JNeurosci
THE JOURNAL OF NEUROSCIENCE

JNeurosci Accepted Manuscript

Research Articles: Behavioral/Cognitive

Dynamic interactions between top-down expectations and conscious awareness

Erik L. Meijs[a,b]**, Heleen A. Slagter**[c,d]**, Floris P. de Lange**[b] **and Simon van Gaal**[b,c,d]

[a]*Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands*

[b]*Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the Netherlands*

[c]*University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands*

[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

Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.

1   **Dynamic interactions between top-down expectations and conscious awareness**

2   *Running title: Interactions between expectations and awareness*

3

4   Erik L. **Meijs**[a,b], Heleen A. **Slagter**[c,d], Floris P. **de Lange**[b], & Simon **van Gaal**[b,c,d]

5

6   [a] Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen,

7   the Netherlands

8   [b] Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6500 HB Nijmegen, the

9   Netherlands

10   [c] University of Amsterdam, Department of Psychology, 1001 NK Amsterdam, the Netherlands

11   [d] University of Amsterdam, Amsterdam Brain and Cognition (ABC), 1001 NK Amsterdam, the Netherlands

12

13   Corresponding author: Simon van Gaal, University of Amsterdam, Department of Psychology, 1001 NK

14   Amsterdam, the Netherlands, simonvangaal@gmail.com

15

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

17

18   The authors declare no competing financial interests.

19

26

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.

2

48 A rapidly growing body of work indicates that sensory processing is strongly influenced by expectations that we

49 have about likely states of the world. Such expectations are shaped by the context in which we are operating,

50 but also by learning, past experience and our genetic makeup (Friston, 2005; Bar, 2009; Summerfield and de

51 Lange, 2014). Expectations are typically thought to originate from higher-level brain regions, such as the

52 (pre)frontal cortex, which may guide information processing in lower-level sensory regions via top-down

53 projections. In this framework, what we consciously see is proposed to be strongly influenced by the brain's

54 expectations about, or its best guess of, the outside world (Gregory, 1980; Hohwy, 2012; Panichello et al.,

55 2013). Initial studies support the idea that the brain uses information in the environment to build expectations

56 of stimulus frequency or conditional probabilities to modify perceptual processing (Bar, 2004; Kok et al., 2012).

57 These ideas have been formalized in theoretical models, such as predictive coding and sequential sampling

58 models (Friston, 2005; Ratcliff and McKoon, 2008; Clark, 2013). Although these frameworks are attractive in

59 their simplicity, how exactly expectations shape conscious perception, and to what extent awareness guides

60 the formation of expectations, is still largely unknown.

61   At present, there are (at least) three issues that need to be resolved to further our understanding of the

62 relationship between expectations and consciousness. The first issue relates to the effect that expectations

63 may have on conscious awareness itself. It has been shown that valid expectations increase the speed of

64 conscious access (Melloni et al., 2011; Pinto et al., 2015; Stein and Peelen, 2015; De Loof et al., 2016) and may

65 help selecting or facilitating stimulus interpretation when (visual) input is ambiguous or noisy (Bar et al., 2006;

66 Denison et al., 2011; Panichello et al., 2013; Chang et al., 2015; Aru et al., 2016). It is yet an open question

67 whether expectations can boost an otherwise unseen stimulus into conscious awareness, thereby enabling the

68 switch from a nonconscious to a conscious stimulus representation, instead of merely facilitating its cognitive

69 interpretation or its speed of appearance in time.

70   Second, it is an open question to what extent prediction errors, arising in a situation of invalid expectations,

71 can be registered outside of conscious awareness. It has been shown that "oddball" stimuli (e.g. simple

72 violations in auditory tone sequences) elicit early mismatch responses in electrophysiological signals: the

73 mismatch negativity (MMN) (Pöppel, 2004; Näätänen et al., 2007). Interestingly, MMN's can even be observed

74   when attention is distracted from the tone sequences (Bekinschtein et al., 2009) or in several reduced states of

75   consciousness, such as sleep (Ruby et al., 2008), anesthesia (Koelsch et al., 2006) and vegetative state

76   (Bekinschtein et al., 2009). This suggests that the MMN reflects a pre-attentive nonconscious prediction error

77   signal (Näätänen et al., 2001; Stefanics et al., 2011; Kimura and Takeda, 2015). However, it remains debated

78   whether these signals originate in model-based comparisons of expectations to new input or merely reflect

79   passive low-level sensory adaptation to repeated inputs (Garrido et al., 2009; Stefanics et al., 2016). The one

80   study in which these mechanisms were dissociated in a nonconscious state showed adaptation remains

81   operative during sleep, whereas prediction error detection disappears (Strauss et al., 2015), questioning the

82   notion that prediction errors may be registered nonconsciously.

83   The final issue concerns the role of awareness in *implementing* expectations. Many expectation-based

84   models assume that expectations are implemented via top-down neural activation. Interestingly, influential

85   theories of consciousness suggest that conscious access requires similar top-down interactions between

86   higher-level (e.g. prefrontal) and lower-level (e.g. visual) brain regions, referred to as feedback or recurrent

87   processing (Lamme and Roelfsema, 2000; Dehaene et al., 2006). Information that does not reach conscious

88   access is thought to only trigger feedforward activity or local recurrent interactions between posterior brain

89   regions. Therefore, it is unclear how nonconscious information, in the absence of feedback signals from higher-

90   order cortical areas, could lead to the implementation of expectations.

91

92   **Materials & Methods**

93   *Participants*

94   We tested 26 participants in Experiment 1 (21 females, age 19.5±1.3 years), 85 participants in Experiment 2 (63

95   females, age 22.0±3.2 years) and 34 participants in Experiment 3 (27 females, age 20.0±1.1 years). All

96   participants were right-handed and had normal or corrected-to-normal vision.

97   For all experiments, participants for whom the minimum number of observations in one or more conditions

98   was lower than 10, were excluded from analysis. Additionally, for Experiment 3 (EEG), we had to exclude 2

99   participants due to problems with the reference electrodes. In the end, this resulted in the inclusion of 25

100 participants for Experiment 1 (20 females, age 19.5±1.3 years), 67 participants for Experiment 2 (49 females,

101 age 21.9±3.0 years) and 29 participants for Experiment 3 (22 females, age 20.0±1.1 years).

102   The studies were approved by the local ethics committee of the University of Amsterdam and written

103 informed consent was obtained from all participants according to the Declaration of Helsinki. Compensation

104 was 20 Euros for Experiment 1, 25 euros for Experiment 2 and 30 Euros for Experiment 3, or equivalents in

105 course credit.

106

107 *Materials*

108 All stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; RRID:SCR_002881) within a

109 MATLAB environment (MathWorks, Natick, MA, USA; RRID:SCR_001622). Stimuli were displayed on an ASUS

110 LCD monitor (1920 x 1080 pixels, 120Hz, 50.9x28.6 cm screen size, 46.3pixels/°) on a "black" (RGB: [0 0 0],

111 ±3cd/m2) background while participants were seated in a dimly lit room, approximately 70 cm away from the

112 screen.

113

114 *Procedure and Stimuli*

115 Participants performed an attentional blink (AB) task (Raymond et al., 1992), in which on every trial a rapid

116 series of visual stimuli was presented consisting of a sequence of 17 uppercase letters drawn from the alphabet

117 but excluding the letters I, L, O, Q, U, and V. Every letter appeared maximally once per trial. Letters were

118 presented at fixation in a mono-spaced font (font size: 40; corresponding to a height of approximately 1.2°) for

119 92 ms each.

120

121 Experiment 1

122 Participants were instructed to detect target letters within the rapid serial visual presentation (RSVP). The first

123 target (T1: G or H) was always presented at the fifth position of the RSVP. On most trials (80%) it was followed

124 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

125   equally likely. T1 was presented in green (RGB: [0 255 0]), while T2 and the distractor letters were white (RGB:

126   [255 255 255], ±320cd/m2).

127   Crucially, there was a predictive relationship between the two targets (Fig. 1A). Namely, in the 80% of trials

128   where a T2 was presented, the identity of T1 (e.g. G) predicted which T2 was likely (75%, e.g. D) or unlikely

129   (25%, e.g. K) to appear. On the 20% remaining trials without a T2 a random distractor letter was presented at

130   the T2-timepoint (every distractor letter was presented maximally once per trial) The mapping of T1 and T2

131   was counterbalanced over participants, so that for half of the participants the most likely target combinations

132   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

133   in the absence of a T2 stimulus, 4 grey squares (RGB: [200 200 200], ±188cd/m2; size: 0.35°; midpoint of each

134   square centered at 1.30° from fixation) were always presented around the stimulus (T2 or distractor) at the T2-

135   timepoint. Participants were instructed to use the timing information this cue provided when making decisions

136   about the presence of a T2 (only the letters "D" or "K", all other letters where distractors).

