Dynamic interactions between top-down expectations and conscious awareness

Erik L. Meijs\textsuperscript{a,b}, Heleen A. Slagter\textsuperscript{c,d}, Floris P. de Lange\textsuperscript{b} and Simon van Gaal\textsuperscript{b,c,d}

\textsuperscript{a}Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands
\textsuperscript{b}Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands
\textsuperscript{c}University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands
\textsuperscript{d}University of Amsterdam, Amsterdam Brain and Cognition (ABC), 1001 NK Amsterdam, the Netherlands

DOI: 10.1523/JNEUROSCI.1952-17.2017
Received: 10 July 2017
Revised: 9 November 2017
Accepted: 26 November 2017
Published: 31 January 2018


Conflict of Interest: The authors declare no competing financial interests.

This work was supported by the Netherlands Organization for Scientific Research (NWO VENI (451-11-007) awarded to SvG; NWO VIDI (452-13-016) awarded to FdL), the European Research Council (ERC-2015-STG_679399 awarded to HAS) and the James S McDonnell Foundation (Understanding Human Cognition, 220020373, awarded to FdL). We thank Doris Dijksterhuis, Sjoerd Manger and Thomas Dolman for their valuable assistance with data acquisition. We thank Timo Stein and Josipa Alilovic for valuable comments on a previous draft of this manuscript.

Corresponding author: Simon van Gaal, University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands, simonvangaal@gmail.com

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.1952-17.2017
Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.
Dynamic interactions between top-down expectations and conscious awareness

Running title: Interactions between expectations and awareness

Erik L. Meijs\textsuperscript{a,b}, Heleen A. Slagter\textsuperscript{c,d}, Floris P. de Lange\textsuperscript{b}, & Simon van Gaal\textsuperscript{b,c,d}

\textsuperscript{a} Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands
\textsuperscript{b} Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands
\textsuperscript{c} University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands
\textsuperscript{d} University of Amsterdam, Amsterdam Brain and Cognition (ABC), 1001 NK Amsterdam, the Netherlands

Corresponding author: Simon van Gaal, University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands, simonvangaal@gmail.com

Number of pages: 29. Number of figures: 4. Number of words: Abstract:152; Introduction:650; Discussion:1476

The authors declare no competing financial interests.

This work was supported by the Netherlands Organization for Scientific Research (NWO VENI (451-11-007) awarded to SvG; NWO VIDI (452-13-016) awarded to FdL), the European Research Council (ERC-2015-STG\textsubscript{679399} awarded to HAS) and the James S McDonnell Foundation (Understanding Human Cognition, 220020373, awarded to FdL). We thank Doris Dijksterhuis, Sjoerd Manger and Thomas Dolman for their valuable assistance with data acquisition. We thank Timo Stein and Josipa Alilovic for valuable comments on a previous draft of this manuscript.
Abstract

It is well known that top-down expectations affect perceptual processes. Yet, remarkably little is known about the relationship between expectations and conscious awareness. We address three crucial questions that are outstanding: 1) How do expectations affect the likelihood of conscious stimulus perception?; 2) Does the brain register violations of expectations nonconsciously?; and 3) Do expectations need to be conscious to influence perceptual decisions? We performed three experiments, in human participants, in which we manipulated stimulus predictability within the attentional blink paradigm, while combining visual psychophysics with electrophysiological recordings. We found that valid stimulus expectations increase the likelihood of conscious access of stimuli. Furthermore, our findings suggest a clear dissociation in the interaction between expectations and consciousness: conscious awareness seems crucial for the implementation of top-down expectations, but not for the generation of bottom-up stimulus-evoked prediction errors. These results constrain and update influential theories about the role of consciousness in the predictive brain.

Significance statement

While the relationship between expectations and conscious awareness plays a major role in many prediction-based theories of brain functioning, thus far few empirical studies have examined this relationship. Here, we address this gap in knowledge in a set of three experiments. Our results suggest that the effect of expectations on conscious awareness varies between different steps of the hierarchy of predictive processing. While the active use of top-down expectations for perceptual decisions requires conscious awareness, prediction errors can be triggered outside of conscious awareness. These results constrain and update influential theories about the role of consciousness in the predictive brain.
A rapidly growing body of work indicates that sensory processing is strongly influenced by expectations that we have about likely states of the world. Such expectations are shaped by the context in which we are operating, but also by learning, past experience and our genetic makeup (Friston, 2005; Bar, 2009; Summerfield and de Lange, 2014). Expectations are typically thought to originate from higher-level brain regions, such as the (pre)frontal cortex, which may guide information processing in lower-level sensory regions via top-down projections. In this framework, what we consciously see is proposed to be strongly influenced by the brain’s expectations about, or its best guess of, the outside world (Gregory, 1980; Hohwy, 2012; Panichello et al., 2013). Initial studies support the idea that the brain uses information in the environment to build expectations of stimulus frequency or conditional probabilities to modify perceptual processing (Bar, 2004; Kok et al., 2012). These ideas have been formalized in theoretical models, such as predictive coding and sequential sampling models (Friston, 2005; Ratcliff and McKoon, 2008; Clark, 2013). Although these frameworks are attractive in their simplicity, how exactly expectations shape conscious perception, and to what extent awareness guides the formation of expectations, is still largely unknown.

At present, there are (at least) three issues that need to be resolved to further our understanding of the relationship between expectations and consciousness. The first issue relates to the effect that expectations may have on conscious awareness itself. It has been shown that valid expectations increase the speed of conscious access (Melloni et al., 2011; Pinto et al., 2015; Stein and Peelen, 2015; De Loof et al., 2016) and may help selecting or facilitating stimulus interpretation when (visual) input is ambiguous or noisy (Bar et al., 2006; Denison et al., 2011; Panichello et al., 2013; Chang et al., 2015; Aru et al., 2016). It is yet an open question whether expectations can boost an otherwise unseen stimulus into conscious awareness, thereby enabling the switch from a nonconscious to a conscious stimulus representation, instead of merely facilitating its cognitive interpretation or its speed of appearance in time.

