercept of the model: $t(13)=2.20$, $p=.0462$; see Fig. 2A), and this effect increased block by block, as confirmed by the BLOCK factor ($F(1,13)=5.71$, $p=.0327$).

Attention affects pupil responses (experiment II)

In the second experiment, a different set of 14 participants (4 females, mean age=28.64 years, SD=5.06) performed two sessions of 3 blocks each. In the first one, a fifth letter was also displayed, with no statistical relationship to the four letters included in the Markovian process, and participants were instructed to report the occurrence of this letter with a mouse click. This

setting ensured that participants maintained attention on the letter stimuli. In the second session, participants were instructed to click whenever the fixation point changed color, allocating attention away from the letters. Both the target letter and the change in the dot color happened randomly once every minute, and were independent of the statistical rules dictating the letters sequence. The order of the two sessions was pseudo-randomized between participants. Interestingly, the GLMM performed on the signed absolute maximum value of the impulse response revealed a dilation of the pupil to the rare events only when attention was allocated to the stream of letters (intercept of the model $t(13)=3.44$, $p=0.0044$, BLOCK main effect: $F(2,26)=3.50$, $p=.0449$) but not when attention was focused on the fixation point (intercept of the model $t(13)=-0.02$, $p=.9855$, BLOCK main effect: $F(2,26)=0.64$, $p=.5329$) as shown in figure 2B. These results suggest that attention plays an important role in the occurrence of the effect, probably by affecting the degree of statistical learning of the participants.

Relationship between awareness of the statistical association and pupil size (experiment III)

In the third experiment, a new pool of 18 participants (12 females, mean age=24.77 years, SD=2.88) performed 6 blocks of the same attention-on-letter condition task as in experiment II. Two awareness tests were performed after the letter detection task in order to assess the awareness of the statistical associations. In the familiarity test, participants had to provide a rate of familiarity (1 to 10) to all of the possible letters' transitions, whereas in the generative test, participants were asked to guess which transition

should follow the presentation of a pair of letters (see methods for further details). The GLMM performed on the peak of the impulse response from the ARX model confirmed the presence of the pupil response to the rare transitions (figure 2C) (intercept $t(17)=2.99$, $p=.0082$). However, the BLOCK effect ($F(1,17)=6.26$, $p=.0228$) showed a decrease in the pupillary response across blocks, in contrast to what we found in the first experiment.

The results from the familiarity test, in which participants were asked to judge the familiarity of the transitions from one letter to the other, are shown in figure 3A. Surprisingly, participants rated the rare transitions as more familiar ($F(1,17)=8.80$, $p=.0087$). In the generative task, in which participants were asked to choose the following letter given a transition, we observed a significant preference for choosing the subsequent letter according to the rules ($F(2,34)=5.98$, $p=.0060$): in other words, participants were more likely to choose a letter that complied with the statistical association (figure 3B), thus showing some degree of learning in retrieving the association (Tunney & Shanks, 2003). Importantly, at the end of the experiment participants were asked to report any rule or regular pattern that they may have noticed, but failed in reporting any transitions.

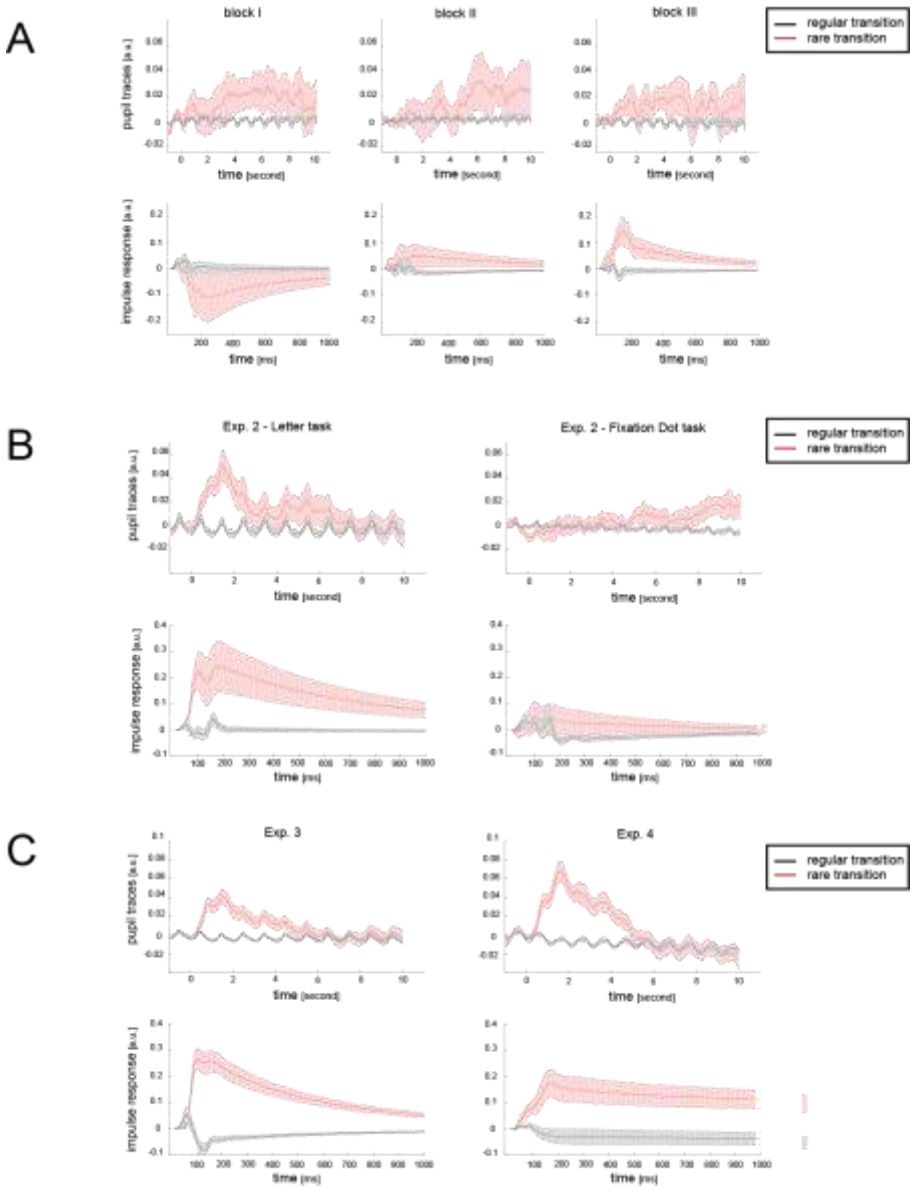


FIGURE 2

Fig.2: Results. A) Average pupil traces (Upper part) and average impulse responses (lower part) in response to rare (red) and frequent events (red) in experiment 1. B) Pupil responses in Experiment 2, same convention. C) Pupil responses in Experiment 3 (left) and 4 (right), same convention.

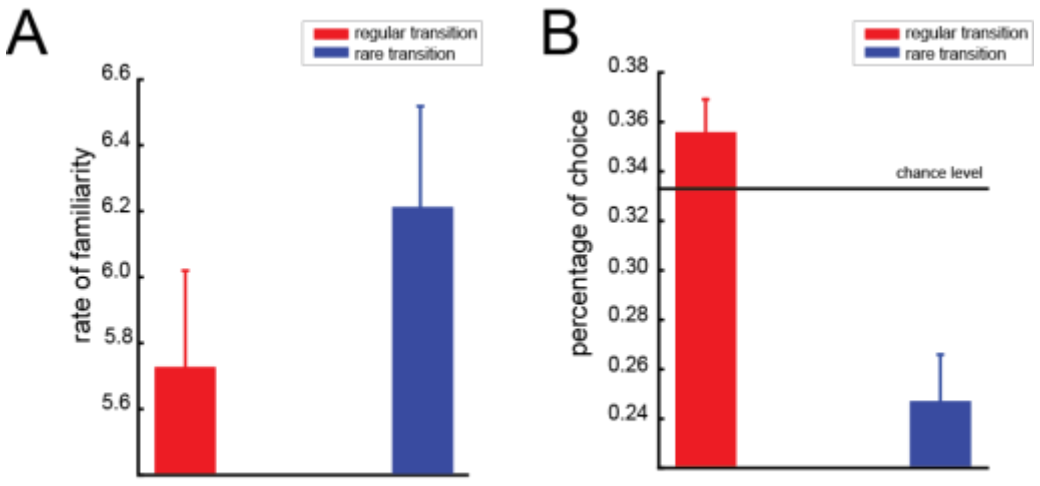


FIGURE 3

Fig.3: A) Results of the Awareness test. Average familiarity rate for the frequent (red) and rare transitions (blue). Error bars represent standard errors of the mean. B) Results of the generative task with the percentage of times participants chose the frequent (red) and rare transition (blue). The transitions that were chosen by the participants (±5% of responses) but never actually occurred are not displayed.

Moreover, we tested between-subject correlations between the pupil responses and the other behavioral measures (RT, generative and familiarity tests) (see figure 4). The amplitude of the impulse response to the rare event correlated with the difference between the familiarity score of the frequent and the rare transitions (fig 4A - Spearman correlation: r=-0.5199, p=.0324), such that larger pupil responses to rare transitions were associated with larger differences between familiarity rates. This suggests that the more the rare transition increased arousal (as tracked by pupillary responses), the more likely the participants were to judge it as familiar. Moreover, we found

that the magnitude of the pupil response to the rare events correlated with the letter detection RT (fig. 4B; Spearman correlation: r=-0.4902, p=.0478). The interpretation of this finding is uncertain but suggests that the degree of engagement in the task may have affected both variables: the more the participants were attentive (i.e. faster responses), the better they learned the statistical transitions (i.e. larger pupil responses). Finally, we found a significant correlation between the response of the pupil to the target and to the rare events (fig. 4C - Spearman correlation: r= 0.7381, p=.0197).