137   Following a 150 ms blank period at the end of the RSVP, participants gave their responses. First, they

138   indicated whether or not they had seen any T2 by pressing the left or right shift key on the keyboard. The

139   mapping between the keys and the response options was randomized per trial to decouple participants'

140   responses from the decision they had to make. Then they were asked to make a forced choice judgment about

141   the T2 letter (D or K) that was presented by typing in this letter. Finally, they made a similar response about the

142   identity of T1 (G or H). We used long response timeout durations of 5s and participants were instructed to

143   value accuracy over response speed. The inter-trial interval, as defined by the time between the last response

144   and the onset of the stream, was 500-750 ms.

145   The experiment consisted of two one-hour sessions on separate days within one week. In the first session,

146   participants received instructions about the task and subsequently performed the task for 6 blocks of 75 trials

147   (total 300 trials). The goal of the training session was to familiarize participants with the task. Besides, since we

148   did not instruct participants about the predictive relationship between T1 and T2, some practice on the task

149   was required for them to (implicitly) learn this relationship. In the second session, participants first received a

150   summary of the instructions, after which the actual experiment started. Participants performed 6 blocks of 90

151   trials (total of 540 trials) of the AB task. The first three participants performed 6 blocks of 105 trials (630 trials).

152   In both sessions participants received summary feedback about their performance at the end of each block,

153   followed by a short break.

154

155   Experiment 2 (EEG)

156   The task in the EEG experiment was the same as in Experiment 1, except that in Experiment 2, we only asked

157   participant to give one response by typing in the target letters they observed. In addition, we only used two

158   different lags: lag 3 (275 ms; 2/3rd of trials) and lag 10 (917 ms; 1/3rd of trials). To further increase the number

159   of trials, the ITI range was reduced to 200-400 ms.

160   Again, the experiment consisted of two different sessions within one week. The first session (1 hour)

161   consisted of instructions followed by extensive training (720 trials over 6 blocks) on the task. Participants were

162   not explicitly informed about the predictive relationship between the targets. In the second session (2 hours)

163   we first prepared the participant for the EEG measurements (see below) and gave brief instructions about the

164   task. Then, participants performed 12 blocks of 120 trials (total 1440 trials) of the AB task.

165

166   Experiment 3

167   To investigate the importance of T1 detection for expectation effects on conscious access, we adjusted the task

168   we used in Experiment 1 to decrease the visibility of T1 (Fig. 3A). We now presented T1 in white instead of

169   green, to make it stand out less among the other stimuli. Furthermore, T1 duration was staircased per

170   participant such that participants could report T1 on roughly 75% of the trials. Starting in the second half of the

171   training and continuing in the experimental session, after each block T1 duration was decreased by one frame

172   (8 ms) if performance was higher than 85% and increased by one frame if performance was lower than 65%. To

173   ensure T1 duration would not deviate too much from the duration of other stimuli, T1 duration was only

174   allowed to be in the range of 42-142 ms (max. 50 ms different from other stimuli). The median duration of T1 in

175   the second session was 125 ms. On 20% of trials no T1 was presented and a random distractor letter was

176 presented instead. When both targets were present, T1 predicted which T2 was likely to follow with 75%

177 accuracy.

178   We made a few changes to the task design to increase the efficiency of the design. The ITI was reduced to

179 values between 300-500 ms. In addition, we only asked participants for one response. They were asked to type

180 in any target letter they had seen during the trial and refrain from typing in a T1 and/or T2 letter when they did

181 not see any. The response was confirmed by pressing the space bar on the keyboard or when a timeout of 4s

182 had passed. To further increase the number of trials per condition, we decided to use only lag 3 (2/3rd of trials)

183 and lag 10 (1/3rd of trials). Because T1 duration was staircased on an individual basis, the T1-T2 SOA differed

184 between participants. On average, lag 3 corresponded to an SOA of 308 ms while lag 10 corresponded to an

185 SOA of 950 ms.

186   Finally, we manipulated the instructions we gave to participants in order to see to what extent explicit

187 knowledge of the relationship between T1 and T2 affected our results. As in Experiment 1, we tested

188 participants during two separate sessions within one week. The first group of the participants (N=25) did not

189 receive any explicit instruction about this relationship, similar to Experiment 1. The second group of

190 participants (N=19) received explicit instructions about the T1-T2 relationship at the start of the second

191 session, and a third group of participants (N=23) received those instructions already at the start of their first

192 session.

193   The first session (1 hour) was used for instructions and training the participants on the task (10 x 75 trials).

194 The experimental session in which participants performed the AB task lasted 1.5 hour and contained 16 blocks

195 of 75 trials (1200 trials).

196

197 *Experimental Design and Statistical Analysis: Behavioral*

198 Preparatory steps were done with in-house MATLAB scripts. Statistical analyses (repeated measures ANOVAs

199 and paired t-tests) were performed using JASP software (Love et al., 2015; RRID:SCR_015823). In situations

200 where a specifically tested hypothesis did not yield a significant result, we used a Bayesian equivalent of the

201 same test to quantify the evidence for the null-hypothesis (Rouder et al., 2012, 2017). In those cases, using

8

202 JASP's default Cauchy prior, Bayes Factors (BF) were computed for each effect. To increase the interpretability

203 in analyses with multiple factors, we used Bayesian model averaging to get a single BF for each effect in

204 ANOVA's. This BF is the change from prior to posterior inclusion odds, and can intuitively be understood as the

205 amount of evidence the data gives for including an experimental factor in a model of the data. The BF will

206 either converge to zero when the factor should not be included, or to infinity when it should be included in the

207 model. Values close to one indicate that there is not enough evidence for either conclusion. We use the

208 conventions from Jeffreys (1967) to interpret the effect sizes of our Bayesian analyses.

209

210 Experiment 1

211 In our behavioral analyses we looked at the T2 detection performance, given that T1 was correctly identified. A

212 response was considered to be correct when (1) the participant indicated no T2 was present when no T2 was

213 presented or (2) the participant correctly indicated a T2 was present and subsequently reported the correct

214 target letter. Since expectation is only a meaningful concept when a T2 target was presented, the T2 absent

215 trials, on which a distractor letter was presented instead of a T2, were not taken into consideration for the

216 main statistical analyses. Trials where one of the responses was missing were deleted from all analyses.

217 Percentage correct was used in a 3 x 2 repeated measures ANOVA with the factors lag (lag 2, lag 4, lag 10) and

218 expectation (valid, invalid). In a control analysis, we repeated our analyses for Experiment 1 based on the T2

219 detection responses (ignoring the accuracy of the T2 identification) as dependent variable (see also Results).

220 Since the seen/miss response is orthogonal to the specific expectations about target letters, this analysis rules

221 out simple response biases as a potential cause of our effects.

222

223 Experiment 2 (EEG)

224 The behavioral analyses for the EEG experiment were similar to those for Experiment 1. However, the factor lag

225 had only 2 levels (lag 3, lag 10). Percentage correct T2 detection was computed as in Experiment 1 using only

226 the trials on which T1 was correctly reported. A response was considered to be correct when the letter a

227 participant entered was the letter that was presented or when a participant refrained from entering a letter

9

228   when none was presented for the T2 absent trials. In addition, we computed a T2 detection measure to use in a

229   control analysis: if a participant typed in any letter, we categorize the response as a "target seen" response,

230   otherwise we call it a "target absent" response. This outcome measure was used in a control analysis.

231

232   Experiment 3

233   In this experiment, participants gave only one response by typing in the target they had perceived. Trials on

234   which no response was given or on which an impossible response was given (e.g. two T1 targets reported) were

235   excluded from analyses. For T1 and T2 separately, we assessed the accuracy of the responses. The definition of

236   correct and incorrect responses was the same as in Experiment 2 and we also used the same T2 detection

237   measure.

238   Subsequently, T2 percentage correct detection was used in a 2 x 2 x 2 x 3 mixed ANOVA with the within-

239   subject factors lag (lag 3, lag 10), expectation (valid, invalid) and T1 visibility (T1 seen, T1 missed) and the

240   between-subject factor instruction (none, start session 2, start session 1). As mentioned before, this between-

241   subject factor was included to find out whether predictive effects would be modulated by explicit knowledge of

242   the relation between T1 and T2. To investigate the effect of T1 visibility in more detail, we followed up the

243   main analyses by other mixed ANOVAs in which we first split up the dataset based on T1 visibility. In situations

244   where we found interactions with the factor instruction, we compared the effects of lag and expectation

245   separately per instruction condition using repeated measures ANOVAs and paired-samples t-tests.