Second, it is an open question to what extent prediction errors, arising in a situation of invalid expectations, can be registered outside of conscious awareness. It has been shown that “oddball” stimuli (e.g. simple violations in auditory tone sequences) elicit early mismatch responses in electrophysiological signals: the mismatch negativity (MMN) (Pöppel, 2004; Näätänen et al., 2007). Interestingly, MMN’s can even be observed
when attention is distracted from the tone sequences (Bekinschtein et al., 2009) or in several reduced states of consciousness, such as sleep (Ruby et al., 2008), anesthesia (Koelsch et al., 2006) and vegetative state (Bekinschtein et al., 2009). This suggests that the MMN reflects a pre-attentive nonconscious prediction error signal (Näätänen et al., 2001; Stefanics et al., 2011; Kimura and Takeda, 2015). However, it remains debated whether these signals originate in model-based comparisons of expectations to new input or merely reflect passive low-level sensory adaptation to repeated inputs (Garrido et al., 2009; Stefanics et al., 2016). The one study in which these mechanisms were dissociated in a nonconscious state showed adaptation remains operative during sleep, whereas prediction error detection disappears (Strauss et al., 2015), questioning the notion that prediction errors may be registered nonconsciously.

The final issue concerns the role of awareness in implementing expectations. Many expectation-based models assume that expectations are implemented via top-down neural activation. Interestingly, influential theories of consciousness suggest that conscious access requires similar top-down interactions between higher-level (e.g. prefrontal) and lower-level (e.g. visual) brain regions, referred to as feedback or recurrent processing (Lamme and Roelfsema, 2000; Dehaene et al., 2006). Information that does not reach conscious access is thought to only trigger feedforward activity or local recurrent interactions between posterior brain regions. Therefore, it is unclear how nonconscious information, in the absence of feedback signals from higher-order cortical areas, could lead to the implementation of expectations.

**Materials & Methods**

**Participants**

We tested 26 participants in Experiment 1 (21 females, age 19.5±1.3 years), 85 participants in Experiment 2 (63 females, age 22.0±3.2 years) and 34 participants in Experiment 3 (27 females, age 20.0±1.1 years). All participants were right-handed and had normal or corrected-to-normal vision. For all experiments, participants for whom the minimum number of observations in one or more conditions was lower than 10, were excluded from analysis. Additionally, for Experiment 3 (EEG), we had to exclude 2 participants due to problems with the reference electrodes. In the end, this resulted in the inclusion of 25 participants.
participants for Experiment 1 (20 females, age 19.5±1.3 years), 67 participants for Experiment 2 (49 females, age 21.9±3.0 years) and 29 participants for Experiment 3 (22 females, age 20.0±1.1 years).

The studies were approved by the local ethics committee of the University of Amsterdam and written informed consent was obtained from all participants according to the Declaration of Helsinki. Compensation was 20 Euros for Experiment 1, 25 euros for Experiment 2 and 30 Euros for Experiment 3, or equivalents in course credit.

Materials

All stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; RRID:SCR_002881) within a MATLAB environment (MathWorks, Natick, MA, USA; RRID:SCR_001622). Stimuli were displayed on an ASUS LCD monitor (1920 x 1080 pixels, 120Hz, 50.9x28.6 cm screen size, 46.3pixels/°) on a “black” (RGB: [0 0 0], ±3cd/m2) background while participants were seated in a dimly lit room, approximately 70 cm away from the screen.

Procedure and Stimuli

Participants performed an attentional blink (AB) task (Raymond et al., 1992), in which on every trial a rapid series of visual stimuli was presented consisting of a sequence of 17 uppercase letters drawn from the alphabet but excluding the letters I, L, O, Q, U, and V. Every letter appeared maximally once per trial. Letters were presented at fixation in a mono-spaced font (font size: 40; corresponding to a height of approximately 1.2°) for 92 ms each.

Experiment 1

Participants were instructed to detect target letters within the rapid serial visual presentation (RSVP). The first target (T1: G or H) was always presented at the fifth position of the RSVP. On most trials (80%) it was followed by a second target (T2: D or K) at lag 2, lag 4 or lag 10 (respectively 183, 367 or 917 ms later). Each lag was
equally likely. T1 was presented in green (RGB: [0 255 0]), while T2 and the distractor letters were white (RGB: [255 255 255], ±320 cd/m²).

Crucially, there was a predictive relationship between the two targets (Fig. 1A). Namely, in the 80% of trials where a T2 was presented, the identity of T1 (e.g. G) predicted which T2 was likely (75%, e.g. D) or unlikely (25%, e.g. K) to appear. On the 20% remaining trials without a T2 a random distractor letter was presented at the T2-timepoint (every distractor letter was presented maximally once per trial). The mapping of T1 and T2 was counterbalanced over participants, so that for half of the participants the most likely target combinations were G-D and H-K while for the other half G-K and H-D were most likely. To be able to distinguish different lags in the absence of a T2 stimulus, 4 grey squares (RGB: [200 200 200], ±188 cd/m²; size: 0.35°; midpoint of each square centered at 1.30° from fixation) were always presented around the stimulus (T2 or distractor) at the T2-timepoint. Participants were instructed to use the timing information this cue provided when making decisions about the presence of a T2 (only the letters “D” or “K”, all other letters where distractors).

Following a 150 ms blank period at the end of the RSVP, participants gave their responses. First, they indicated whether or not they had seen any T2 by pressing the left or right shift key on the keyboard. The mapping between the keys and the response options was randomized per trial to decouple participants’ responses from the decision they had to make. Then they were asked to make a forced choice judgment about the T2 letter (D or K) that was presented by typing in this letter. Finally, they made a similar response about the identity of T1 (G or H). We used long response timeout durations of 5s and participants were instructed to value accuracy over response speed. The inter-trial interval, as defined by the time between the last response and the onset of the stream, was 500-750 ms.

The experiment consisted of two one-hour sessions on separate days within one week. In the first session, participants received instructions about the task and subsequently performed the task for 6 blocks of 75 trials (total 300 trials). The goal of the training session was to familiarize participants with the task. Besides, since we did not instruct participants about the predictive relationship between T1 and T2, some practice on the task was required for them to (implicitly) learn this relationship. In the second session, participants first received a summary of the instructions, after which the actual experiment started. Participants performed 6 blocks of 90
trials (total of 540 trials) of the AB task. The first three participants performed 6 blocks of 105 trials (630 trials). In both sessions participants received summary feedback about their performance at the end of each block, followed by a short break.