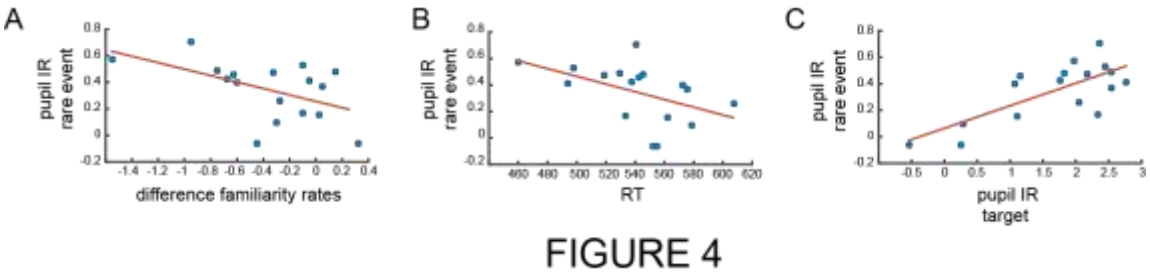


FIGURE 4

Fig.4: Correlations. The figure shows the correlation (A) between the impulse response (IR) to the rare events and the familiarity difference, (B) between the IR to the rare event and the reaction time, and (C) the IR to the rare event and to the target.

Electrophysiological results (experiment IV)

In the fourth experiment, 17 participants (11 females, mean age=22.41 years, SD=2.58) performed 3 blocks of the letter detection task (as in exp.2 and 3). The objective of the last study was to investigate the possible neural correlates of the pupillary response. As in the previous

experiments, we found a pupil response to the rare transitions (figure 2C) (intercept t(16)=2.579, p=.0199), but no effect of the BLOCK factor (F(1,16)=2.54, p>.10).

We investigated both the temporal and the time-frequency domain of the EEG recordings. First, we tested a GLMM considering as dependent variable the signed area of the Evoked Response Potential (ERP) elicited by the rare and regular transitions (i.e. factor CONDITION) in the time window 50-200ms, in the right and left frontal clusters of electrodes (i.e. factor REGION – see methods for details). The results (see figure 5A) revealed a significant interaction between the two factors (F(1,16)=12.57, p=.0027). The post hoc analysis showed a difference between the evoked-related potentials to the rare and frequent conditions specifically in the right frontal hemisphere (t(16)=2.66, p=.0170), revealing a larger negativity for the traces elicited by the rare transitions. This result is in accordance with recent findings that highlight the role of the frontal regions (Karuza et al., 2013; Turk-Browne, Scholl, Chun, & Johnson, 2009) and specifically of the right hemisphere in statistical learning (Janacsek, Ambrus, Paulus, Antal, & Nemeth, 2015; Roser, Fiser, Aslin, & Gazzaniga, 2011). We also tested a similar model in the time window 250ms-550ms considering the electrodes Oz and POz (factor REGION) and the factor CONDITION. The results showed a significant effect of CONDITION (F(1,16)=6.43, p=.0220), revealing a larger positivity for the trace elicited by the rare transitions. Moreover, we found that the difference between signed areas (rare - frequent) in the right frontal electrodes correlated with the peaks of the impulse response of the pupil dilation (Pearson: r=.7178, p=.0012; Spearman r=.5980, p=.0129 – see figure 5B), indicating that larger differences in the EEG traces were accompanied by larger pupillary responses. No correlation was found regarding the frontal region of the left hemisphere (p>0.10).

Regarding the time-frequency domain, for each subject we computed spectrograms in the electrodes of the frontal right region highlighted in the time-domain analysis, aligned on the onset of frequent and rare letter transitions (see figure 6A). We tested a GLMM considering the average power for all the frequency bands (factor BAND: θ , α , β and γ) in the time interval 50-200ms (same time-interval considered in the time-domain analysis) for the rare and frequent transitions (factor CONDITION). As expected, the results revealed a difference between bands ($F(3,54)=18.82$, $p<0.0001$), and a significant interaction between the factors ($F(3,54)=4.61$, $p=.0061$), whereas the factor CONDITION revealed a trend ($F(1,18)=3.89$, $p=.0642$). A post-hoc analysis on the interaction revealed a larger power for the rare transitions in the theta ($t(54)=-3.49$, $p=0.0010$), alpha ($t(54)=-2.18$, $p=0.0340$) and gamma band ($t(54)=-2.94$, $p=0.0048$). Furthermore, we found a significant positive correlation between the difference in the gamma band in the 50-200ms time-window and the pupillary response to the rare transitions (Pearson: $r=0.5627$, $p=0.0187$; Spearman $r=0.5074$, $p=0.0397$ – see figure 6B).

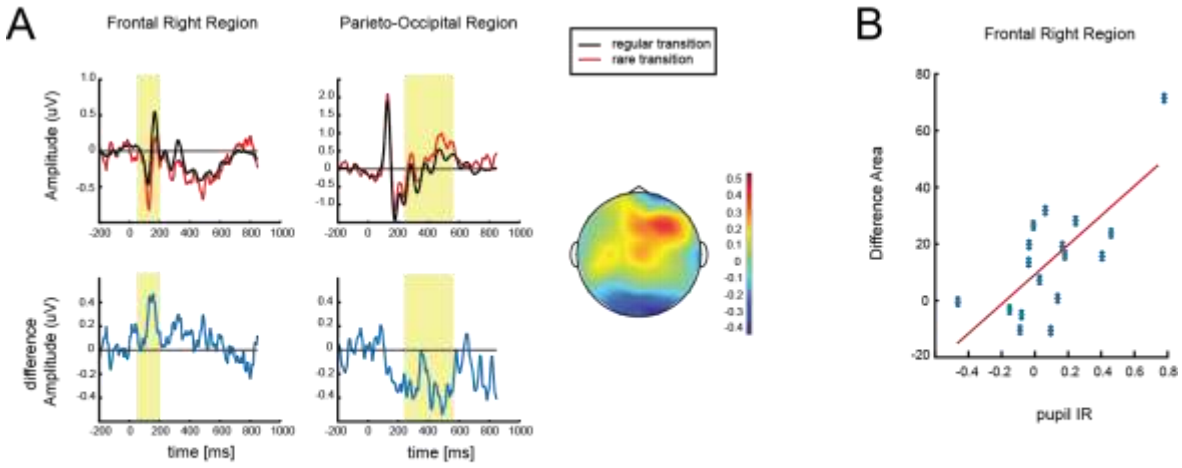


FIGURE 5

Fig.5: Electrophysiological results: ERPs. A) In the first row the ERPs elicited by the rare (red) and frequent (black) transitions are shown for the frontal right region (first column) and occipital (second column) regions. In the second row the difference between the response to frequent and rare letters is shown for each region. In both graphs the time window included in the analysis is highlighted in yellow. The scalp topography is shown for the first time-window (50ms – 200ms). B) Correlation between the difference in the EEG responses in the frontal right region (on the Y axis) and the peak of the impulse response of the pupil to the rare transition (on the X axis).

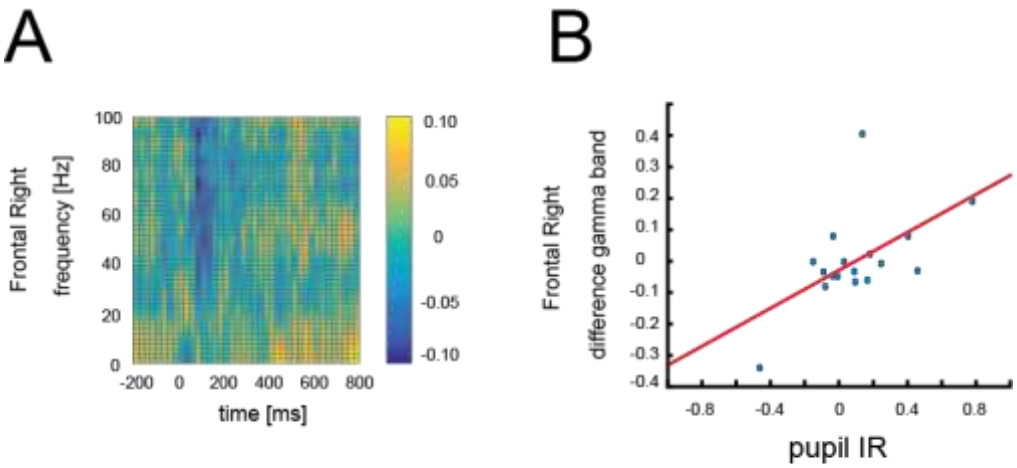


FIGURE 6

Fig.6: Electrophysiological results: Spectrograms. A) The figure shows the difference between the spectrograms obtained from the right frontal region's electrodes in the trials in which there was a frequent transition and those in which there was a rare transition. B) Correlation between the peak of the impulse response of the pupil to the rare transition (on the X axis) and the averaged difference (on the X axis) obtained from the spectrogram in the time window 50-200ms in the gamma band.