246   Finally, to test for an interaction between expectation validity and lag, we combined behavioral data from

247   all experiments in a post-hoc analysis. Only trials on which T1 was correctly identified were used. For

248   Experiment 1 we averaged data for lag 2 and lag 4 to create an average "short lag" condition. Subsequently,

249   these data were entered into a 2 x 2 x 3 mixed ANOVA with the within-subject factors lag (short, long) and

250   expectation (valid, invalid) and the between-subject factor experiment (Experiment 1, Experiment 2,

251   Experiment 3).

252

253   *Electroencephalographic Measurements*

254   EEG was recorded with a BioSemi ActiveTwo system and sampled at 512 Hz (BioSemi, Amsterdam, The

255   Netherlands). Potentials were measured from sixty-four scalp electrodes, along with two reference electrodes

256   on the earlobes and four electrodes measuring horizontal and vertical eye movements. After data acquisition,

257   EEG data was pre-processed with the FieldTrip toolbox for MATLAB (Oostenveld et al., 2011;

258   RRID:SCR_004849). First, data were re-referenced to the linked earlobes, high-pass filtered at 0.01 Hz, and

259   epoched from -0.750 to 1s surrounding the onset of T2. Data were visually inspected and trials and/or channels

260   containing artifacts not related to eye blinks were manually removed, resulting in deletion of on average 9.1%

261   (±3.9%) of trials and 2.0 (±1.7) channels. Independent component analysis was used to identify components

262   related to eye blinks or other artifacts that could easily be distinguished from other EEG signals. After the

263   independent component analysis, previously deleted channels were reconstructed based on a nearest neighbor

264   approach. Trials were baseline corrected to the average amplitude prior to T1 onset (-0.750 to -0.275). As a

265   final step, we applied a 40Hz low-pass filter to the trial data, after which ERPs were created separately for each

266   condition of interest.

267

268   *Experimental Design and Statistical Analysis: Electroencephalography*

269   All EEG analyses are based exclusively on trials where T2 appeared at lag 3 and T1 was correctly identified. We

270   used a combination of Fieldtrip (Oostenveld et al., 2011) and in-house MATLAB scripts to perform our analyses.

271   As a first step, we performed cluster-based permutation tests (Maris and Oostenveld, 2007) on the time-

272   window 0-750 ms from stimulus onset to isolate significant ERP events relating to expectation validity (valid,

273   invalid; irrespective of T2 visibility) or T2 visibility (seen, missed; irrespective of validity) or the interaction

274   between these factors. Next, we used an in-house built MATLAB script to isolate the significant events as

275   clusters in time and space. For this purpose, we computed an average difference wave over all channels that

276   were part of the cluster at any point in time. Subsequently, the onset and offset of a cluster were defined as

277   the time period around the maximum difference where the difference did not drop below 50% of this

278   maximum and where at least one channel showed a significant effect. We then selected the 10 channels that

279   showed the largest effect in this time-window. One of the observed events reflected a mixture of the

280 traditionally observed P3a and P3b components (Sergent et al., 2005; Volpe et al., 2007). Therefore, we split

281 the event into two clusters by manually selecting either the 32 most anterior or 32 most posterior EEG

282 channels (from the central midline) before running the cluster selection procedure.

283   As an alternative way to establish potential interactions between T2 detection and validity, we inspected

284 the clusters that were isolated in the previous step in more detail. This may be a more powerful (but also less

285 sensitive) way to detect small effects, because data is averaged over more time-points and channels. Within

286 each of the clusters, we performed a 2 x 2 repeated measures ANOVA (and its Bayesian equivalent, see also

287 Behavioral Analysis) with the factors T2 detection (seen, missed) and expectation validity (valid, invalid) on the

288 cluster data averaged over channels and time. To prevent double dipping, in each analysis we only considered

289 the effects orthogonal to the one that was used to define the cluster (e.g. not testing the effect of expectation

290 in a cluster defined based on the expectation effect).

291

292 {Insert Figure 1 about here}

293

294 **Results**

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

296 In the first experiment we addressed the question if expectations about the likelihood of stimulus identity

297 modulate the likelihood of conscious access, and if so, in what direction. To do so, we used the attentional blink

298 paradigm (Raymond et al., 1992). The attentional blink is an impairment in the conscious perception of the

299 second of two target stimuli that are presented in rapid succession when the initial target was correctly

300 perceived Here we modified the paradigm in such a way that the first target (T1: the letter G or H, in green)

301 predicted which of the second targets would most be likely to appear in case a T2 target was presented (T2: the

302 letter D or K, predicted=75%, unpredicted=25%, in white, Fig. 1A). On 20% of trials we presented a random

303 distractor letter instead of a T2 target. At the end of each stream of letters, participants gave three responses.

304 First, they indicated whether or not they had seen any of the two T2 targets ("seen"/"unseen" response).

305 Second, they were prompted to make a forced-choice judgment about the identity of T2 (whether the letter D

306   or K was presented). Third, participants had to make a similar forced-choice decision about the identity of T1

307   (whether the letter G or H was presented) (see Methods for details). Participants were not explicitly instructed

308   about the predictive relationship between T1 and T2.

309   In Figure 1 we plot the percentage of trials in which T2 was correctly detected and T1 discrimination was

310   also correct (average T1 accuracy was 94.20%, sd=5.77%) for the three different lags (lag 2, 4 and 10). T2 was

311   considered to be detected correctly when participants indicated to have seen it (based on the first response)

312   and correctly identified it (based on the second response). Overall, there was a clear attentional blink, as

313   reflected by reduced T2 detection when the time (i.e. lag) between T1 and T2 was shorter (Figure 1B, main

314   effect of lag: $F_{2,48}$=48.15, p<0.001). Importantly, expectations modulated T2 detection rate. T2 detection was

315   significantly better when T1 validly predicted T2 (black lines) compared to when the expectation was invalid

316   (gray lines, main effect of validity: $F_{1,24}$=7.10, p=0.014, no significant interaction between lag and validity:

317    $F_{2,48}$=1.30, p=0.283). These results extend several previous studies (Chang et al., 2015; Melloni et al., 2011;

318   Pinto et al., 2015; Stein et al., 2015; Stein & Peelen, 2015) by showing that conscious perception is (partly)

319   determined by the transitional probability of the input the brain receives.

320   While these data support the notion that valid expectations trigger access to consciousness, it has been

321   recognized that such findings may not solely be due to changes in perception, but perhaps (also) due to

322   changes in decision criteria or response biases (Gayet et al., 2014b; Yang et al., 2014; Attarha and Moore,

323   2015). To rule out the possibility that our effects could be explained by a response bias in which people simply

324   report the target letter that they expected based on T1, irrespective of whether they consciously perceived T2,

325   we performed an analysis with T2 detection (instead of T2 discrimination, see Methods) as the dependent

326   variable. This analysis takes into account only participants' first response (the "seen"/"unseen" response),

327   regardless of whether subsequent T2 letter identification was correct or not. Crucially, this analysis cannot be

328   influenced by any decision/response biases because the response was orthogonal to the participants'

329   expectation. Information about the most likely letter to appear cannot predispose participants to better

330   determine whether a target letter was presented at all. Still, we observed a qualitatively similar pattern of

331   results (main effect of validity: $F_{1,24}$=5.47, p=0.028). This finding suggests that validity indeed boosted

332   participants conscious access of T2, instead of merely eliciting a shift in response bias.

333

334   **Experiment 2: EEG markers of conscious and nonconscious expectation violations**

335   Subsequently, we tested whether expectation violations can be elicited by nonconsciously processed

336   unpredicted stimuli or whether conscious perception of a stimulus is a prerequisite for it to trigger neural

337   expectation error responses. To test this, we measured subjects' brain activity with EEG while they performed a

338   similar task as in Experiment 1. First, we replicated the behavioral effects of Experiment 1 (Fig. 1C). Overall, T1

339   performance was high (M=93.61%, sd=7.31%) and T2 detection was higher at lag 10 than at lag 3 (main effect

340   of lag: $F_{1,28}$=128.72, p<0.001), reflecting a robust attentional blink. More importantly, validly predicted T2s were

341   discriminated better than invalidly predicted T2s (main effect of validity: $F_{1,28}$=9.49, p=0.005). The effects were

342   similar in a control analysis where we considered the percentage of T2 seen responses (regardless of the exact

343   letter participants entered), making it less likely that our effect can be explained by a response bias (main

344   effect of validity: $F_{1,28}$=4.23, p=0.049). In this experiment, the validity effect was significantly modulated by lag

345   (validity x lag: $F_{1,28}$=5.86, p=0.022), an effect that was numerically similar, but not significant in Experiment 1.