Experiment 2 (EEG)

The task in the EEG experiment was the same as in Experiment 1, except that in Experiment 2, we only asked participants to give one response by typing in the target letters they observed. In addition, we only used two different lags: lag 3 (275 ms; 2/3rd of trials) and lag 10 (917 ms; 1/3rd of trials). To further increase the number of trials, the ITI range was reduced to 200-400 ms.

Again, the experiment consisted of two different sessions within one week. The first session (1 hour) consisted of instructions followed by extensive training (720 trials over 6 blocks) on the task. Participants were not explicitly informed about the predictive relationship between the targets. In the second session (2 hours) we first prepared the participant for the EEG measurements (see below) and gave brief instructions about the task. Then, participants performed 12 blocks of 120 trials (total 1440 trials) of the AB task.

Experiment 3

To investigate the importance of T1 detection for expectation effects on conscious access, we adjusted the task we used in Experiment 1 to decrease the visibility of T1 (Fig. 3A). We now presented T1 in white instead of green, to make it stand out less among the other stimuli. Furthermore, T1 duration was staircased per participant such that participants could report T1 on roughly 75% of the trials. Starting in the second half of the training and continuing in the experimental session, after each block T1 duration was decreased by one frame (8 ms) if performance was higher than 85% and increased by one frame if performance was lower than 65%. To ensure T1 duration would not deviate too much from the duration of other stimuli, T1 duration was only allowed to be in the range of 42-142 ms (max. 50 ms different from other stimuli). The median duration of T1 in the second session was 125 ms. On 20% of trials no T1 was presented and a random distractor letter was presented.
presented instead. When both targets were present, T1 predicted which T2 was likely to follow with 75% accuracy.

We made a few changes to the task design to increase the efficiency of the design. The ITI was reduced to values between 300-500 ms. In addition, we only asked participants for one response. They were asked to type in any target letter they had seen during the trial and refrain from typing in a T1 and/or T2 letter when they did not see any. The response was confirmed by pressing the space bar on the keyboard or when a timeout of 4s had passed. To further increase the number of trials per condition, we decided to use only lag 3 (2/3rd of trials) and lag 10 (1/3rd of trials). Because T1 duration was staircased on an individual basis, the T1-T2 SOA differed between participants. On average, lag 3 corresponded to an SOA of 308 ms while lag 10 corresponded to an SOA of 950 ms.

Finally, we manipulated the instructions we gave to participants in order to see to what extent explicit knowledge of the relationship between T1 and T2 affected our results. As in Experiment 1, we tested participants during two separate sessions within one week. The first group of the participants (N=25) did not receive any explicit instruction about this relationship, similar to Experiment 1. The second group of participants (N=19) received explicit instructions about the T1-T2 relationship at the start of the second session, and a third group of participants (N=23) received those instructions already at the start of their first session.

The first session (1 hour) was used for instructions and training the participants on the task (10 x 75 trials). The experimental session in which participants performed the AB task lasted 1.5 hour and contained 16 blocks of 75 trials (1200 trials).

Experimental Design and Statistical Analysis: Behavioral

Preparatory steps were done with in-house MATLAB scripts. Statistical analyses (repeated measures ANOVAs and paired t-tests) were performed using JASP software (Love et al., 2015; RRID:SCR_015823). In situations where a specifically tested hypothesis did not yield a significant result, we used a Bayesian equivalent of the same test to quantify the evidence for the null-hypothesis (Rouder et al., 2012, 2017). In those cases, using
JASP’s default Cauchy prior, Bayes Factors (BF) were computed for each effect. To increase the interpretability in analyses with multiple factors, we used Bayesian model averaging to get a single BF for each effect in ANOVA’s. This BF is the change from prior to posterior inclusion odds, and can intuitively be understood as the amount of evidence the data gives for including an experimental factor in a model of the data. The BF will either converge to zero when the factor should not be included, or to infinity when it should be included in the model. Values close to one indicate that there is not enough evidence for either conclusion. We use the conventions from Jeffreys (1967) to interpret the effect sizes of our Bayesian analyses.

Experiment 1

In our behavioral analyses we looked at the T2 detection performance, given that T1 was correctly identified. A response was considered to be correct when (1) the participant indicated no T2 was present when no T2 was presented or (2) the participant correctly indicated a T2 was present and subsequently reported the correct target letter. Since expectation is only a meaningful concept when a T2 target was presented, the T2 absent trials, on which a distractor letter was presented instead of a T2, were not taken into consideration for the main statistical analyses. Trials where one of the responses was missing were deleted from all analyses.

Percentage correct was used in a 3 x 2 repeated measures ANOVA with the factors lag (lag 2, lag 4, lag 10) and expectation (valid, invalid). In a control analysis, we repeated our analyses for Experiment 1 based on the T2 detection responses (ignoring the accuracy of the T2 identification) as dependent variable (see also Results). Since the seen/miss response is orthogonal to the specific expectations about target letters, this analysis rules out simple response biases as a potential cause of our effects.

Experiment 2 (EEG)

The behavioral analyses for the EEG experiment were similar to those for Experiment 1. However, the factor lag had only 2 levels (lag 3, lag 10). Percentage correct T2 detection was computed as in Experiment 1 using only the trials on which T1 was correctly reported. A response was considered to be correct when the letter a participant entered was the letter that was presented or when a participant refrained from entering a letter
when none was presented for the T2 absent trials. In addition, we computed a T2 detection measure to use in a control analysis: if a participant typed in any letter, we categorize the response as a “target seen” response, otherwise we call it a “target absent” response. This outcome measure was used in a control analysis.

Experiment 3

In this experiment, participants gave only one response by typing in the target they had perceived. Trials on which no response was given or on which an impossible response was given (e.g. two T1 targets reported) were excluded from analyses. For T1 and T2 separately, we assessed the accuracy of the responses. The definition of correct and incorrect responses was the same as in Experiment 2 and we also used the same T2 detection measure.