Discussion

Our results show that pupil size increases in response to the violation of expectations, even when the predictions occur unconsciously. While pupil responses to surprising or novel events have been reported in previous studies (Kloosterman et al., 2015; Lavín, San Martín, & Rosales Jubal, 2014; O'Reilly et al., 2013; Preuschoff, 't Hart, & Einh??user, 2011), the present findings allow us to conclude that this pupil response does not arise from the conscious feeling of surprise, or from the need to adapt behavior to new

contingencies, but rather from fundamental prediction error signals computed in the brain. In the study from Nassar and colleagues (Nassar et al., 2012), pupil diameter increased in response to unexpected surprise, i.e. when participants detected a change in the environment. O'Reilly and colleagues (O'Reilly et al., 2013) later showed that pupil size increased in response to surprising events while this response was diminished when the internal model of the task needed to be updated, already hinting that pupil size does not index behavioral adaptation, but rather surprise per se, in agreement with the present findings.

The pupillary response to prediction error, while occurring outside the scope of awareness, was modulated by attentional allocation. This finding may be interpreted either as evidence that statistical learning occurs only when stimuli are under the focus of attention (Turk-browne, Junge, & Scholl, 2005), or as a sign that the pupil response to prediction error is itself modulated by attentional allocation. This last proposition would fit with the framework of predictive coding, according to which attention modulates the precision of the sensory signal, and applies a gain on the prediction errors (Den Ouden, Kok, & de Lange, 2012; Feldman & Friston, 2010). Given that the attentional manipulation in the present study was applied both during learning and during the assessment of the pupil response to the violation of the rule, we are unable to disentangle these two hypotheses. Nevertheless, and as already hypothesized elsewhere (Koch & Tsuchiya, 2007; Tsuchiya & Koch, 2009), these findings provide further evidence in favor of the hypothesis that attention and consciousness have to be thought of as two different cognitive processes, possibly based on different neuronal mechanisms.

Interestingly, the familiarity test in the third experiment revealed that the rare events were reported as more familiar by the participants in our study. The increased pupil response to rare events provides a clue as to the possible origin of this counter-intuitive finding: violations of expectations increased

arousal, and arousal has a strong influence on the ability to encode new events in memory (Cruciani, Berardi, Cabib, & Conversi, 2011; Naber, Frässle, Rutishauser, & Einhäuser, 2013; Nielson & Bryant, 2005). It is noteworthy that this preferential memorization of the rare event did not impact the responses in the generative task, suggesting that the two tasks rely on different types of memory encoding, akin to the distinction between familiarity and recollection in recognition memory (Jacoby, 1991).

Finally, the fourth experiment aimed at investigating the neural substrates responsible for the pupil dilation. Previous studies have shed some light on the neural correlates of statistical learning, both in an auditory segmentation task (Karuza et al., 2013) and in a visual task (Turk-Browne et al., 2009). In the fMRI study from the Turk Browne's group, sub-cortical (i.e. caudate and hippocampus) and cortical regions (i.e. ventral occipital-temporal cortex and lateral occipital cortex) were found activated in a block-analysis contrasting regularities versus non-regularities. Conversely, in the work of Karuza and colleagues, the main cortical area responding to statistical regularities was the left inferior frontal gyrus, i.e. Broca area. Interestingly, this area, which is notoriously associated with language processing (Novick, Trueswell, & Thompson-Schill, 2010), seems to play a crucial role also in hierarchical processing (Alamia et al., 2016; Tettamanti & Weniger, 2006). Two more studies, one with callosotomy patients (Roser et al., 2011) and one using a tDCS approach (Janacsek et al., 2015), have already advocated the crucial role of the frontal right hemisphere in statistical learning, in agreement with our results.

A possible neural correlate, responsible for the larger negativity we found in the right-frontal electrodes in response to rare transitions, could be the Anterior Cingulate Cortex (ACC). ACC is notoriously linked to error related negativity (ERN) (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter, 1998; Holroyd & Coles, 2002; Ridderinkhof, Ullsperger, Crone, &

Nieuwenhuis, 2004). Several studies have already associated ERN with errors that escaped the scrutiny of awareness (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; Orr & Hester, 2012; Wessel, 2012). Furthermore, ACC is densely connected to Locus Coeruleus (LC, the main noradrenergic nucleus in the human brain), and it is well-known to encode surprise-related signals (Carter, 1998). The LC is currently thought to be one of the main actors driving the responses in the pupil (Aston-Jones & Cohen, 2005; Eldar, Cohen, & Niv, 2013; Gabay et al., 2011), even though the cholinergic basal forebrain system is also thought to play an important role (Reimer et al., 2016).

Importantly, our findings in the time-frequency domain shed some light on the computational mechanisms that drive the pupillary response. It has been recently shown that gamma-band activity (30-100hz) can be related to prediction-error computations, whereas beta-band activity (12-30Hz) has been associated with prediction signals (Sedley et al., 2016; van Pelt et al., 2016). Further than the effect we reported in the gamma-band, the correlation between the pupillary responses and the difference in the theta band between rare and frequent transitions, provides robust evidence in favor of the hypothesis that pupillary response is tracking prediction errors' computations, supposedly driven by ACC activity (Brown & Braver, 2005; Garrison, Erdeniz, & Done, 2013). Moreover, our results in the temporal domain (N100) in the right-frontal region, are also reminiscent of another effect known as mismatch negativity (MMN) (Pekkonen, Rinne, & Näätänen, 1995; Shelley et al., 1991). MMN has been usually associated with predictability in auditory stimuli (Garrido, Kilner, Stephan, & Friston, 2009), and interpreted as a prediction error signal (Stefanics, Kremláček, & Czigler, 2014). Recently, it has also been reported using visual stimuli (Kimura, 2012), and a recent study from Koelsch and colleagues showed an effect in frontal electrodes related to transitional probabilities (Koelsch, 2016). Finally,

our electrophysiological results have remarkable similarities with the findings reported by Sanders and colleagues (Sanders, Newport, & Neville, 2002) in a study investigating the effects of statistical learning in discriminating meaningless words. In their work, these authors reported a larger ERP following a training session both in the frontal and in the posterior electrodes, with latencies similar to the ones we investigated (around 100ms for the frontal and around 400ms for the posterior electrodes). Moreover, in a different study, Sanders and colleagues obtained the same enhancement in the N100 negativity, interpreted as a marker for speech segmentation also elicited by non-sense words (Sanders & Neville, 2003). Similarly, Abla and Okanoya (Abla & Okanoya, 2009), who investigated the learning of triplets of visually displayed shapes, reported an effect of statistical regularities in the frontal electrodes with a latency of 400ms only in a subset of participants (defined as good learners) and specifically at the beginning of the session, whereas the subset labeled as middle learners reported a similar effect at the end of the experiment. Unfortunately, in our study the scarcity of rare transitions prevented us from investigating the dynamic of the learning block by block. All in all, our findings suggest that pupillary responses may be considered as a marker of prediction-error computations in frontal regions.

Conclusion

In conclusion, we show that pupil size responds to the violation of unconscious statistical regularities, and that this response presumably originates from right frontal cortex. This response does not require awareness but is affected by attention allocation. Since behavioral markers of implicit statistical learning consist usually of modest differences in RT or familiarity judgements (Barakat, Seitz, & Shams, 2013; Turk-browne et al.,

2005), the present results provide us with a robust marker of statistical learning processes, unbiased by intentional, behavioral responses.

Methods

Participants

Sixty-three healthy participants (exp 1: 14 participants, 7 females, mean age=29.53 years, SD=3.72; exp 2: 14 participants, 4 females, mean age=28.64 years, SD=5.06; exp 3: 18 participants, 12 females, mean age=24.77 years, SD=2.88; exp 4: 17 participants, 11 females, mean age=22.41 years, SD=2.58) took part in the study, receiving monetary compensation for their participation. The initial sample size of 14 participants for the first experiment was chosen *a priori.* We used the same sample size in the second experiment but a larger sample size in Experiment 3 and 4, to ensure statistical power given that we expected negative results in the familiarity and generative tests, and to compensate for the limited number of trials in the rare condition for the electrophysiological analysis. In all the experiments, we stopped acquiring subjects when we reached the sample size planned prior to starting the experiments. All participants reported normal or corrected-to-normal vision. All experiments were carried out according to the Declaration of Helsinki and were approved by the Ethics Committee of the Université Catholique de Louvain. Written informed consents were obtained from all the participants, who received monetary compensation for their participation.

Procedure

The experiment took place in a dim room, with the participants sitting in front of a 19” CRT screen with a 100Hz refresh rate. The distance between the screen and the chin support was 58 cm, while the height of the chin support was adjusted to each participant to ensure a comfortable position. The task was implemented using the version 3.0.9 of the Psychotoolbox (Brainard, 1997) in Matlab 7.5 (The MathWorks, Inc. Natick, Massachusetts, United States).

In all experiments, each block began with the display of a black fixation point in the center of the screen, over-imposed on a letter (figure 1A). The letters were 1° wide and 2° high. In experiment 2, participants were pseudo-randomly assigned to one of two sets of 4 letters (set 1: ZYKW, set 2: RLMQ), whereas in all the other experiment they were assigned to the first set of letters. The letter, displayed in yellow (RGB: .9 .9 .7) over a gray background (RGB .7 .7 .7), changed with a constant rate of one letter per second. The stream of letters followed a probabilistic Markovian process, as reported in figure 1B: two transitions had a 0.475 probability of occurring (from L1 to L2 and to L3), whereas the last one had a probability of 0.050 (L1 to L4).