346   Participants performed better for valid than invalid trials at lag 3, but there was no convincing evidence for an

347   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).

348   Thus, effects of expectations were larger in the time window in which T2 more often goes unperceived.

349

350   {Insert Figure 2 here}

351

352   Next, we investigated potential differences in the neural processing of predicted and unpredicted stimuli, as a

353   function of stimulus awareness. To this end, we contrasted invalidly and validly predicted T2s and tested this

354   difference using cluster-based permutation testing, correcting for multiple comparisons across both time (0-

355   750 ms) and (electrode) space (see Fig. 2 and Methods) (Maris and Oostenveld, 2007). We found one

356   significant difference over fronto-central electrode channels, which reflected greater T2-elicited negativity for

14

357   invalid compared to valid trials between 174-314 ms (p=0.015, Fig. 2B), therefore potentially reflecting some

358   type of mismatch response. We then further analyzed this event to test whether the difference was modulated

359   by, or dependent on, conscious perception of T2. Crucially, the size of this fronto-central mismatch component

360   was independent of T2 awareness ($F_{1,28}$=0.04, p=.850, BF=0.254, Fig. 2C), indicating that both seen and unseen

361   T2's generated a fronto-central mismatch response.

362   Additionally, analyses of T2 visibility effects (irrespective of expectation validity) replicated previously

363   reported findings (Kranczioch et al., 2003; Sergent et al., 2005; Harris et al., 2013). In these analyses, we

364   examined the difference in ERPs following seen and missed T2s using a cluster-based permutation test (Fig. 3),

365   revealing two significant events. First, a significant negative difference could be observed over (left) posterior

366   electrodes from 170-355 ms after T2 onset (p=0.010; Fig 3A). This event was followed by a significant long-

367   lasting positive event (p<0.001), reflecting a mixture of the P3a and P3b components, extending over frontal

368   and central electrodes.

369

370   {Insert Figure 3 here}

371

372   Subsequently, we had a closer look at the interactions between conscious access and expectation validity.

373   Therefore, we analyzed the ERP events that were isolated in the previous step in more detail (Fig. 3B-G). For

374   this analysis we first isolated the traditionally observed AB-related P3a and P3b ERP components from the long-

375   lasting positive ERP event that differentiated between seen and missed T2s (Sergent et al., 2005). Doing so

376   resulted in an early positive P3a cluster (Fig. 3D) over fronto-central channels that was significant between 395-

377   586 ms and a somewhat later positive P3b cluster (Fig. 3F) over more posterior parietal channels, which was

378   significant between 445-611 ms. Within each of these clusters we performed repeated measures ANOVAs with

379   the factors validity and T2 detection. For none of the events we found evidence that the T2 detection effect

380   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,

381   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

382   absence of such interactions is moderate at best. This is somewhat surprising, because especially the late

383 positive events have previously been related to conscious access (Sergent et al., 2005; Rutiku et al., 2015) and

384 metacognition (Desender et al., 2016). However, recent investigations show it may also reflect even later more

385 cognitive processes, merely arising as a consequence of becoming consciously aware of information (Pitts et al.,

386 2014; Silverstein et al., 2015). We did not find evidence that the amplitude of these ERP events was modulated

387 by expectation validity, which may suggest that once a stimulus has been perceived consciously, it is irrelevant

388 whether or not the expectation was valid.

389 Finally, we directly tested for an interaction between conscious access and expectation by comparing the

390 validity ERP effect (invalid-valid) for T2 seen and T2 missed trials in a cluster-based permutation test (this

391 analysis takes into account the entire scalp topography). Again, no significant interactions between these

392 factors were observed (all clusters p>0.10).

393

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

395 In our final experiment, we addressed the question whether expectation formation itself can unfold in the

396 absence of awareness and subsequently influence conscious access (Fig. 4). To address this question, we

397 changed the color of T1 from green to white and for each subject staircased T1 duration in such a way that T1

398 was correctly identified on approximately 75% of the trials (actual T1 identification performance: M=76.03%;

399 sd=8.65%). T1 duration did not differ between trials where T2 was seen and trials where T2 was missed (T2

400 detection: t_{66}=0.31, p=0.752; T2 seen: M=117.42 ms; T2 missed: M=117.46 ms), which indicates that T1

401 visibility was not determined by stimulus duration. Likely, internal fluctuations in the system (e.g. variability in

402 attention) must be causing participants to sometimes see T1 and sometimes miss it. Moreover, on 20% of trials

403 no T1 was presented (but replaced by a distractor). Further, to test to what extent explicit knowledge of the

404 predictive relationships between stimuli would increase the validity effects, we varied the moment at which

405 explicit information about the predictive relations between T1 and T2 was provided. The experiment consisted

406 of a training session and a test session on separate days. A first group of subjects received no explicit

407 instructions about the predictive relations in either session and had to learn them implicitly through experience

408 with the task; the second group received explicit instructions about the T1-T2 relations in the test session only,

409   but not in the first training session; and the third group received explicit instructions already from the start of

410   the experiment.

411

412   {Insert Figure 4 here}

413

414   T1 visibility strongly affected T2 detection. When T1 was seen, T2 detection was markedly lower than when T1

415   was missed (main effect of T1 awareness: $F_{1,64}$=4.62, p=0.035), in particular at short lags (T1 awareness x lag:

416    $F_{1,64}$=72.95, p<0.001). Validly predicted targets were detected more often (main effect of validity: $F_{1,64}$=33.39,

417   p<0.001). The effect of expectation validity on T2 detection varied as a function of T1 awareness and

418   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

419   awareness x validity x instruction: $F_{2,64}$=11.33, p<0.001). When T1 was seen (Fig. 4B), a clear attentional blink

420   was observed (main effect of lag: $F_{1,64}$=170.01, p<0.001) and validly predicted targets were more often

421   detected than invalidly predicted targets (main effect of validity: $F_{1,64}$=64.97, p<0.001) (as in Experiment 1 and

422   2). Like in the previous experiments, a control analysis considering only the percentage of T2 seen responses

423   (regardless of the exact letter participants entered) also revealed a significant effect of validity (main effect of

424   validity: $F_{1,64}$=65.83, p<0.001), making it unlikely that response biases are causing the effect. Interestingly, we

425   also observed a significant attentional blink for missed T1's, reflecting a nonconsciously elicited AB (main effect

426   of lag: $F_{1,64}$=74.42, p<0.001). This AB effect cannot be explained by an overall T2 detection performance benefit

427   for targets that are presented later in the trial because the AB was larger for trials on which T1 was presented

428   but missed compared to trials on which no T1 was presented in the trial (lag x T1 presence: $F_{1,66}$=24.19,

429   p<0.001). However, although missed T1's triggered an AB, expectation validity did not affect T2 detection

430   performance for missed T1's (main effect of validity: $F_{1,64}$=0.35, p=0.554), regardless of the type of instruction

431   participants received about the predictive relation between T1 and T2 (validity x instruction: $F_{2,64}$=0.64,

432   p=0.533). A Bayesian equivalent of the repeated-measures analysis strongly suggested validity should not be

433   included in a model of the data (BF=.084).

434   The above results highlight that only when T1 was seen, valid expectations facilitated T2 detection. A post-

435   hoc analysis on T1-seen trials only revealed that this effect was modulated by how explicitly we instructed

436   participants about the predictive relationship between T1 and T2 (validity x instruction: $F_{2,64}$=14.83, p<0.001).

437   The validity effect, as defined by the difference between valid and invalid trials, averaged across the two lags,

438   increased with more explicit instructions (group 1: 1.87%, group 2: 19.53%, group 3: 26.27%). These results

439   indicate that, not only does the visibility of T1 define the predictive impact on T2 detection, but also the extent

440   to which these predictive relations are (explicitly) known affects the impact of expectations on conscious

441   access. This may also explain why the validity effect appeared more pronounced in Experiment 3 compared to

442   Experiments 1 and 2, because in experiment 1 an d 2 subjects were not explicitly instructed about the

443   predictive relations between T1 and T2.

444   Finally, in contrast to Experiment 2, on T1 seen trials the validity effect was independent of lag (validity x

445   lag: $F_{1,64}$=1.750, p=0.191). Since we anticipated stronger expectation effects at short lags, behavioral data from

446   all three experiments was combined in a post-hoc analysis. Only trials on which T1 was correctly identified

447   were used and for Experiment 1 we averaged data for lag 2 and lag 4 to create an average "short lag"

448   condition. A significant interaction between validity and lag showed that across all experiments, the

449   expectation effect was stronger at short lags compared to the long lags (validity x lag: $F_{1,118}$=5.73, p=0.018; no

450   validity x lag x experiment interaction: $F_{2,118}$=0.065, p=0.937).