Subsequently, T2 percentage correct detection was used in a 2 x 2 x 2 x 3 mixed ANOVA with the within-subject factors lag (lag 3, lag 10), expectation (valid, invalid) and T1 visibility (T1 seen, T1 missed) and the between-subject factor instruction (none, start session 2, start session 1). As mentioned before, this between-subject factor was included to find out whether predictive effects would be modulated by explicit knowledge of the relation between T1 and T2. To investigate the effect of T1 visibility in more detail, we followed up the main analyses by other mixed ANOVAs in which we first split up the dataset based on T1 visibility. In situations where we found interactions with the factor instruction, we compared the effects of lag and expectation separately per instruction condition using repeated measures ANOVAs and paired-samples t-tests.

Finally, to test for an interaction between expectation validity and lag, we combined behavioral data from all experiments in a post-hoc analysis. Only trials on which T1 was correctly identified were used. For Experiment 1 we averaged data for lag 2 and lag 4 to create an average “short lag” condition. Subsequently, these data were entered into a 2 x 2 x 3 mixed ANOVA with the within-subject factors lag (short, long) and expectation (valid, invalid) and the between-subject factor experiment (Experiment 1, Experiment 2, Experiment 3).

Electroencephalographic Measurements
EEG was recorded with a BioSemi ActiveTwo system and sampled at 512 Hz (BioSemi, Amsterdam, The Netherlands). Potentials were measured from sixty-four scalp electrodes, along with two reference electrodes on the earlobes and four electrodes measuring horizontal and vertical eye movements. After data acquisition, EEG data was pre-processed with the FieldTrip toolbox for MATLAB (Oostenveld et al., 2011; RRID:SCR_004849). First, data were re-referenced to the linked earlobes, high-pass filtered at 0.01 Hz, and epoched from -0.750 to 1s surrounding the onset of T2. Data were visually inspected and trials and/or channels containing artifacts not related to eye blinks were manually removed, resulting in deletion of on average 9.1% (±3.9%) of trials and 2.0 (±1.7) channels. Independent component analysis was used to identify components related to eye blinks or other artifacts that could easily be distinguished from other EEG signals. After the independent component analysis, previously deleted channels were reconstructed based on a nearest neighbor approach. Trials were baseline corrected to the average amplitude prior to T1 onset (-0.750 to -0.275). As a final step, we applied a 40Hz low-pass filter to the trial data, after which ERPs were created separately for each condition of interest.

Experimental Design and Statistical Analysis: Electroencephalography

All EEG analyses are based exclusively on trials where T2 appeared at lag 3 and T1 was correctly identified. We used a combination of Fieldtrip (Oostenveld et al., 2011) and in-house MATLAB scripts to perform our analyses. As a first step, we performed cluster-based permutation tests (Maris and Oostenveld, 2007) on the time-window 0-750 ms from stimulus onset to isolate significant ERP events relating to expectation validity (valid, invalid; irrespective of T2 visibility) or T2 visibility (seen, missed; irrespective of validity) or the interaction between these factors. Next, we used an in-house built MATLAB script to isolate the significant events as clusters in time and space. For this purpose, we computed an average difference wave over all channels that were part of the cluster at any point in time. Subsequently, the onset and offset of a cluster were defined as the time period around the maximum difference where the difference did not drop below 50% of this maximum and where at least one channel showed a significant effect. We then selected the 10 channels that showed the largest effect in this time-window. One of the observed events reflected a mixture of the...
traditionally observed P3a and P3b components (Sergent et al., 2005; Volpe et al., 2007). Therefore, we split the event into two clusters by manually selecting either the 32 most anterior or 32 most posterior EEG channels (from the central midline) before running the cluster selection procedure.

As an alternative way to establish potential interactions between T2 detection and validity, we inspected the clusters that were isolated in the previous step in more detail. This may be a more powerful (but also less sensitive) way to detect small effects, because data is averaged over more time-points and channels. Within each of the clusters, we performed a 2 x 2 repeated measures ANOVA (and its Bayesian equivalent, see also Behavioral Analysis) with the factors T2 detection (seen, missed) and expectation validity (valid, invalid) on the cluster data averaged over channels and time. To prevent double dipping, in each analysis we only considered the effects orthogonal to the one that was used to define the cluster (e.g. not testing the effect of expectation in a cluster defined based on the expectation effect).

Results

Experiment 1: (how) do expectations affect conscious access?

In the first experiment we addressed the question if expectations about the likelihood of stimulus identity modulate the likelihood of conscious access, and if so, in what direction. To do so, we used the attentional blink paradigm (Raymond et al., 1992). The attentional blink is an impairment in the conscious perception of the second of two target stimuli that are presented in rapid succession when the initial target was correctly perceived Here we modified the paradigm in such a way that the first target (T1: the letter G or H, in green) predicted which of the second targets would most be likely to appear in case a T2 target was presented (T2: the letter D or K, predicted=75%, unpredicted=25%, in white, Fig. 1A). On 20% of trials we presented a random distractor letter instead of a T2 target. At the end of each stream of letters, participants gave three responses. First, they indicated whether or not they had seen any of the two T2 targets (“seen”/”unseen” response). Second, they were prompted to make a forced-choice judgment about the identity of T2 (whether the letter D
or K was presented). Third, participants had to make a similar forced-choice decision about the identity of T1 (whether the letter G or H was presented) (see Methods for details). Participants were not explicitly instructed about the predictive relationship between T1 and T2.

In Figure 1 we plot the percentage of trials in which T2 was correctly detected and T1 discrimination was also correct (average T1 accuracy was 94.20%, sd=5.77%) for the three different lags (lag 2, 4 and 10). T2 was considered to be detected correctly when participants indicated to have seen it (based on the first response) and correctly identified it (based on the second response). Overall, there was a clear attentional blink, as reflected by reduced T2 detection when the time (i.e. lag) between T1 and T2 was shorter (Figure 1B, main effect of lag: $F_{2,48}=48.15$, $p<0.001$). Importantly, expectations modulated T2 detection rate. T2 detection was significantly better when T1 validly predicted T2 (black lines) compared to when the expectation was invalid (gray lines, main effect of validity: $F_{1,24}=7.10$, $p=0.014$, no significant interaction between lag and validity: $F_{2,48}=1.30$, $p=0.283$). These results extend several previous studies (Chang et al., 2015; Melloni et al., 2011; Pinto et al., 2015; Stein et al., 2015; Stein & Peelen, 2015) by showing that conscious perception is (partly) determined by the transitional probability of the input the brain receives.