In the first experiment, composed of three blocks of 5 minutes each (300 transitions), participants were instructed to passively look at the stream of letters, staring at the fixation point. They were informed about the purpose of the experiment, and the presence of statistical rules driving the stream of letters. Nevertheless, none of the participants at the end of the experiment could report the rules, or any specific pattern. In experiment 2, participants performed 2 sessions: each session was composed of three blocks of 5 minutes each (300 transitions), and a short break of few minutes was provided between sessions. The order was pseudo-randomized across subjects. In the first session, participants were instructed to report the

appearance of a fifth letter by clicking a mouse button (the target letter was ‘A’); in the second session, they were instructed to click whenever the fixation point turned from black to white for 300ms, concurrently with the onset of a new trial (i.e. the onset of a new letter). No target letter (i.e. ‘A’) was displayed during this version of the task. In each block, only 5 targets (either the ‘A’ or the change in the fixation point color) were displayed during the experiment such that the target events occurred each at a random point within successive epochs lasting one fifth of the total task duration. In the third and fourth experiments, participants performed respectively 6 and 3 blocks of the letter detection task, in which they were instructed to detect the letter ‘A’, similarly to experiment 2. In the third experiment, a short pause of a few minutes was granted between the third and fourth block. In experiments 2, 3 and 4, participants were not informed of the presence of the rules and, during the final debriefing at the end of the experiment, none of them was able to report the nature of the rules. In the third experiment, after the end of the last block, participants were informed about the presence of the probabilistic rules, and were asked to perform one familiarity and one generative task. The familiarity task consisted in providing a rate of familiarity from 1 (not familiar) to 10 (very familiar) to all the possible letter transitions. During the generative task, participants were instructed to guess which letter would be the most likely to follow an initial 2-letter transition. In both experiments, every transition was presented 7 times in random order.

Pupillometry

The pupil diameter was acquired with an Eyelink 1000+ eye tracker video-based system (SR Research Ltd., Kanata, Ontario, Canada), recording monocularly pupil size (in arbitrary units) and eye movements with a sampling frequency of 500 Hz.

Before analyzing the data, pupil traces were preprocessed to remove eye-blinks, which were identified by the blink detection algorithm implemented in the Eyelink system, and replaced by linear interpolations. Furthermore, traces were down-sampled to 10Hz, to facilitate the model fitting (see below). The pupil response was modeled according to an autoregressive model with exogenous input (ARX), as described in Zénon, 2017. This approach disentangles the spontaneous low frequency oscillations of the pupil from the responses due to external stimulations. Briefly (see Zénon, 2017 for details), the autoregressive approach can be viewed as being composed of two parts. The auto-regressive part accounts for the variance explained by previous values (i.e. autocorrelation within the signal). The second part accounts for the exogenous inputs, which models external factors. In our study, the design matrix included the onset of the letter, the occurrence of the rare transition and the onset of the target when present. The influence of each factor in the model is regulated by its order: the higher the order, the higher the number of samples used to predict the next sample. These orders were fitted independently for each subject. The output of the model is an impulse response for each considered factor and the innovation error, which is the difference between the actual pupil data and the prediction of the model. Once the model was fitted on each participant's data, we analyzed the impulse responses to the factor modeling the onset of the rare letter. Specifically, we extracted the signed absolute maximum value (which could be either positive or negative, depending on whether the pupil is shrinking or dilating) and its latency. The data were finally analyzed by means of Generalized Linear Mixed Models (GLMM), which were implemented in the SAS 9.3 Software (SAS Institute, Cary NC).

Electrophysiology

In experiment 4, we recorded EEG signals from 32 actively shielded Ag-AgCl electrodes, mounted in an elastic cap in accordance with the extended 10-20 systems (Waveguard32 cap; Advanced Neuro Technologies). The frontal AFz electrode was used as ground, and common reference average. All the impedances were kept below 5kΩ. Signals were pre-processed applying a zero-phase Butterworth bandpass filter (0.5 to 30 Hz) regarding the ERPs analysis, whereas no filtering was applied in the time-frequency analysis. Epochs were aligned on stimulus display, starting 200ms prior to onset and lasting until the onset of the next stimulus (i.e. the subsequent letter). The signal was baseline-corrected (200 ms pre-stimulus baseline) and an independent component analysis was used to remove eye movements after visual inspection, and before rejecting epochs whose amplitude was higher than 50 μV.

The analyses were performed in MATLAB (The Mathworks, MA) using Letswave6 toolbox (http:/www.nocions.org/letswave; see also (Mouraux & Iannetti, 2008)). In order to analyze the data we gathered the electrodes in 3 clusters based on regions of interest (Abla & Okanoya, 2009): we considered frontal left (F7, FC5, F3 and FC1), frontal right (F8, FC6, F4 and FC2) and central occipital regions (POz, Oz).

Regarding the frontal regions, we considered a time-window between 50 and 200ms post stimulus onset to include N100 and P200 peaks, whereas for the posterior region we considered a time window between 250 and 550ms post-stimulus, to include the P3 component. These regions and temporal windows have been already used in previous studies investigating the electrophysiological correlates of statistical learning (Abla & Okanoya, 2009; Karuza et al., 2013; Koelsch, 2016). We tested a GLMM on each temporal window considering as dependent variable the signed area of the event-

related potential in each participant, and as factors REGION (frontal left and frontal right for the first time window, and POz and Oz for the second time window) and CONDITION, defining whether the trace was elicited by the rare or by the regular transition. SUBJECT was considered as a random effect. All the interactions between the factors were included in the fixed and random parts of the models.

Regarding the time-frequency analysis, we computed short-time Fourier transform spectrogram aligned at the onset of each letters. We used a Hamming time-window of 50ms with an overlap of 25ms from -200ms to 800ms, whereas the frequency ranged from 0Hz to 256Hz, with a resolution of 0.013Hz. We tested a GLMM, extracting for each participant the average power amplitude in the 50ms – 200ms time window, separately for the frequent and the rare transition trials (factor CONDITION). The factor BAND modeled the 4 physiological bands: $0Hz < \theta < 8Hz$, $8< \alpha < 12Hz$, $12< \beta < 30Hz$ and $30< \gamma < 100Hz$. In the GLMM we considered SUBJECT in the random model, as well as the full interactions of the other factors.

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences : An ERP study, *64*, 185–190. http://doi.org/10.1016/j.neures.2009.02.013

Alamia, A., Solopchuk, O., D'Ausilio, A., Van Bever, V., Fadiga, L., Olivier, E., & Zénon, A. (2016). Disruption of Broca's Area Alters Higher-order Chunking Processing during Perceptual Sequence Learning. *Journal of Cognitive Neuroscience*, *28*(3), 402–417. http://doi.org/10.1162/jocn_a_00911

Alamia, A., & Zénon, A. (2016). Statistical Regularities Attract Attention when Task-Relevant. *Frontiers in Human Neuroscience*, *10*(February), 1–10. http://doi.org/10.3389/fnhum.2016.00042

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The

Need for Longitudinal Data. *Frontiers in Psychology*, *3*(August), 324. http://doi.org/10.3389/fpsyg.2012.00324
Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–50. http://doi.org/10.1146/annurev.neuro.28.061604.135709
Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations : Predictable items are enhanced even when not predicted. *COGNITION*, *129*(2), 205–211. http://doi.org/10.1016/j.cognition.2013.07.003
Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*(12), 3046–3052. http://doi.org/10.1152/jn.00502.2014
Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. http://doi.org/10.1037/0033-295X.108.3.624
Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. http://doi.org/10.1111/j.1469-8986.2008.00654.x
Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952
Carter, C. S. (1998). Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance. *Science*, *280*(5364), 747–749. http://doi.org/10.1126/science.280.5364.747
Cohen, J., & Aston-Jones, G. (2005). Decision Amid Uncertainty. *Advanced Materials*, *17*(16), 1977–1981. http://doi.org/10.1002/adma.200401726
de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(5), E618-25. http://doi.org/10.1073/pnas.1317557111
De Groot, S. G., & Gebhard, J. W. (1952). Pupil size as determined by adapting luminance. *Journal of the Optical Society of America*, *42*(7), 492–495. http://doi.org/10.1364/JOSA.42.000492
Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, *16*(8), 1146–53. http://doi.org/10.1038/nn.3428
Foote, S. L., Berridge, C. W., Adams, L. M., & Pineda, J. A. (1991). Electrophysiological evidence for the involvement of the locus coeruleus in alerting, orienting, and attending. *Progress in Brain Research*, *88*, 521–532.
Gabay, S., Pertzov, Y., & Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception &*

Psychophysics, *73*(1), 123–9. http://doi.org/10.3758/s13414-010-0015-4

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*. http://doi.org/10.1016/j.clinph.2008.11.029

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 252–69. http://doi.org/10.3758/CABN.10.2.252

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science (New York, N.Y.)*, *132*(3423), 349–50. http://doi.org/10.1126/science.132.3423.349

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science (New York, N.Y.)*, *143*(3611), 1190–1192. http://doi.org/10.1126/science.143.3611.1190

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. http://doi.org/10.1037//0033-295X.109.4.679