451

452   **Discussion**

453   In this report we investigated three important questions regarding the intricate relationship between top-down

454   expectations and conscious awareness. The first question that we addressed was how prior information about

455   the identity of an upcoming stimulus influences the likelihood of that stimulus entering conscious awareness.

456   Using a novel attentional blink paradigm in which the identity of T1 cued the likelihood of the identity of T2, we

457   showed that stimuli that confirm our expectation have a higher likelihood of gaining access to conscious

458   awareness than stimuli that violate our expectations, especially at short lags. The expectation effect was

459   qualitatively similar across all three experiments, regardless of subtle experimental differences in task design

460 and overall performance between those experiments. Furthermore, it could not be explained by simple shifts in

461 response criterion, because it was also present for a dependent measure orthogonal to the expectation

462 manipulation. Together, this suggests that valid expectations amplify the perceptual strength of a stimulus and

463 therefore increase the chance of conscious access, possibly due to the sharpening of its neural representations

464 (Kok et al., 2012). This interpretation is supported by previous experiments that have shown varying effects of

465 expectations on (subjective) perception, such as studies showing prior knowledge increases the speed (Chang

466 et al., 2015; De Loof et al., 2016; Melloni et al., 2011; Pinto et al., 2015) and accuracy (Stein and Peelen, 2015)

467 of stimulus detection. Furthermore, our findings complement recent studies showing that the AB can be

468 reduced when there is knowledge about temporal statistics of the task (Lasaponara et al., 2015; Visser et al.,

469 2015) or when the latency of T2 targets is explicitly cued (Martens & Johnson, 2005; Nieuwenstein, Chun, van

470 der Lubbe, & Hooge, 2005). In addition, two recent reports have shown that, in some subjects, expecting the

471 presentation of a stimulus can even elicit an illusory stimulus percept when no stimulus is presented, when

472 attention is diverted (Mack et al., 2016; Aru and Bachmann, 2017). Future experimentation is required to shed

473 light on the generalizability of our effect to simpler tasks. Such experiments may also consider using other

474 measures of subjective perception (e.g. perceptual awareness scale) (Overgaard et al., 2006).

475   The second question that we addressed was related to the extent to which nonconscious stimuli can trigger

476 prediction error responses, as measured with EEG. Over the last 20 years, we and others have shown that

477 nonconscious information processing is rather sophisticated (Dehaene and Changeux, 2011; van Gaal and

478 Lamme, 2012), and that a diverse range of high-level cognitive processes can unfold nonconsciously (Dehaene

479 et al., 2001; Custers and Aarts, 2005; Lau and Passingham, 2007; Pessiglione et al., 2007, 2008, van Gaal et al.,

480 2010, 2012, 2014). Interestingly, in Experiment 2 we found that expectations that are violated by a

481 nonconscious stimulus trigger a stronger negative fronto-central ERP component than expectations that are

482 confirmed. This neural event was similar for trials on which T2 was seen and on trials where T2 was missed,

483 highlighting that conscious awareness of a stimulus is not a prerequisite for it to trigger a prediction error

484 response (Mathews et al., 2014; Malekshahi et al., 2016). This effect may reflect a mismatch signal, similar to

485 the mismatch negativity (Näätänen et al., 2007), which is a negative deflection following oddball stimuli that

486 develops 100-200 milliseconds after stimulus onset. Sometimes this effect lasts longer, in some experiments

487 until ~400 milliseconds, depending on the specifics of the task and stimulus material (Pöppel, 2004; Stefanics et

488 al., 2011; Kimura and Takeda, 2015). While in terms of interpretation this effect is similar to a mismatch effect,

489 its topography is slightly different than a typical visually evoked MMN, which generally peaks more posteriorly,

490 although considerable variation in its topography has been reported (Pöppel, 2004). Alternatively, it is possible

491 that the higher activation for valid compared to invalid trials corresponds to the frontal selection positivity

492 (FSP), which is a well-known marker of non-spatial attentional processes (Kenemans et al., 1993). In our

493 paradigm, this could be explained as improved attentional selection when expectations are confirmed.

494 Although the exact nature of the observed component deserves future experimentation, the key finding is that

495 the effect was independent of T2 perception and purely depends on the validity of the expectation. This is in

496 line with studies that have shown context influences on nonconscious information processing (Nakamura et al.,

497 2007; Van Opstal et al., 2011; Gayet et al., 2014a; Rohaut et al., 2016), studies that have shown that the MMN

498 can be observed when the expectation violations are unattended (Bekinschtein et al., 2009; Stefanics et al.,

499 2011; King et al., 2013; Dykstra and Gutschalk, 2015; Kimura and Takeda, 2015) and more generally evidence

500 for relatively high-level processing of nonconscious stimuli (Luck et al., 1996; van Gaal and Lamme, 2012;

501 Silverstein et al., 2015). Nevertheless, the absence of interactions in the ERP is also somewhat surprising (but

502 see also Rutiku et al., 2016), because as noted earlier such interactions between expectation validity and

503 conscious T2 detection were present in behavior. A neural basis for this effect should exist, but may be very

504 subtle. Recently, a study by Aru et al. (2016) found early (<100 ms) differences in signal amplitude over

505 posterior channels that predicted the behavioral benefit of prior knowledge on the detection of stimuli

506 presented at the threshold of perception. Another potentially interesting signature to investigate could be the

507 onset of components related to conscious perception (Melloni et al., 2011) and how they relate to

508 expectations. Moreover, it is possible that instead of signal strength, it is the signal-to-noise ratio or sharpness

509 of the representation that is improved (Kok et al., 2012). Possibly, valid expectations do not modulate the

510 amplitude of the neural response, but increase the specificity of the neural representation.

511   In the final experiment, we showed that conscious perception of T1, initiating the expectation, is a

512   prerequisite for influences on conscious access to occur. On the subset of trials where subjects did not see T1,

513   there was no expectation effect T1 on T2 detection performance. This result contrasts with findings from a

514   recent study that suggested that some priors may operate nonconsciously (Chang, Baria, Flounders, & He,

515   2016). Chang and colleagues presented participants with masked grey-scale natural scene images and found

516   that the nonconscious processing of these images improved subsequent recognition of their degraded

517   counterparts, so-called "Mooney images", presented seconds later. One explanation for this difference is that

518   the priors on which the effects of Chang et al. relied may be more automatic and hard-wired than the relatively

519   arbitrary relationships that people have to learn and actively use in our experiments. It is possible that lower-

520   level, automatic expectations are more easily processed outside of awareness compared to the more active

521   ones studied here.

522   Further, it is also possible that with more training we would find nonconscious expectation effects.

523   However, since subjects were already trained on the task on a separate day before performing the

524   experimental session, this possibility seems unlikely. We did observe greater validity effects when subjects

525   were made explicitly aware of the predictive nature of T1, suggesting that explicit knowledge of stimulus

526   associations can facilitate the effects of stimulus-induced expectations. Finally, it should be noted that we did

527   not test the full range of timing intervals between T1 and T2. It has been shown and proposed that the

528   processing of nonconscious stimuli is relatively fleeting (Greenwald et al., 1996; Dehaene et al., 2006; but see

529   King et al., 2016), so it is conceivable that the T1-T2 lags that we have used here may have been too long to

530   observe expectation effects triggered by unseen T1's. Further, a significant attentional blink was observed on

531   trials on which T1 was missed, indicating that attention was still captured by a missed T1 at the T1-T2 lags used

532   here. This latter result is in line with evidence showing that nonconscious stimuli are able to trigger attentional

533   capture (Ansorge et al., 2009; Mulckhuyse and Theeuwes, 2010; Hsieh et al., 2011) and with a study showing

534   lower T2 detection for T1's that were missed compared to trials without a T1 (in that experiment this effect

535   was independent of lag (Nieuwenstein, Van der Burg, Theeuwes, Wyble, & Potter, 2009)).

536   In summary, three main conclusions can be drawn from the present series of studies. First, expectation

537 confirmation, compared to violation, increases the likelihood of conscious awareness, suggesting that valid

538 expectations amplify the perceptual strength of a stimulus. Second, nonconscious violations of conscious

539 expectations are registered in the human brain Third, however, expectations need to be implemented

540 consciously to subsequently modulate conscious access. These results suggest a differential role of conscious

541 awareness in the hierarchy of predictive processing, in which the active implementation of top-down

542 expectations requires conscious awareness, whereas a conscious expectation and a nonconscious stimulus can

543 interact to generate prediction errors. How these nonconscious prediction errors are used for updating future

544 behavior and shaping trial-by-trial learning is a matter for future experimentation.