While these data support the notion that valid expectations trigger access to consciousness, it has been recognized that such findings may not solely be due to changes in perception, but perhaps (also) due to changes in decision criteria or response biases (Gayet et al., 2014b; Yang et al., 2014; Attarha and Moore, 2015). To rule out the possibility that our effects could be explained by a response bias in which people simply report the target letter that they expected based on T1, irrespective of whether they consciously perceived T2, we performed an analysis with T2 detection (instead of T2 discrimination, see Methods) as the dependent variable. This analysis takes into account only participants’ first response (the “seen”/“unseen” response), regardless of whether subsequent T2 letter identification was correct or not. Crucially, this analysis cannot be influenced by any decision/response biases because the response was orthogonal to the participants’ expectation. Information about the most likely letter to appear cannot predispose participants to better determine whether a target letter was presented at all. Still, we observed a qualitatively similar pattern of
results (main effect of validity: $F_{1,24}=5.47$, $p=0.028$). This finding suggests that validity indeed boosted participants conscious access of T2, instead of merely eliciting a shift in response bias.

Experiment 2: EEG markers of conscious and nonconscious expectation violations

Subsequently, we tested whether expectation violations can be elicited by nonconsciously processed unpredicted stimuli or whether conscious perception of a stimulus is a prerequisite for it to trigger neural expectation error responses. To test this, we measured subjects’ brain activity with EEG while they performed a similar task as in Experiment 1. First, we replicated the behavioral effects of Experiment 1 (Fig. 1C). Overall, T1 performance was high ($M=93.61\%$, $sd=7.31\%$) and T2 detection was higher at lag 10 than at lag 3 (main effect of lag: $F_{1,28}=128.72$, $p<0.001$), reflecting a robust attentional blink. More importantly, validly predicted T2s were discriminated better than invalidly predicted T2s (main effect of validity: $F_{1,28}=9.49$, $p=0.005$). The effects were similar in a control analysis where we considered the percentage of T2 seen responses (regardless of the exact letter participants entered), making it less likely that our effect can be explained by a response bias (main effect of validity: $F_{1,28}=4.23$, $p=0.049$). In this experiment, the validity effect was significantly modulated by lag (validity x lag: $F_{1,28}=5.86$, $p=0.022$), an effect that was numerically similar, but not significant in Experiment 1.

Participants performed better for valid than invalid trials at lag 3, but there was no convincing evidence for an effect of expectations at lag 10 (lag 3 validity effect: $t_{28}=3.40$, $p=0.002$; lag 10 validity effect: $t_{28}=0.98$, $p=0.334$). Thus, effects of expectations were larger in the time window in which T2 more often goes unperceived.

Next, we investigated potential differences in the neural processing of predicted and unpredicted stimuli, as a function of stimulus awareness. To this end, we contrasted invalidly and validly predicted T2s and tested this difference using cluster-based permutation testing, correcting for multiple comparisons across both time (0-750 ms) and (electrode) space (see Fig. 2 and Methods) (Maris and Oostenveld, 2007). We found one significant difference over fronto-central electrode channels, which reflected greater T2-elicited negativity for
invalid compared to valid trials between 174-314 ms (p=0.015, Fig. 2B), therefore potentially reflecting some
type of mismatch response. We then further analyzed this event to test whether the difference was modulated
by, or dependent on, conscious perception of T2. Crucially, the size of this fronto-central mismatch component
was independent of T2 awareness (F_{1,28}=0.04, p=.850, BF=.254, Fig. 2C), indicating that both seen and unseen
T2’s generated a fronto-central mismatch response.

Additionally, analyses of T2 visibility effects (irrespective of expectation validity) replicated previously
reported findings (Kranzioch et al., 2003; Sergent et al., 2005; Harris et al., 2013). In these analyses, we
examined the difference in ERPs following seen and missed T2s using a cluster-based permutation test (Fig. 3),
revealing two significant events. First, a significant negative difference could be observed over (left) posterior
electrodes from 170-355 ms after T2 onset (p=0.010; Fig 3A). This event was followed by a significant long-
lasting positive event (p<0.001), reflecting a mixture of the P3a and P3b components, extending over frontal
and central electrodes.

Subsequently, we had a closer look at the interactions between conscious access and expectation validity.
Therefore, we analyzed the ERP events that were isolated in the previous step in more detail (Fig. 3B-G). For
this analysis we first isolated the traditionally observed AB-related P3a and P3b ERP components from the long-
lasting positive ERP event that differentiated between seen and missed T2s (Sergent et al., 2005). Doing so
resulted in an early positive P3a cluster (Fig. 3D) over fronto-central channels that was significant between 395-
586 ms and a somewhat later positive P3b cluster (Fig. 3F) over more posterior parietal channels, which was
significant between 445-611 ms. Within each of these clusters we performed repeated measures ANOVAs with
the factors validity and T2 detection. For none of the events we found evidence that the T2 detection effect
was modulated by expectation validity (early left-posterior event: F_{1,28}=0.29, p=0.597, BF=.260; P3a: F_{1,28}=1.56,
p=0.222, BF=0.230; P3b: F_{1,28}=2.10, p=0.159, BF=0.296), though the BF values suggest that the evidence for the
absence of such interactions is moderate at best. This is somewhat surprising, because especially the late
positive events have previously been related to conscious access (Sergent et al., 2005; Rutiku et al., 2015) and metacognition (Desender et al., 2016). However, recent investigations show it may also reflect even later more cognitive processes, merely arising as a consequence of becoming consciously aware of information (Pitts et al., 2014; Silverstein et al., 2015). We did not find evidence that the amplitude of these ERP events was modulated by expectation validity, which may suggest that once a stimulus has been perceived consciously, it is irrelevant whether or not the expectation was valid.

Finally, we directly tested for an interaction between conscious access and expectation by comparing the validity ERP effect (invalid-valid) for T2 seen and T2 missed trials in a cluster-based permutation test (this analysis takes into account the entire scalp topography). Again, no significant interactions between these factors were observed (all clusters p>0.10).