Janacsek, K., Ambrus, G. G., Paulus, W., Antal, A., & Nemeth, D. (2015). Right hemisphere advantage in statistical learning: Evidence from a probabilistic sequence learning task. *Brain Stimulation*, *8*(2), 277–282. http://doi.org/10.1016/j.brs.2014.11.008

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. http://doi.org/10.1016/j.neuron.2015.11.028

Kang, O. E., Huffer, K. E., & Wheatley, T. P. (2014). Pupil dilation dynamics track attention to high-level information. *PloS One*, *9*(8), e102463. http://doi.org/10.1371/journal.pone.0102463

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54. http://doi.org/10.1016/j.bandl.2012.11.007

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Neuroscience Letters Testing assumptions of statistical learning : Is it long-term and implicit ?, *461*, 145–149. http://doi.org/10.1016/j.neulet.2009.06.030

Kimura, M. (2012). Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *International Journal of Psychophysiology*. http://doi.org/10.1016/j.ijpsycho.2011.11.010

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068–

1078. http://doi.org/10.1111/ejn.12859
Koelsch, S. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports*, (August 2015), 1–11. http://doi.org/10.1038/srep11677
Lavín, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*(January), 218. http://doi.org/10.3389/fnbeh.2013.00218
Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, *26*(7), 1041–1054. http://doi.org/10.1016/j.mri.2008.01.011
Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11. http://doi.org/10.1167/13.2.11.doi
Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. http://doi.org/10.1038/nn.3130
Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, a. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. http://doi.org/10.1111/1469-8986.3850752
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca' s Area and Language Processing: Evidence for the Cognitive Control Connection. *Language and Linguistics Compass*, *4*, 906–924. http://doi.org/10.1111/j.1749-818X.2010.00244.x
O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), E3660--9. http://doi.org/10.1073/pnas.1305373110
Orr, C., & Hester, R. (2012). Error-related anterior cingulate cortex activity and the prediction of conscious error awareness. *Frontiers in Human Neuroscience*, *6*(June), 1–12. http://doi.org/10.3389/fnhum.2012.00177
Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1–2), 185–198. http://doi.org/10.1016/S1071-5819(03)00017-X
Pekkonen, E., Rinne, T., & Näätänen, R. (1995). Variability and replicability

of the mismatch negativity. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *96*(6), 546–554. http://doi.org/10.1016/0013-4694(95)00148-R

Preuschoff, K., 't Hart, B. M., & Einh??user, W. (2011a). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*(SEP), 1–12. http://doi.org/10.3389/fnins.2011.00115

Preuschoff, K., 't Hart, B. M., & Einh??user, W. (2011b). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, (SEP), 1–12. http://doi.org/10.3389/fnins.2011.00115

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science (New York, N.Y.)*, *306*(5695), 443–447. http://doi.org/10.1126/science.1100301

Roser, M. E., Fiser, J., Aslin, R. N., & Gazzaniga, M. S. (2011). Right hemisphere dominance in visual statistical learning. *Journal of Cognitive Neuroscience*, *23*(5), 1088–1099. http://doi.org/10.1162/jocn.2010.21508

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

Saffran, J. R., & Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. http://doi.org/10.1111/1467-8721.01243

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–227. http://doi.org/10.1016/S0926-6410(02)00194-5

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–3. http://doi.org/10.1038/nn873

Shelley, A. M., Ward, P. B., Catts, S. V., Michie, P. T., Andrews, S., & McConaghy, N. (1991). Mismatch negativity: An index of a preattentive processing deficit in schizophrenia. *Biological Psychiatry*, *30*(10), 1059–1062. http://doi.org/10.1016/0006-3223(91)90126-7

Siagian, C., & Itti, L. (2017). Computational and Cognitive Neuroscience of Vision, 141–169. http://doi.org/10.1007/978-981-10-0213-7

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, *6*(3). http://doi.org/10.1371/journal.pone.0018298

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, *8*(September), 666. http://doi.org/10.3389/fnhum.2014.00666
Tettamanti, M., & Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex*, *42*, 491–494.
Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition*, *31*(7), 1060–1071. http://doi.org/10.3758/BF03196127
Turk-browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional Visual Statistical Learning, *34*(2), 399–407. http://doi.org/10.1037/0278-7393.34.2.399
Turk-browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning, *134*(4), 552–564. http://doi.org/10.1037/0096-3445.134.4.552
Turk-browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2010). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness, *21*(10), 1934–1945. http://doi.org/10.1162/jocn.2009.21131.Neural
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), 1934–45. http://doi.org/10.1162/jocn.2009.21131
Van Gerven, P. W. M., Paas, F., Van Merri??nboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167–174. http://doi.org/10.1111/j.1469-8986.2003.00148.x
Vankov, A., Herve-Minvielle, A., & Sara, S. J. (1995). Response to novelty and its rapid habituation in locus coeruleus neurons of the freely exploring rat. *European Journal of Neuroscience*, *7*(6), 1180–1187. http://doi.org/10.1111/j.1460-9568.1995.tb01108.x
Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6*(April), 1–16. http://doi.org/10.3389/fnhum.2012.00088
Yellin, D., Berkovich-Ohana, A., & Malach, R. (2015). Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. *NeuroImage*, *106*, 414–427. http://doi.org/10.1016/j.neuroimage.2014.11.034
Yu, A. J. (2012). Change is in the eye of the beholder. *Nat Neurosci*, *15*(7), 933–935. Retrieved from http://dx.doi.org/10.1038/nn.3150
Zénon, A. (2017). Time-domain analysis for extracting fast-paced pupil responses. *Scientific Reports*, *7*(December 2016), 41484. http://doi.org/10.1038/srep41484
Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate

with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*(August), 286. http://doi.org/10.3389/fnbeh.2014.00286
Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–77. http://doi.org/10.1177/0956797612460407

Chapter VI: General discussion

Overview of the experimental studies

A plethora of diverse studies has investigated several aspects of unconscious processing from different points of view: implicit learning, subliminal learning, priming, and so on. Nevertheless, in 1994 Shanks and St Johns questioned the existence of unconscious decision making (and more generally of all unconscious processes) introducing criteria that put the whole field into question (Newell & Shanks, 2014; Shanks & Stjohn, 1994). Has unconscious learning been proven so far? Do we have unequivocal evidence in favor of the existence of unconscious learning, satisfying the most stringent criteria? This work tries to address this question definitively, providing conclusive evidence in favor of the existence of unconscious processing in a simple perceptual decision making task (chapter II). Following the experimental demonstration of the existence of unconscious processing, this work investigates further two fundamental questions: how unconscious mechanisms affect visual attention (chapter III and IV), and how pupil size can reliably track unconscious processes (chapter V).

CHAPTER	RESULTS
Chapter II	Provide evidence in favor of the existence of unconscious learning.
Chapter III and IV	Investigating the relationship between unconscious learning and eye movements.
Chapter V	Investigating how the eye pupil responds to unconscious surprising events.

Table 1 – Overview of the work

All in all, this work provides an experimental framework for studying perceptual and statistical unconscious learning, and how it affects cognitive processes, paving the way for further studies rooted in a methodologically robust experimental approach.

One of the main difficulties in proving the existence of unconscious learning and at the same time addressing the criteria suggested by Shanks and colleagues (Newell & Shanks, 2014; Shanks & Stjohn, 1994) lies in disentangling the conscious and unconscious information that drives the behavioral effect (i.e. information and sensitivity criteria). One way to address this issue is to design an experiment in which the unconscious knowledge is simple to test, being a straightforward association (i.e. between motion and color) or a simple rule. Notably, the awareness of such association is straightforward to test with the most sensitive approaches (familiarity and generative tests in addition to the questionnaire). Our results, in agreement with the definition of consciousness discussed in the introduction, proved that the learnt association is truly unconscious: 18 out of 23 participants showed a behavioral effect (i.e. learnt the color-motion association) but failed in reporting the contingency during the questionnaire (ability to verbally access the acquired knowledge) and performed at random in the generative and familiarity tasks (ability to recollect the acquired knowledge). Unfortunately we did not test directly the third definition suggested by Destrebecqz and Peigneux (Destrebecqz & Peigneux, 2005) which defines consciousness as acquisition of meta-knowledge: an interesting future study may investigate how confidence is affected during this kind of learning by the different contingencies. All in all, the first main result is therefore the factual existence of unconscious learning, demonstrating that implicit processes may occur in the absence of complex rules. Consequently, an interesting subsequent question to test could be: would it be possible to prove the existence of

implicit learning in the context of a more complex design? In other words: would it be possible to address Shanks's criteria considering the paradigms already explored in the literature? So far the results from the literature seem to be open to interpretation. The most conservative approach would cautiously deny the existence of unconscious learning in many paradigms: in the case of contextual cuing for example, in which a visual search task is influenced by implicit knowledge, an extensive testing of the actual level of awareness of the participants revealed a lack of truly implicit learning (Smyth & Shanks, 2008). An important step for future studies will be to show convincing results regarding the existence of unconscious learning in more complex paradigms, addressing both the theoretical issues (Destrebecqz & Peigneux, 2005) and the methodological ones.