545

546 **References**

547 Ansorge U, Kiss M, Eimer M (2009) Goal-driven attentional capture by invisible colors: Evidence from event-

548   related potentials. Psychon Bull Rev 16:648–653.

549 Aru J, Bachmann T (2017) Expectation creates something out of nothing: The role of attention in iconic memory

550   reconsidered. Conscious Cogn 53:203–210.

551 Aru J, Rutiku R, Wibral M, Singer W, Melloni L (2016) Early effects of previous experience on conscious

552   perception. Neurosci Conscious:niw004.

553 Attarha M, Moore CM (2015) Onset rivalry: factors that succeed and fail to bias selection. Attention,

554   Perception, Psychophys 77:520–535.

555 Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629.

556 Bar M (2009) Predictions: a universal principle in the operation of the human brain. Philos Trans R Soc B Biol Sci

557   364:1181–1182.

558 Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hamalainen MS, Marinkovic K, Schacter DL,

559   Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103:449–

560   454.

561 Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious

562   processing of auditory regularities. Proc Natl Acad Sci U S A 106:1672–1677.

563   Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

564   Chang AYC, Kanai R, Seth AK (2015) Cross-modal prediction changes the timing of conscious access during the

565   motion-induced blindness. Conscious Cogn 31:139–147.

566   Chang R, Baria AT, Flounders MW, He BJ (2016) Unconsciously elicited perceptual prior. Neurosci Conscious

567   2016:niw008.

568   Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav

569   Brain Sci 36:181–204.

570   Custers R, Aarts H (2005) Positive Affect as Implicit Motivator: On the Nonconscious Operation of Behavioral

571   Goals. J Pers Soc Psychol 89:129–142.

572   De Loof E, Van Opstal F, Verguts T (2016) Predictive information speeds up visual awareness in an individuation

573   task by modulating threshold setting, not processing efficiency. Vision Res 121:104–112.

574   Dehaene S, Changeux J-P (2011) Experimental and Theoretical Approaches to Conscious Processing. Neuron

575   70:200–227.

576   Dehaene S, Changeux J-P, Naccache L, Sackur J, Sergent C (2006) Conscious, preconscious, and subliminal

577   processing: a testable taxonomy. Trends Cogn Sci 10:204–211.

578   Dehaene S, Naccache L, Cohen L, Bihan DL, Mangin JF, Poline JB, Rivière D (2001) Cerebral mechanisms of word

579   masking and unconscious repetition priming. Nat Neurosci 4:752–758.

580   Denison R, Piazza E, Silver M (2011) Predictive context biases perceptual selection during binocular rivalry. J Vis

581   11:312–312.

582   Desender K, Van Opstal F, Hughes G, Van den Bussche E (2016) The temporal dynamics of metacognition:

583   Dissociating task-related activity from later metacognitive processes. Neuropsychologia 82:54–64.

584   Dykstra AR, Gutschalk A (2015) Does the mismatch negativity operate on a consciously accessible memory

585   trace? Sci Adv 1:e1500677.

586   Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815–836.

587   Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying

588   mechanisms. Clin Neurophysiol 120:453–463.

589   Gayet S, Van der Stigchel S, Paffen CLE (2014a) Seeing is believing: Utilization of subliminal symbols requires a

590   visible relevant context. Attention, Perception, Psychophys 76:489–507.

591   Gayet S, Van Der Stigchel S, Paffen CLE (2014b) Breaking continuous flash suppression: Competing for

592   consciousness on the pre-semantic battlefield. Front Psychol 5:1–10.

593   Greenwald AG, Draine SC, Abrams RL (1996) Three cognitive markers for unconscious semantic activation.

594   Science (80-) 173:1699–1702.

595   Gregory RL (1980) Perceptions as Hypotheses. Philos Trans R Soc B Biol Sci 290:181–197.

596   Harris JA, McMahon AR, Woldorff MG (2013) Disruption of Visual Awareness during the Attentional Blink Is

597   Reflected by Selective Disruption of Late-stage Neural Processing. J Cogn Neurosci 25:1863–1874.

598   Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. Front Psychol 3:96.

599   Hsieh P-J, Colas JT, Kanwisher N (2011) Pop-Out Without Awareness. Psychol Sci 22:1220–1226.

600   Jeffreys H (1967) Theory of Probability, 3rd ed. Oxford: Oxford University Press.

601   Kenemans JL, Kok A, Smulders FTY (1993) Event-related potentials to conjunctions of spatial frequency and

602   orientation as a function of stimulus parameters and response requirements. Electroencephalogr Clin

603   Neurophysiol 88:51–63.

604   Kimura M, Takeda Y (2015) Automatic prediction regarding the next state of a visual object:

605   Electrophysiological indicators of prediction match and mismatch. Brain Res 1626:31–44.

606   King J-R, Faugeras F, Gramfort A, Schurger A, El Karoui I, Sitt JD, Rohaut B, Wacongne C, Labyt E, Bekinschtein T,

607   Cohen L, Naccache L, Dehaene S (2013) Single-trial decoding of auditory novelty responses facilitates the

608   detection of residual consciousness. Neuroimage 83:726–738.

609   King J-R, Pescetelli N, Dehaene S (2016) Brain Mechanisms Underlying the Brief Maintenance of Seen and

610   Unseen Sensory Information. Neuron 92:1122–1134.

611   Koelsch S, Heinke W, Sammler D, Olthoff D (2006) Auditory processing during deep propofol sedation and

612   recovery from unconsciousness. Clin Neurophysiol 117:1746–1759.

613   Kok P, Jehee JFM, de Lange FP (2012) Less Is More: Expectation Sharpens Representations in the Primary Visual

614   Cortex. Neuron 75:265–270.

615   Kranczioch C, Debener S, Engel AK (2003) Event-related potential correlates of the attentional blink

616   phenomenon. Cogn Brain Res 17:177–187.

617   Lamme VAF, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent

618   processing. Trends Neurosci 23:571–579.

619   Lasaponara S, Dragone A, Lecce F, Di Russo F, Doricchi F (2015) The "serendipitous brain": Low expectancy and

620   timing uncertainty of conscious events improve awareness of unconscious ones (evidence from the

621   Attentional Blink). Cortex 71:15–33.

622   Lau HC, Passingham RE (2007) Unconscious Activation of the Cognitive Control System in the Human Prefrontal

623   Cortex. J Neurosci 27:5805–5811.

624   Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen AJ, Ly A, Gronau QF, Smira M, Epskamp S, Matzke

625   D, Wild A, Knight P, Rouder JN, Morey RD, Wagenmakers E-J (2015) JASP.

626   Luck SJ, Vogel EK, Shapiro KL (1996) Word meanings can be accessed but not reported during the attentional

627   blink. Nature 383:616–618.

628   Mack A, Erol M, Clarke J, Bert J (2016) No iconic memory without attention. Conscious Cogn 40:1–8.

629   Malekshahi R, Seth A, Papanikolaou A, Mathews Z, Birbaumer N, Verschure PFMJ, Caria A (2016) Differential

630   neural mechanisms for early and late prediction error detection. Sci Rep 6:24350.

631   Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods

632   164:177–190.

633   Martens S, Johnson A (2005) Timing attention: Cuing target onset interval attenuates the attentional blink.

634   Mem Cognit 33:234–240.

635   Mathews Z, Cetnarski R, Verschure PFMJMJ (2014) Visual anticipation biases conscious decision making but not

636   bottom-up visual processing. Front Psychol 5:1–14.

637   Melloni L, Schwiedrzik CM, Muller N, Rodriguez E, Singer W (2011) Expectations Change the Signatures and

638   Timing of Electrophysiological Correlates of Perceptual Awareness. J Neurosci 31:1386–1396.

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

640   literature. Acta Psychol (Amst) 134:299–309.

641 Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central

642   auditory processing: A review. Clin Neurophysiol 118:2544–2590.

643 Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) "Primitive intelligence" in the auditory

644   cortex. Trends Neurosci 24:283–288.

645 Nakamura K, Dehaene S, Jobert A, Le Bihan D, Kouider S (2007) Task-specific change of unconscious neural

646   priming in the cerebral language network. Proc Natl Acad Sci U S A 104:19643–19648.

647 Nieuwenstein M, Van der Burg E, Theeuwes J, Wyble B, Potter M (2009) Temporal constraints on conscious

648   vision: On the ubiquitous nature of the attentional blink. J Vis 9:18–18.