Experiment 3: the role of conscious awareness in implementing top-down expectations

In our final experiment, we addressed the question whether expectation formation itself can unfold in the absence of awareness and subsequently influence conscious access (Fig. 4). To address this question, we changed the color of T1 from green to white and for each subject stairsed T1 duration in such a way that T1 was correctly identified on approximately 75% of the trials (actual T1 identification performance: M=76.03%; sd=8.65%). T1 duration did not differ between trials where T2 was seen and trials where T2 was missed (T2 detection: t_{66}=0.31, p=0.752; T2 seen: M=117.42 ms; T2 missed: M=117.46 ms), which indicates that T1 visibility was not determined by stimulus duration. Likely, internal fluctuations in the system (e.g. variability in attention) must be causing participants to sometimes see T1 and sometimes miss it. Moreover, on 20% of trials no T1 was presented (but replaced by a distractor). Further, to test to what extent explicit knowledge of the predictive relationships between stimuli would increase the validity effects, we varied the moment at which explicit information about the predictive relations between T1 and T2 was provided. The experiment consisted of a training session and a test session on separate days. A first group of subjects received no explicit instructions about the predictive relations in either session and had to learn them implicitly through experience with the task; the second group received explicit instructions about the T1-T2 relations in the test session only,
but not in the first training session; and the third group received explicit instructions already from the start of the experiment.

(Insert Figure 4 here)

T1 visibility strongly affected T2 detection. When T1 was seen, T2 detection was markedly lower than when T1 was missed (main effect of T1 awareness: $F_{1,64}=4.62$, $p=0.035$), in particular at short lags (T1 awareness x lag: $F_{1,64}=72.95$, $p<0.001$). Validly predicted targets were detected more often (main effect of validity: $F_{1,64}=33.39$, $p<0.001$). The effect of expectation validity on T2 detection varied as a function of T1 awareness and instructions (T1 awareness x validity: $F_{1,64}=40.55$, $p<0.001$; validity x instruction: $F_{1,64}=5.91$, $p=0.004$; T1 awareness x validity x instruction: $F_{2,64}=11.33$, $p<0.001$). When T1 was seen (Fig. 4B), a clear attentional blink was observed (main effect of lag: $F_{1,64}=170.01$, $p<0.001$) and validly predicted targets were more often detected than invalidly predicted targets (main effect of validity: $F_{1,64}=64.97$, $p<0.001$) (as in Experiment 1 and 2). Like in the previous experiments, a control analysis considering only the percentage of T2 seen responses (regardless of the exact letter participants entered) also revealed a significant effect of validity (main effect of validity: $F_{1,64}=65.83$, $p<0.001$), making it unlikely that response biases are causing the effect. Interestingly, we also observed a significant attentional blink for missed T1’s, reflecting a nonconsciously elicited AB (main effect of lag: $F_{1,64}=74.42$, $p<0.001$). This AB effect cannot be explained by an overall T2 detection performance benefit for targets that are presented later in the trial because the AB was larger for trials on which T1 was presented but missed compared to trials on which no T1 was presented (lag x T1 presence: $F_{1,66}=24.19$, $p<0.001$). However, although missed T1’s triggered an AB, expectation validity did not affect T2 detection performance for missed T1’s (main effect of validity: $F_{1,64}=0.35$, $p=0.554$), regardless of the type of instruction participants received about the predictive relation between T1 and T2 (validity x instruction: $F_{2,64}=0.64$, $p=0.533$). A Bayesian equivalent of the repeated-measures analysis strongly suggested validity should not be included in a model of the data ($BF=.084$).
The above results highlight that only when T1 was seen, valid expectations facilitated T2 detection. A post-hoc analysis on T1-seen trials only revealed that this effect was modulated by how explicitly we instructed participants about the predictive relationship between T1 and T2 (validity x instruction: F_{2,64}=14.83, p<0.001). The validity effect, as defined by the difference between valid and invalid trials, averaged across the two lags, increased with more explicit instructions (group 1: 1.87%, group 2: 19.53%, group 3: 26.27%). These results indicate that, not only does the visibility of T1 define the predictive impact on T2 detection, but also the extent to which these predictive relations are (explicitly) known affects the impact of expectations on conscious access. This may also explain why the validity effect appeared more pronounced in Experiment 3 compared to Experiments 1 and 2, because in experiment 1 and 2 subjects were not explicitly instructed about the predictive relations between T1 and T2.

Finally, in contrast to Experiment 2, on T1 seen trials the validity effect was independent of lag (validity x lag: F_{1,64}=1.750, p=0.191). Since we anticipated stronger expectation effects at short lags, behavioral data from all three experiments was combined in a post-hoc analysis. Only trials on which T1 was correctly identified were used and for Experiment 1 we averaged data for lag 2 and lag 4 to create an average “short lag” condition. A significant interaction between validity and lag showed that across all experiments, the expectation effect was stronger at short lags compared to the long lags (validity x lag: F_{1,118}=5.73, p=0.018; no validity x lag x experiment interaction: F_{2,118}=0.065, p=0.937).

Discussion

In this report we investigated three important questions regarding the intricate relationship between top-down expectations and conscious awareness. The first question that we addressed was how prior information about the identity of an upcoming stimulus influences the likelihood of that stimulus entering conscious awareness. Using a novel attentional blink paradigm in which the identity of T1 cued the likelihood of the identity of T2, we showed that stimuli that confirm our expectation have a higher likelihood of gaining access to conscious awareness than stimuli that violate our expectations, especially at short lags. The expectation effect was qualitatively similar across all three experiments, regardless of subtle experimental differences in task design.
and overall performance between those experiments. Furthermore, it could not be explained by simple shifts in response criterion, because it was also present for a dependent measure orthogonal to the expectation manipulation. Together, this suggests that valid expectations amplify the perceptual strength of a stimulus and therefore increase the chance of conscious access, possibly due to the sharpening of its neural representations (Kok et al., 2012). This interpretation is supported by previous experiments that have shown varying effects of expectations on (subjective) perception, such as studies showing prior knowledge increases the speed (Chang et al., 2015; De Loof et al., 2016; Melloni et al., 2011; Pinto et al., 2015) and accuracy (Stein and Peelen, 2015) of stimulus detection. Furthermore, our findings complement recent studies showing that the AB can be reduced when there is knowledge about temporal statistics of the task (Lasaponara et al., 2015; Visser et al., 2015) or when the latency of T2 targets is explicitly cued (Martens & Johnson, 2005; Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005). In addition, two recent reports have shown that, in some subjects, expecting the presentation of a stimulus can even elicit an illusory stimulus percept when no stimulus is presented, when attention is diverted (Mack et al., 2016; Aru and Bachmann, 2017). Future experimentation is required to shed light on the generalizability of our effect to simpler tasks. Such experiments may also consider using other measures of subjective perception (e.g. perceptual awareness scale) (Overgaard et al., 2006).