Interestingly, our results can be considered in the light of another computational approach, namely the distinction between model-free and model-based processes (Daw, Niv, & Dayan, 2005; Dayan & Berridge, 2014; Wan Lee, Shimojo, & O'Doherty, 2014). Briefly, the difference between the two methods is that model-based learning consists in updating a model of the world (state transition matrix) which drives the decision of the agent, leading him from one state to another; whereas model-free learning assumes that the agent acts coherently with a process which does not encode the transitions between states, but is based on simple stimulus-response associations. Notably, both are supervised learning (i.e. feedback dependent), but the model-free process is usually associated with a lower amount of cognitive resources (Otto, Raio, Chiang, Phelps, & Daw, 2013) and is usually related to unconscious learning, whereas model-based processes generally require an higher amount of cognitive resources, and are mostly explicit. In our study, in line with these theoretical considerations, it is legitimate to assume that participants implement a model-free type of

learning. It is important to remark though that in motor learning, for example, model-free and model-based do not equate implicit and explicit learning: an example is the case of motor adaptation in which, despite being model-based, the changes in the motor behavior remain implicit (Orban de Xivry & Lefèvre, 2015; Taylor, Krakauer, & Ivry, 2014). However, further studies may provide further insight within this computational framework.

After having proved the actual lack of awareness of the association in our perceptual task, it is possible to investigate how unconscious processing affects other cognitive functions, such as visual attention (chapter III and IV). Here we addressed this question investigating two different paradigms: on the one hand one task was based on the same perceptual decision making task investigated in the first study; whereas on the other hand the second study was based on an implicit statistical learning framework, in which some stimuli were predictable and others were not. It is legitimate to assume that the behavioral effect reported in the first study (i.e. the learning of the association) remained implicit: the questionnaire responses and the similarity of the design with respect to the original study provide robust evidence in favor of this hypothesis, even though we didn't test directly the generative and familiarity tasks in the attention study. Conversely, the truly unconscious nature of the statistical learning task could not be proven rigorously (i.e. addressing Shanks' criteria): our operative definition of consciousness in this case will be limited to the lack of verbal report. Importantly, in both studies we found that visual attention was affected by unconscious knowledge, in such a way that attention was biased toward the most informative stimuli.

In the last experimental study (chapter V), we finally investigated whether it is possible to track unconscious processing using pupillometry measures. The main result of this study is that the surprise effect does not need to be conscious in order to elicit a pupillary response: as attention can

be affected by unconscious information without being consciously processed (previous two studies), in the same fashion physiological responses, such as pupil size, may bypass the access to conscious level and be driven by unconscious mechanisms.

As already mentioned previously (see chapter V), in the framework of predictive coding, attention affects prediction error (PE), increasing the precision of the perceptual inferences. In the light of previous studies investigating pupillary responses to cognitive processes, and considering our results showing that unconscious processes may affect pupil size, it is possible to suggest an alternative hypothesis regarding pupillary response: pupil is modulated by changes in precision-weighted prediction errors. Prediction error can be defined as the difference between the actual information received and the prediction about such information: precision-weighted PE are simply weighted by the reliability (i.e. precision) of the prediction. More specifically, in a very volatile environment, in which predictions are unreliable and uncertain, prediction errors will have less impact than in a stable and certain environment. Moreover, it is possible to distinguish between perceptual, motivational and cognitive prediction errors (Den Ouden, Kok, & de Lange, 2012), depending on the brain regions in which they are computed (e.g. sensory areas produce perceptual prediction errors, etc.). Preliminary results from a pilot study we conducted recently suggest that the pupil may indeed reflect changes in precision-weighted prediction errors: specifically, we manipulated the averaged life time of the dots on the screen (considered as a proxy for prediction error) and observed indeed a significant change in pupil size. Importantly, our hypothesis suggests that pupil would respond differently whether the change in prediction errors occurs from a volatile to a more certain environment (higher to lower PE) or the other way around. Further studies, combining neuroimaging techniques with pupillometry, will reveal which mechanisms

are at the basis of pupillary modulations, whether prediction errors can be related to the pupil size also from a functional and physiological perspective, and what is the role of consciousness in driving such physiological response.

Attention and awareness

The results from the two studies investigating the relationship between overt attention and unconscious learning (chapter III and IV) allow us to examine further two important points: the controversial relationship between attention and awareness, and the specific role of attention in our tasks. Regarding the first issue, some authors have argued that awareness and attention are cognitive processes which share the same mechanisms (Cohen, Cavanagh, Chun, & Nakayama, 2012; Hogarth, Dickinson, Austin, Brown, & Duka, 2008), and it is challenging –if not impossible- to disentangle the two processes. Nevertheless, more and more pieces of evidence are showing rather the opposite trend, disentangling attention and awareness in several tasks (Koch & Tsuchiya, 2007, 2012; Tsuchiya & Koch, 2009). Particularly, Koch and Tsuchiya attributed different functional roles to each process: on one side attention acts as an information filter, thus reducing the information overload that daily overwhelms the brain; on the other side consciousness is supposedly summarizing relevant information, planning actions, accomplishing complex decision making and further processes (such as language, setting long term goals, etc.). Therefore, an effect on attention allocation which escapes the scrutiny of awareness, as shown in our results, is in line with this interpretation. On a similar note, a recent study from Kok and colleagues (Kok, Rahnev, Jehee, Lau, & De Lange, 2012) disentangles the effect of prediction (i.e. expectation) and attention, in the

framework of predictive coding (Kok & De Lange, 2015). Specifically, in an attentional cueing paradigm (such as the Posner task, excellently modeled by Feldman and Friston in predictive coding terms (Feldman & Friston, 2010)), prediction increases overall sensory processing, instead of reducing it (as expected). This effect can be explained considering that attention and prediction synergistically improve the precision of prediction errors (Vossel, Geng, & Friston, 2014). Therefore it is possible to see attention as a mechanism that reduces uncertainty in the sensorium, increasing the precision of the perceptual inference, and therefore sampling preferentially informative stimuli (Mackintosh, 1975).

After accepting that attention and consciousness are two distinct mechanisms, a second issue may be: what is the role of attention, specifically in our tasks? To answer this question, it is possible to hypothesize that unconscious and conscious information are integrated optimally (in Bayesian sense, in which each source of information is weighted as a function of its reliability; see fig.1) in the brain (Angelaki, Gu, & DeAngelis, 2009; Francken, Meijs, Hagoort, van Gaal, & de Lange, 2015; Mudrik, Faivre, & Koch, 2014; Yuille, Bulthoff, Kersten, & Mamassian, 1996).

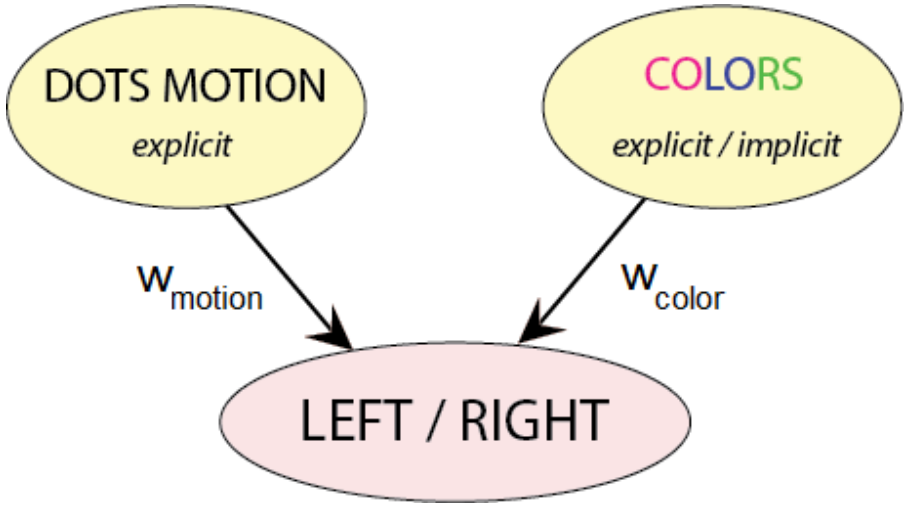


Figure 1 – Integration of explicit and implicit information

Consequently, unconscious knowledge would affect eye movements in an optimal fashion: the more the explicit information is difficult to decode (difficult motion perception), the more weight (i.e. importance) is given to the implicit information, thus affecting eye movements. This interpretation can also shed some light on the actual relative small size of the effect we reported in our data: it is reasonable to assume that a larger effect on eye movements would have been observed with more challenging and difficult tasks. In other words, since both the motion detection task in the first study and the color detection task in the second study were relatively easy (performance around 80/95% in the difficult/easy condition in the first study and close to 100% in the second one) unconscious information was not crucial to perform the task and therefore in driving eye movements.

The role of attention emerges clearly also in the last experimental study: pupil size dilates significantly when attention is focused on the stream of letters (presumably favoring the occurrence of statistical learning) whereas such effect vanishes when attention is diverted elsewhere. Nonetheless, it is also important to remark the conceptually different nature of the attentional mechanisms involved in chapter III and IV, and the ones in the last study (chapter V). In the first case, attention is rather influenced by top-down mechanisms, i.e. related to task-relevant information: participants are actively sampling one out of two possible stimuli (i.e. selective attention), choosing the one that in that condition (i.e. difficult motion perception in chapter III) optimize their performance; conversely, in the second case attention is driven by bottom-up mechanisms, that is the occurrence of a single –unconscious- salient event (i.e. a rare transition) which supposedly elicits an attentional response.