649 Nieuwenstein MR, Chun MM, van der Lubbe RHJ, Hooge ITC (2005) Delayed Attentional Engagement in the

650   Attentional Blink. J Exp Psychol Hum Percept Perform 31:1463–1475.

651 Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of

652   MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011.

653 Overgaard M, Rote J, Mouridsen K, Ramsøy TZ (2006) Is conscious perception gradual or dichotomous? A

654   comparison of report methodologies during a visual task. Conscious Cogn 15:700–708.

655 Panichello MF, Cheung OS, Bar M (2013) Predictive feedback and conscious visual experience. Front Psychol

656   3:620.

657 Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD (2008) Subliminal Instrumental

658   Conditioning Demonstrated in the Human Brain. Neuron 59:561–567.

659 Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau H, Dolan RJ, Frith CD (2007) How the Brain Translates

660   Money into Force: A Neuroimaging Study of Subliminal Motivation. Science (80-) 316:904–906.

661 Pinto Y, Gaal V, de Lange FP, Lamme VAF, Seth AK (2015) Expectations accelerate entry of visual stimuli into

662   awareness. J Vis 15:1–15.

663 Pitts MA, Padwal J, Fennelly D, Martínez A, Hillyard SA (2014) Gamma band activity and the P3 reflect post-

664   perceptual processes, not visual awareness. Neuroimage 101:337–350.

665 Pöppel E (2004) Lost in time: A historical frame, elementary processing units and the 3-second window. Acta

666   Neurobiol Exp (Wars) 64:295–301.

667 Ratcliff R, McKoon G (2008) The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks.

668   Neural Comput 20:873–922.

669 Raymond JE, Shapiro KL, Arnell KM (1992) Temporary suppression of visual processing in an RSVP task: An

670   attentional blink? J Exp Psychol Hum Percept Perform 18:849–860.

671 Rohaut B, Alario F-X, Meadow J, Cohen L, Naccache L (2016) Unconscious semantic processing of polysemous

672   words is not automatic. Neurosci Conscious 2016:niw010.

673 Rouder JN, Morey RD, Speckman PL, Province JM (2012) Default Bayes factors for ANOVA designs. J Math

674   Psychol 56:356–374.

675 Rouder JN, Morey RD, Verhagen J, Swagman AR, Wagenmakers E-J (2017) Bayesian analysis of factorial designs.

676   Psychol Methods 22:304–321.

677 Ruby P, Caclin A, Boulet S, Delpuech C, Morlet D (2008) Odd sound processing in the sleeping brain. J Cogn

678   Neurosci 20:296–311.

679 Rutiku R, Martin M, Bachmann T, Aru J (2015) Does the P300 reflect conscious perception or its consequences?

680   Neuroscience 298:180–189.

681 Rutiku R, Tulver K, Aru J, Bachmann T (2016) Visual masking with frontally applied pre-stimulus TMS and its

682   subject-specific neural correlates. Brain Res 1642:136–145.

683 Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the

684   attentional blink. Nat Neurosci 8:1391–1400.

685 Silverstein BH, Snodgrass M, Shevrin H, Kushwaha R (2015) P3b, consciousness, and complex unconscious

686   processing. Cortex 73:216–227.

687 Stefanics G, Kimura M, Czigler I (2011) Visual mismatch negativity reveals automatic detection of sequential

688   regularity violation. Front Hum Neurosci 5:46.

689 Stefanics G, Kremláček J, Czigler I (2016) Mismatch negativity and neural adaptation: Two sides of the same

690   coin. Response: Commentary: Visual mismatch negativity: a predictive coding view. Front Hum Neurosci

691   10:925–939.

692   Stein T, Kaiser D, Peelen M V. (2015) Interobject grouping facilitates visual awareness. J Vis 15:10.

693   Stein T, Peelen M V. (2015) Content-specific expectations enhance stimulus detectability by increasing

694   perceptual sensitivity. J Exp Psychol Gen 144:1089–1104.

695   Strauss M, Sitt JD, King J-R, Elbaz M, Azizi L, Buiatti M, Naccache L, van Wassenhove V, Dehaene S (2015)

696   Disruption of hierarchical predictive coding during sleep. Proc Natl Acad Sci U S A 112:E1353–E1362.

697   Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational

698   mechanisms. Nat Rev Neurosci 15:745–756.

699   van Gaal S, de Lange FP, Cohen MX (2012) The role of consciousness in cognitive control and decision making.

700   Front Hum Neurosci 6:121.

701   van Gaal S, Lamme VAF (2012) Unconscious High-Level Information Processing. Neurosci 18:287–301.

702   van Gaal S, Lamme VAF, Ridderinkhof KR (2010) Unconsciously triggered conflict adaptation. PLoS One 5:2–7.

703   van Gaal S, Naccache L, Meuwese JDI, van Loon AM, Leighton AH, Cohen L, Dehaene S (2014) Can the meaning

704   of multiple words be integrated unconsciously? Philos Trans R Soc B Biol Sci 369:20130212–20130212.

705   Van Opstal F, Calderon CB, Gevers W, Verguts T (2011) Setting the stage subliminally: Unconscious context

706   effects. Conscious Cogn 20:1860–1864.

707   Visser TW, Ohan JL, Enns JT (2015) Temporal cues derived from statistical patterns can overcome resource

708   limitations in the attentional blink. Attention, Perception, Psychophys 77:1585–1595.

709   Volpe U, Mucci A, Bucci P, Merlotti E, Galderisi S, Maj M (2007) The cortical generators of P3a and P3b: A

710   LORETA study. Brain Res Bull 73:220–230.

711   Yang E, Brascamp J, Kang MS, Blake R (2014) On the use of continuous flash suppression for the study of visual

712   processing outside of awareness. Front Psychol 5:1–17.

713

714   **Figure Legends**

715   **Figure 1. Task design and behavioral results of Experiment 1 and 2. (A)** The trial structure of the attentional

716   blink task used in Experiment 1 and 2. Each trial consisted of a stream of rapidly presented letters in which

717   predefined target letters had to be detected and then reported at the end of the stream. The first target (T1: a

718   green G or H) always appeared at the fifth position. The second target (T2: D or K), was presented at varying

719   SOAs (lags) after the first one and was marked by placeholders. The identity of T1 predicted which of the T2

720   targets was most likely to appear, thereby introducing validly and invalidly predicted T2 targets. On 20% of the

721   trials no second target was presented and a random distractor letter was presented instead. **(B)** Percentage

722   correct T2 target detection at each of the T1-T2 lags, for valid expectations, invalid expectations and T2 absent

723   trials in Experiment 1. Validly predicted T2's were significantly more often perceived than invalidly predicted

724   T2's. **(C)** Percentage of T2 target detection at each of the T1-T2 lags, after a valid or invalid expectation or on a

725   T2 absent trial for Experiment 2. Again, validly predicted T2's were more often perceived, in particular at short

726   lags. Error bars represent SEM.

727

728   **Figure 2. ERP effects related to T2 prediction validity. (A)** Topographic maps of the difference between validly

729   and invalidly predicted T2s over time (0 = T2 onset). Cluster-based permutation tests were used to isolate the

730   significant events, while correcting for multiple comparisons across time and (electrode) space. On each head

731   map, channels with a significant effect for at least 50% of its time window are highlighted. **(B)** The average ERP

732   time-course of the 10 channels shown on the headmap on the left, shown seperately for each validity

733   condition. The significant time-window is marked by a black line above the x-axis. Invalidly predicted T2s were

734   associated with greater fronto-central negativity than validly predicted T2s. **(C)** Bar graphs showing the average

735   amplitude of the four conditions (visibility x prediciton) for the significant neural event shown in B. In all plots

736   error bars represent SEM.