The second question that we addressed was related to the extent to which nonconscious stimuli can trigger prediction error responses, as measured with EEG. Over the last 20 years, we and others have shown that nonconscious information processing is rather sophisticated (Dehaene and Changeux, 2011; van Gaal and Lamme, 2012), and that a diverse range of high-level cognitive processes can unfold nonconsciously (Dehaene et al., 2001; Custers and Aarts, 2005; Lau and Passingham, 2007; Pessiglione et al., 2007, 2008, van Gaal et al., 2010, 2012, 2014). Interestingly, in Experiment 2 we found that expectations that are violated by a nonconscious stimulus trigger a stronger negative fronto-central ERP component than expectations that are confirmed. This neural event was similar for trials on which T2 was seen and on trials where T2 was missed, highlighting that conscious awareness of a stimulus is not a prerequisite for it to trigger a prediction error response (Mathews et al., 2014; Malekshahi et al., 2016). This effect may reflect a mismatch signal, similar to the mismatch negativity (Näätänen et al., 2007), which is a negative deflection following oddball stimuli that
develops 100-200 milliseconds after stimulus onset. Sometimes this effect lasts longer, in some experiments
till ~400 milliseconds, depending on the specifics of the task and stimulus material (Pöppel, 2004; Stefanics et
al., 2011; Kimura and Takeda, 2015). While in terms of interpretation this effect is similar to a mismatch effect,
its topography is slightly different than a typical visually evoked MMN, which generally peaks more posteriorly,
although considerable variation in its topography has been reported (Pöppel, 2004). Alternatively, it is possible
that the higher activation for valid compared to invalid trials corresponds to the frontal selection positivity
(FSP), which is a well-known marker of non-spatial attentional processes (Kenemans et al., 1993). In our
paradigm, this could be explained as improved attentional selection when expectations are confirmed.

Although the exact nature of the observed component deserves future experimentation, the key finding is that
the effect was independent of T2 perception and purely depends on the validity of the expectation. This is in
line with studies that have shown context influences on nonconscious information processing (Nakamura et al.,
2007; Van Opstal et al., 2011; Gayet et al., 2014a; Rohaut et al., 2016), studies that have shown that the MMN
can be observed when the expectation violations are unattended (Bekinschtein et al., 2009; Stefanics et al.,
2011; King et al., 2013; Dykstra and Gutschalk, 2015; Kimura and Takeda, 2015) and more generally evidence
for relatively high-level processing of nonconscious stimuli (Luck et al., 1996; van Gaal and Lamme, 2012;
Silverstein et al., 2015). Nevertheless, the absence of interactions in the ERP is also somewhat surprising (but
see also Rutiku et al., 2016), because as noted earlier such interactions between expectation validity and
conscious T2 detection were present in behavior. A neural basis for this effect should exist, but may be very
subtle. Recently, a study by Aru et al. (2016) found early (<100 ms) differences in signal amplitude over
posterior channels that predicted the behavioral benefit of prior knowledge on the detection of stimuli
presented at the threshold of perception. Another potentially interesting signature to investigate could be the
onset of components related to conscious perception (Melloni et al., 2011) and how they relate to
expectations. Moreover, it is possible that instead of signal strength, it is the signal-to-noise ratio or sharpness
of the representation that is improved (Kok et al., 2012). Possibly, valid expectations do not modulate the
amplitude of the neural response, but increase the specificity of the neural representation.
In the final experiment, we showed that conscious perception of T1, initiating the expectation, is a prerequisite for influences on conscious access to occur. On the subset of trials where subjects did not see T1, there was no expectation effect T1 on T2 detection performance. This result contrasts with findings from a recent study that suggested that some priors may operate nonconsciously (Chang, Baria, Flounders, & He, 2016). Chang and colleagues presented participants with masked grey-scale natural scene images and found that the nonconscious processing of these images improved subsequent recognition of their degraded counterparts, so-called “Mooney images”, presented seconds later. One explanation for this difference is that the priors on which the effects of Chang et al. relied may be more automatic and hard-wired than the relatively arbitrary relationships that people have to learn and actively use in our experiments. It is possible that lower-level, automatic expectations are more easily processed outside of awareness compared to the more active ones studied here.

Further, it is also possible that with more training we would find nonconscious expectation effects. However, since subjects were already trained on the task on a separate day before performing the experimental session, this possibility seems unlikely. We did observe greater validity effects when subjects were made explicitly aware of the predictive nature of T1, suggesting that explicit knowledge of stimulus associations can facilitate the effects of stimulus-induced expectations. Finally, it should be noted that we did not test the full range of timing intervals between T1 and T2. It has been shown and proposed that the processing of nonconscious stimuli is relatively fleeting (Greenwald et al., 1996; Dehaene et al., 2006; but see King et al., 2016), so it is conceivable that the T1-T2 lags that we have used here may have been too long to observe expectation effects triggered by unseen T1’s. Further, a significant attentional blink was observed on trials on which T1 was missed, indicating that attention was still captured by a missed T1 at the T1-T2 lags used here. This latter result is in line with evidence showing that nonconscious stimuli are able to trigger attentional capture (Ansorge et al., 2009; Mulckhuyse and Theeuwes, 2010; Hsieh et al., 2011) and with a study showing lower T2 detection for T1’s that were missed compared to trials without a T1 (in that experiment this effect was independent of lag (Nieuwenstein, Van der Burg, Theeuwes, Wyble, & Potter, 2009)).
In summary, three main conclusions can be drawn from the present series of studies. First, expectation confirmation, compared to violation, increases the likelihood of conscious awareness, suggesting that valid expectations amplify the perceptual strength of a stimulus. Second, nonconscious violations of conscious expectations are registered in the human brain. Third, however, expectations need to be implemented consciously to subsequently modulate conscious access. These results suggest a differential role of conscious awareness in the hierarchy of predictive processing, in which the active implementation of top-down expectations requires conscious awareness, whereas a conscious expectation and a nonconscious stimulus can interact to generate prediction errors. How these nonconscious prediction errors are used for updating future behavior and shaping trial-by-trial learning is a matter for future experimentation.