Limitations and future directions

Undoubtedly, several methodological and conceptual limitations in the studies described in the present manuscript should be discussed further. The first point that could be addressed regards the different ways consciousness has been measured across the studies. In the first study, the behavioral effect was measured as a difference in the accuracy (i.e. number of correct responses) between the informative and the non-informative colors. Even though in the second experiment of the first study we reported an effect on RT, the design of the whole experiment was such that participants were forced to provide an answer quickly (within 1 second), thus fostering the occurrence of any effect in terms of accuracy rather than RT. Conversely, in the second and third studies (chapter III and IV respectively) we reported differences between conditions in RT rather than in the accuracy. In this case, since we needed participants to localize the two stimuli and then perform a saccade, we lengthen the response time, thus providing more time to answer. This experimental constrain allowed us to measure eye movements in an ecological situation (participants had the time to move their eyes without time pressure) but moved the measured variable from accuracy (nearly optimal, particularly in the third study) to RT. Finally, in the last study we had no behavioral measure of awareness, but simply pupillometry data. Even though it would have been possible to add a letter-detection task, showing a slower RT for rare transitions, this may have polluted the pupillometry data, adding noise to the requested measure

(response to surprise): each target would have elicited a pupillary response. In conclusion, even though we quantify awareness with different measures, in all the studies (except the third one), we comply with the criteria raised by Shanks and colleagues.

Another important issue may regard the statistical validity of our study, from the point of view of replication and sample size. Even though we replicated from the findings from the first and last study four times (4 slight variations of the same design), confirming each time the effect, we used relatively small sample size (in all experiments $14<N<20$), thus casting a shadow of doubt regarding the reliability of the results. Specifically in unconscious and implicit learning, replicability represents an important and often argued problem (Vadillo, Konstantinidis, & Shanks, 2015) which deserved full consideration: future experiments will be needed to replicate and expand the results with larger sample size ($N>20$).

Lastly, it is possible to argue that the fourth criterion raised by Shank, i.e. the immediacy criterion, is not entirely fulfilled in our studies. As a matter of fact it is possible that participants –during the time between the experiment and the questionnaire- forget the condition learnt during the task (e.g. the color motor associations in the first two studies). Even though it is not possible to discard this case a priori, it seems implausible, given the extreme simplicity of the association, that participants aware of the contingency would forget it. Conversely, considering consciousness as a graded rather than an all-of-nothing phenomenon, this could be possible in the case in which the association would be not fully conscious, thus –at least partially- unconscious.

The role of feedback: dopaminergic pathways and an alternative model

The important role played by dopamine in the occurrence of learning is well-established in the scientific literature (Dayan & Balleine, 2002; Wise, 2004), as well as its role in implicit learning mechanisms (Karabanov et al., 2010). In line with the literature, the third experiment of the first study suggests as well that the learning in such task relies on some sort of reinforcement-learning mechanisms: the lack of feedback impaired the occurrence of learning. This result leads to some possible speculation about the neurotransmitters involved in the learning process itself. It has been shown that dopamine signals reward-prediction error (W. Schultz, 2001; W Schultz, 1998; Wolfram Schultz, 1999) and we could thus speculate that, in our task, it would be responsible for reinforcing the link between color and motion directions, or between the color and the motor response (as highlighted in the fourth experiment, both associations are learnt: see (Mostert, Kok, & de Lange, 2015) for a review about the dissociation between sensory and motor decision in a perceptual task). Moreover, a recent review about the pharmacology of implicit learning advocates a similar role for the dopaminergic systems (Uddén, Folia, & Petersson, 2010). An alternative experimental model that would address this question could involve patients affected by Parkinson disease, in which a lack of dopamine is observed in the midbrain (specifically in the substantia nigra). By comparing the performance of patient with and without treatment (L-dopa, or with dopamine agonist) with healthy subjects, it would be possible to investigate whether

dopamine plays a crucial role in this perceptual learning task. The effect of dopamine has been already studied on other implicit learning tasks considering as model patients affected by Parkinson disease (PD): the results revealed that motor sequence learning tasks are affected by degeneration of basal ganglia in the pathological populations (Hayes, Hunsaker, & Dibble, 2015), but the results are less clear in other paradigms in which the motor component does not play an important role. A study considering the contextual cueing task reported no differences between healthy participants and PD patients (van Asselen et al., 2009), whereas a verbal version of the serial reaction time task reported a significant difference between the two groups (Westwater, McDowall, Siegert, Mossman, & Abernethy, 1998). Our results seem to show that feedback is needed for the learning of the association to occur. An alternative explanation could be that the presence of the feedback engages more the participants in processing the features displayed on the screen: this would favor learning without altering the overall performance (similar between the group with and without feedback). From a computational point a view, it would be possible to test the data from our study with the prediction provided by the Rescorla-Wagner model, in which a conditioned stimulus (i.e. the color) predicts the unconditioned stimulus (i.e. motion direction). This popular and influential model, proposed for the first time in 1972 (Rescorla & Wagner, 1972) has been updated and modified through the years, and notably a study related the phasic activity of dopaminergic neurons with the prediction error encoded in this model (Hazy, Frank, & O'Reilly, 2010; P. Read Montague, Hyman, & Cohen, 2004; P R Montague, Dayan, & Sejnowski, 1996). This computational approach would further validate and test the role of the feedback in our task.

Models of consciousness and our findings

In the introduction, several models have been presented to describe the mechanisms of consciousness and conscious cognition. In this last paragraph of the discussion our results will be interpreted in the light of these models.

Regarding the Global Workspace theory, in which smaller modules process information locally (and unconsciously) and transmit the relevant bits to a global workspace, it is possible to imagine that the color-motion association we observed in our first study does not reach a conscious level because it remains processed in specific, unconscious, visual modules of the brain, and doesn't upgrade to the global workspace spontaneously. Similarly, it is legitimate to assume that the pupillary response is triggered by information encoded at lower levels, which does not reach the global workspace. This framework raises at least two theoretical questions. Firstly why some pieces of information would be constrained to lower levels whereas others would reach the Global Workspace: in other words, which features are required in order to access the higher levels and becoming conscious? Secondly, why in few subjects the same information leaked into the global workspace leading them to become conscious of the association, whereas for other participants it remained confined to the lower modules, and stayed unconscious? What's the role of attention is such framework (assuming that attention and awareness are different mechanisms, see previous paragraph)? Possibly, some participants were more prone to looking for patterns, sampling information and generating hypothesis consciously. Another possible explanation could be that some participants payed more

attention to the relevant feature (i.e. the colors) than others, thus increasing the likelihood to notice consciously the association. Moreover, it is possible that a specific series of trials (i.e. a series of colors and motion directions) eventually led to a salient event: twice the same association would boost the saliency of the next trial having the opposite association (e.g. blue/left; blue/left; red/right): such bottom-up effect may drive the information to a conscious level. In other words, in the Global Workspace model, the role of attention can be seen as a modulator of the probability for the information processed in each specific module to enter in the global workspace. On a similar note lays the Higher Order Thought theory, based on the presence of different, hierarchical levels. The assumption would be that the lower levels (indeed reminiscent of the specific modules in the global workspace theory) learn the contingencies in the environment (e.g. the association between colors and motion directions), thus affecting the behavior of the participants. Nevertheless, the higher levels fail in learning that the lower levels are using such information (i.e. have learnt such information). Similarly, the pupillary response may be triggered by the statistical learning which occurs at the lowest levels of the hierarchy, escaping the monitoring of the higher levels.
Also considering an agent conscious when his own internal states are represented in an internal model leads to similar interpretations. De facto, it is possible to reason that the color-motion association, even though not encoded in the internal model of the agent, may still affect behavior from the lower levels. Alternatively, even though encoded in the hypothetical model of the agent, it doesn't reach consciousness because submerged by other information considered as more salient.

All in all, the ultimate computational model of consciousness needs to take into account (and should be able to model properly) the experimental results showed in this work and in the current literature. A trustable and worth

model of consciousness needs to be able to explain non-conscious behavior and unconsciousness as well.

References

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*. http://doi.org/10.1016/j.conb.2009.06.008

Cohen, M. a, Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411–7. http://doi.org/10.1016/j.tics.2012.06.013

Daw, N. D., Niv, Y., & Dayan, P. (2005). Actions, policies, values and the basal ganglia. *Recent Breakthroughs in Basal Ganglia Research*, (February), 91–106. http://doi.org/10.1.1.103.4593

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*. http://doi.org/10.1016/S0896-6273(02)00963-7

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 473–492. http://doi.org/10.3758/s13415-014-0277-8

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*(DEC), 1–12. http://doi.org/10.3389/fpsyg.2012.00548

Destrebecqz, A., & Peigneux, P. (2005). Methods for studying unconscious learning. *Progress in Brain Research*. http://doi.org/10.1016/S0079-6123(05)50006-2

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*(December), 215. http://doi.org/10.3389/fnhum.2010.00215

Francken, J. C., Meijs, E. L., Hagoort, P., van Gaal, S., & de Lange, F. P. (2015). Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific Reports*, *5*, 17725. http://doi.org/10.1038/srep17725

Hayes, H. A., Hunsaker, N., & Dibble, L. E. (2015). Implicit motor sequence learning in individuals with Parkinson disease: A meta-analysis. *Journal of Parkinson's Disease*, *5*(3), 549–560. http://doi.org/10.3233/JPD-140441

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*. http://doi.org/10.1016/j.neubiorev.2009.11.019