References


Mulckhuys M, Theeuwes J (2010) Unconscious attentional orienting to exogenous cues: A review of the


Neurobiol Exp (Wars) 64:295–301.


Figure Legends

Figure 1. Task design and behavioral results of Experiment 1 and 2. (A) The trial structure of the attentional blink task used in Experiment 1 and 2. Each trial consisted of a stream of rapidly presented letters in which predefined target letters had to be detected and then reported at the end of the stream. The first target (T1: a
green G or H) always appeared at the fifth position. The second target (T2: D or K), was presented at varying SOAs (lags) after the first one and was marked by placeholders. The identity of T1 predicted which of the T2 targets was most likely to appear, thereby introducing validly and invalidly predicted T2 targets. On 20% of the trials no second target was presented and a random distractor letter was presented instead. (B) Percentage correct T2 target detection at each of the T1-T2 lags, for valid expectations, invalid expectations and T2 absent trials in Experiment 1. Validly predicted T2’s were significantly more often perceived than invalidly predicted T2’s. (C) Percentage of T2 target detection at each of the T1-T2 lags, after a valid or invalid expectation or on a T2 absent trial for Experiment 2. Again, validly predicted T2’s were more often perceived, in particular at short lags. Error bars represent SEM.

Figure 2. ERP effects related to T2 prediction validity. (A) Topographic maps of the difference between validly and invalidly predicted T2s over time (0 = T2 onset). Cluster-based permutation tests were used to isolate the significant events, while correcting for multiple comparisons across time and (electrode) space. On each head map, channels with a significant effect for at least 50% of its time window are highlighted. (B) The average ERP time-course of the 10 channels shown on the headmap on the left, shown separately for each validity condition. The significant time-window is marked by a black line above the x-axis. Invalidly predicted T2s were associated with greater fronto-central negativity than validly predicted T2s. (C) Bar graphs showing the average amplitude of the four conditions (visibility x prediction) for the significant neural event shown in B. In all plots error bars represent SEM.

Figure 3. ERP effects related to T2 visibility analyses. (A) Topographic maps showing the difference between seen and missed T2s over time (0 = T2 onset). Cluster-based permutation tests were used to isolate the significant events while correcting for multiple comparisons across time and (electrode) space. On each head map, channels showing a significant difference for at least 50% of its time window are highlighted. Three events were isolated based on the permutation tests. (B,D,F) For each of the events individually, the average ERP time-course of the 10 channels shown on the headmap on the left, seperately for T2 seen and T2 missed
conditions is shown. The significant time-window is marked by a black line above the x-axis. (C,E,G) Bar graphs showing the average amplitude of the four conditions (visibility x prediction) for the significant neural events shown in B,D,F. In all plots error bars represent SEM.

Figure 4. Task design and behavioral results of Experiment 3. (A) Trial structure of the task used in Experiment 3. T1 visibility was staircased at approximately 75% correct by manipulating its duration (on 20% of trials no T1 was presented). (B) Percentage of correct T2 target detection at each of the T1-T2 lags, after a valid or invalid expectation and on a T2 absent trials, for trials where T1 was correctly reported (T1 seen). As in Experiment 1 and 2, when T1 was seen, validly predicted T2’s were more often detected than invalidly predicted T2’s. (C) Solid lines show percentage of T2 target detection at each of the T1-T2 lags, after a valid or invalid expectation and on a T2 absent trials, for trials where T1 was presented but missed. In contrast to T1 seen trials (B), when T1 was not seen, validity did not enhance T2 detection. However, a missed T1 still triggered a significant attentional blink, as compared to trials on which no T1 was presented (dotted line). Error bars represent SEM.
A. Task Design: Exp 1 & 2 (EEG)

**T1-T2 relationship**
- 60% valid prediction
- 20% invalid prediction
- 20% no T2 target

**T2:** D, K or distractor

**T1:** G or H

**T1-T2 SOA**
- Exp 1:
  - lag 2: 183 ms
  - lag 4: 367 ms
  - lag 10: 917 ms
- Exp 2 (EEG):
  - lag 3: 275 ms
  - lag 10: 917 ms

**Durations**
- All stimuli: 92 ms

B. Exp 1

![Graph showing percentage of correct T2 detection vs lag]

- $p=0.014$
- $p=0.283$

C. Exp 2 (EEG)

![Graph showing percentage of correct T2 detection vs lag]

- Valid
- Invalid
- T2 absent

**Condition**
- $p=0.005$
- $p=0.022$
A. Difference topographies (Invalid - Valid)

B. Early central negativity

C. Comparison of average amplitudes

- Valid
- Invalid

seen
miss
A. Difference topographies (Seen - Miss)

B. Early left-posterior negativity

C. Average amplitude (µV)

D. Anterior positivity (P3a)

E. Average amplitude (µV)

F. Posterior positivity (P3b)

G. Average amplitude (µV)
A. Task Design: Exp 3
T1-T2 relationship
48% valid prediction
16% invalid prediction
36% no T1 or T2 target

T2: D, K or distractor
T1: G, H or distractor

T1-T2 SOA
lag 3: ±308 ms
lag 10: ±950 ms

Durations
T1: 42-142 ms
T2 & distractors: 92 ms

B. Exp 3: T1 seen

% correct T2 detection

\[ \text{lag 3} \quad \text{lag 10} \]

= p<0.001
\( \times \) p=0.191

C. Exp 3: T1 missed / absent

% correct T2 detection

\[ \text{lag 3} \quad \text{lag 10} \]

Condition
- Valid
- Invalid
- T2 absent
- T1 absent

\[ \text{p=0.554} \quad \text{p=0.104} \]