Hogarth, L., Dickinson, A., Austin, A., Brown, C., & Duka, T. (2008). Attention and expectation in human predictive learning: the role of uncertainty. *Quarterly Journal of Experimental Psychology (2006)*, *61*(11), 1658–68. http://doi.org/10.1080/17470210701643439

Karabanov, A., Cervenka, S., de Manzano, Ö., Forssberg, H., Farde, L., & Ullén, F. (2010). Dopamine D2 receptor density in the limbic striatum is related to implicit but not explicit movement sequence learning. *Proceedings of the National Academy of Sciences*, *107*(16), 7574–7579. http://doi.org/10.1073/pnas.0911805107

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. http://doi.org/10.1016/j.tics.2006.10.012

Koch, C., & Tsuchiya, N. (2012). Attention and consciousness: Related yet different. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2011.11.012

Kok, P., & De Lange, F. P. (2015). *Predictive coding in sensory cortex*. *An Introduction to Model-Based Cognitive Neuroscience*. http://doi.org/10.1007/978-1-4939-2236-9_11

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*(9), 2197–2206. http://doi.org/10.1093/cercor/bhr310

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*. http://doi.org/10.1037/h0076778

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947. http://doi.org/10.1.1.156.635

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767. http://doi.org/10.1038/nature03015

Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*(November), 18253. http://doi.org/10.1038/srep18253

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496. http://doi.org/10.1016/j.tics.2014.04.009

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behavioral and Brain Sciences*, *37*, 1–19. http://doi.org/10.1017/S0140525X12003214

Orban de Xivry, J.-J., & Lefèvre, P. (2015). Formation of model-free motor memories during motor adaptation depends on perturbation schedule.

Journal of Neurophysiology, *113*(7), 2733–2741. http://doi.org/10.1152/jn.00673.2014

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. http://doi.org/10.1073/pnas.1312011110

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II Current Research and Theory*, *21*(6), 64–99. http://doi.org/10.1101/gr.110528.110

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. http://doi.org/10.1007/s00429-010-0262-0

Schultz, W. (1999). The Reward Signal of Midbrain Dopamine Neurons. *News in Physiological Sciences*, *14*(December), 249–255. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11390860

Schultz, W. (2001). Book Review: Reward Signaling by Dopamine Neurons. *The Neuroscientist*, *7*(4), 293–302. http://doi.org/10.1177/107385840100700406

Shanks, D., & Stjohn, M. (1994). Characteristics of Dissociable Human Learning-Systems, (February 2010). http://doi.org/10.1017/S0140525X00035032

Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403–415. http://doi.org/10.3758/MC.36.2.403

Taylor, J. a, Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(8), 3023–32. http://doi.org/10.1523/JNEUROSCI.3619-13.2014

Tsuchiya, N., & Koch, C. (2009). The relationship between consciousness and attention. In *The Neurology of Consciousness* (pp. 63–77). http://doi.org/10.1016/B978-0-12-374168-4.00006-X

Uddén, J., Folia, V., & Petersson, K. M. (2010). The neuropharmacology of implicit learning. *Current Neuropharmacology*, *8*, 367–381. http://doi.org/10.2174/157015910793358178

Vadillo, M. a., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-015-0892-6

van Asselen, M., Almeida, I., Andre, R., Januário, C., Gonçalves, A. F., & Castelo-Branco, M. (2009). The role of the basal ganglia in implicit contextual learning: A study of Parkinson's disease. *Neuropsychologia*, *47*(5), 1269–1273. http://doi.org/10.1016/j.neuropsychologia.2009.01.008

Vossel, S., Geng, J. J., & Friston, K. J. (2014). Attention, predictions and expectations, and their violation: attentional control in the human brain. *Frontiers in Human Neuroscience*, *8*(July), 3389. http://doi.org/10.3389/fnhum.2014.00490

Wan Lee, S., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. http://doi.org/10.1016/j.neuron.2013.11.028

Westwater, H., McDowall, J., Siegert, R., Mossman, S., & Abernethy, D. (1998). Implicit learning in Parkinson's disease: evidence from a verbal version of the serial reaction time task. *Journal of Clinical and Experimental Neuropsychology*, *20*(June 2014), 413–418. http://doi.org/10.1076/jcen.20.3.413.826

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews. Neuroscience*, *5*, 483–494. http://doi.org/10.1038/nrn1406

Yuille, A. L., Bulthoff, H. H., Kersten, D., & Mamassian, P. (1996). Perception as Bayesian Inference. *Annual Review of Psychology*, *55*, 271–304. http://doi.org/10.1146/psych.2004.55.issue-1

Acknowledgements

Finally, and fortunately, it took 5 years to complete my phd. It has been the hell of a long ride, the most intense period of my life, in which I grew a lot. The list of persons I'd like to thank is very long: in this 5 years I had the good chance to meet amazing and important people. Here, in a very rough chronological order, I want to take my time to thank them all.

Firstly my supervisors. Etienne has been a mentor, a guide, a reference point. His death was one of the saddest moment of my conscious life, the first time I understood how devastating it is to miss suddenly a person you love. We had created a positive loop, a reinforcement feedback: his incredible trust in my capacity and his appreciation for my hard working and motivation was bursting even more my will to work hard and be motivated in what I was doing. I wish I had the chance to see the pride in his eyes at the end of this path we started together.

Alex has been one of the most important person here in Brussels for me, and I can't thank the Gods enough for having put him in my life. Scientifically he thought me everything. As a human being, believe it or not, even more. He keeps showing me how to be amazing without losing a tiny bit of his immense humbleness. During these years I managed to extensively test his boundless patience, to admire his intelligence and vast erudition at work, and enjoy his very good humor and sarcasm. In 5 years I can't remember once in which he was upset for something. Really. If I managed to do anything good in all these years, most of the honors go to him.

A grateful thanks to my real first Belgian friend, the inspiring Pierre-Alexandre, who not only introduced me to the Belgian world, and avoid that

loneliness invaded me the first months in Brussels, but also showed me how inspiration and bravery can move things and change the world.

Four great postdocs saw me during these years ranging from euphoric happiness to the darkest sadness. Two friends enjoyed my moves on different dance floors (from Brussels to Krakow and back..), and tasted together the best meals in the greatest (phd-student affordable) restaurants in Bruxelles (turjleeeek!). Thanks to my dearest friends Diana and Emanuel. And two other amazing friends with whom I joined the Nespresso coffee gang in the last years, in the one-machine bar at the Roman Offices of UCL: a huge thank to my dearest friends Giulia and Emanuele. The first one a beautiful real Roman flower in Brussels, the second one my squash teacher and my EEG mentor. Two friends I'm happy I can trust as a beautiful part of my life.

I had the very good luck to have amazing office-mates, who managed to bear with my craziness for all these years without killing me. I would like to thank Vincent, my computational friend and amazing dad (sweet kiss to Livia and Cate!), for all the constructive interactions we had, for our shared craziness and for being able to become red 'on command' (big hug!); and a special and deep thank to the tenacious, beautiful and Hungarian (yes, Budapest is -one of- the best city in the world!) Monika, with whom I shared incredible moments during these years, side by side, in bad and in good times.

I specially thank my friend Oleg, with whom I spent most of my time in the last years, inside and outside the lab. From teaching me trivial things, like how to make the presentation smoother and clearer, to show me how to build a tent in the dark, no matter if you are in the middle of Norway, in the Alps or simply in Oostende. But mostly for showing me how is possible to

stay positive and optimistic ever. Even on top of the highest mountain of the world, when just a leather jacket is protecting you against hibernation, at 4000m. Or in the best restaurant of Batumi, in Georgia, or around the corners of Singapore, between a Mosque and a Buddhist temple. Or in Nepal, Ukraine, Malaysia.. Or while training some breakdance and acro move. Thanks for having spent nights in the lab with me, improving my (very poor) Russian between pizza slices, while discussing about pretty much everything. And a special thanks to his lovely family, for their warm and sincere hospitality: part of the review of this work happened in a beautiful but remote village in Ukraine.

A huge thank to André, for adopting me a bit in their group (not only at their lab meetings), and for being an example of very good scientist and group leader, and for carrying on the legacy of Etienne in supervising me. A special thank to Valery, for his good humor, his Facebook posts, and for letting me work with the amazing Lieve -we did a great job with the Bayesian TOJ, didn't we? And a grateful thanks to prof. Axel Cleeremans, for helping me framing the context in the first study and sharing his enthusiasm about implicit learning.

A deep thank to all the people in the lab who shared their ideas and their smile in our corridor meetings: Julie and all his group (Gerard, Charles-Etienne, Caroline, Julien, Emmanuelle, Charlotte), as well as all the members of the NOCIONS team, past and present (too many to list them here). Leila, the impeccable secretary (what else?). Laurence, not only for her stunning beauty but also for her kindness and humanity; and Ben, for his extremely contagious laugh who fills the corridors with good mood and healthy craziness. You all played an important role in my life.

Finally, a special and dutiful thanks to all the members of the committee. Specially to prof. Floris de Lange, for having accepted to be the external member: your expertise and knowledge is very much appreciated. To Prof. Jean Jacques Orban de Xivry (alias JJ) for his help during these years, and particularly in the first study. Prof. Wim Vanduffel for the punctual and precious interaction in these years: even though my work finally didn't investigate Superior Colliculi, I appreciated his valuable comments and interactions at our annual meetings. Finally I'd like to thank prof. Marcus Missal for having accepted to be the president of the committee on-the-flight. I'm glad I have the chance to work with you all.