737

738   **Figure 3. ERP effects related to T2 visibility analyses. (A)** Topographic maps showing the difference between

739   seen and missed T2s over time (0 = T2 onset). Cluster-based permutation tests were used to isolate the

740   significant events while correcting for multiple comparisons across time and (electrode) space. On each head

741   map, channels showing a significant difference for at least 50% of its time window are highlighted. Three

742   events were isolated based on the permutation tests. **(B,D,F)** For each of the events individually, the average

743   ERP time-course of the 10 channels shown on the headmap on the left, seperately for T2 seen and T2 missed
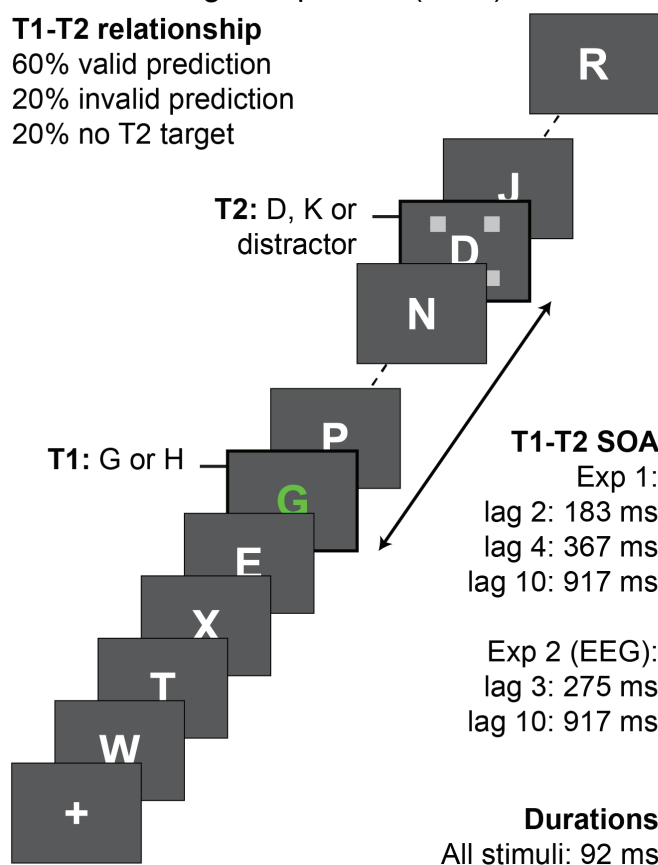
29

744   conditions is shown.. The significant time-window is marked by a black line above the x-axis. **(C,E,G)** Bar graphs

745   showing the average amplitude of the four conditions (visibility x prediciton) for the significant neural events

746   shown in B,D,F. In all plots error bars represent SEM.

747

748   **Figure 4. Task design and behavioral results of Experiment 3. (A)** Trial structure of the task used in Experiment

749   3. T1 visibility was staircased at approximately 75% correct by manipulating its duration (on 20% of trials no T1

750   was presented). **(B)** Percentage of correct T2 target detection at each of the T1-T2 lags, after a valid or invalid

751   expectation and on a T2 absent trials, for trials where T1 was correctly reported (T1 seen). As in Experiment 1

752   and 2, when T1 was seen, validly predicted T2's were more often detected than invalidly predicted T2's. **(C)**

753   Solid lines show percentage of T2 target detection at each of the T1-T2 lags, after a valid or invalid expectation

754   and on a T2 absent trials, for trials where T1 was presented but missed. In contrast to T1 seen trials (B), when

755   T1 was not seen, validity did not enhance T2 detection. However, a missed T1 still triggered a significant

756   attentional blink, as compared to trials on which no T1 was presented (dotted line). Error bars represent SEM.
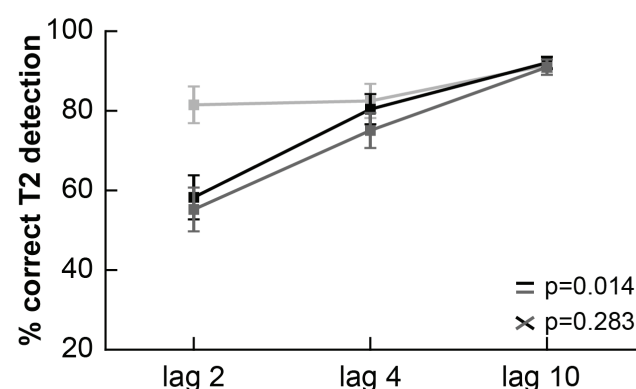
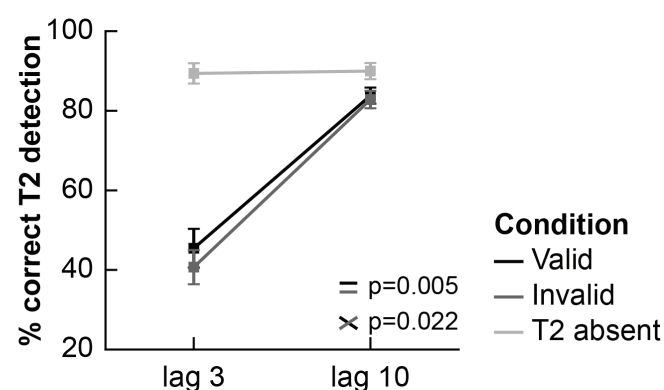A. Task Design: Exp 1 & 2 (EEG)

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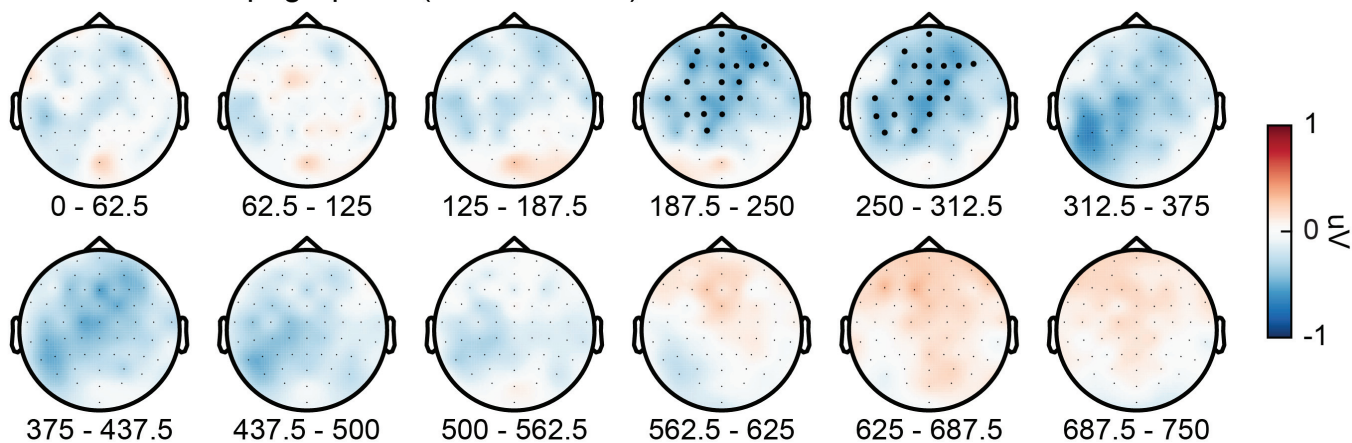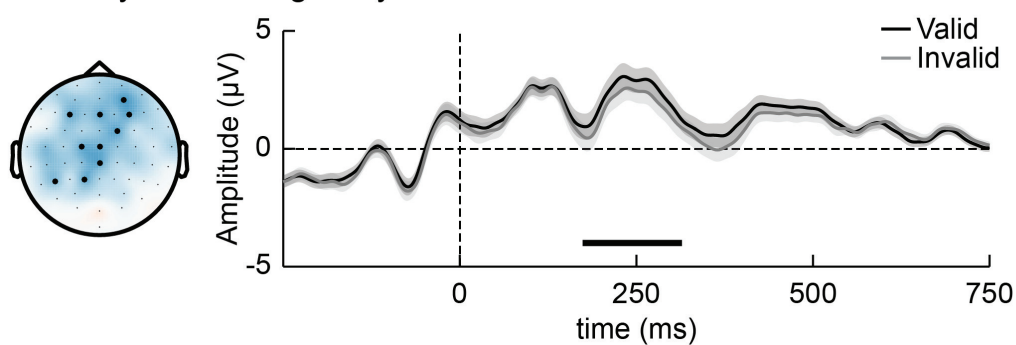20% invalid prediction
20% no T2 target

R

J

T2: D, K or
distractor

D

N

P

T1: G or H

G

E

X

T

W

+

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

p=0.014
p=0.283

C. Exp 2 (EEG)

p=0.005
p=0.022

Condition
— Valid
— Invalid
— T2 absent

A. Difference topographies (Invalid - Valid)

0 - 62.5   62.5 - 125   125 - 187.5   187.5 - 250   250 - 312.5   312.5 - 375

375 - 437.5   437.5 - 500   500 - 562.5   562.5 - 625   625 - 687.5   687.5 - 750

B. Early central negativity

C.
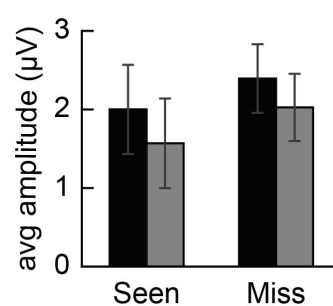
A. Difference topographies (Seen - Miss)

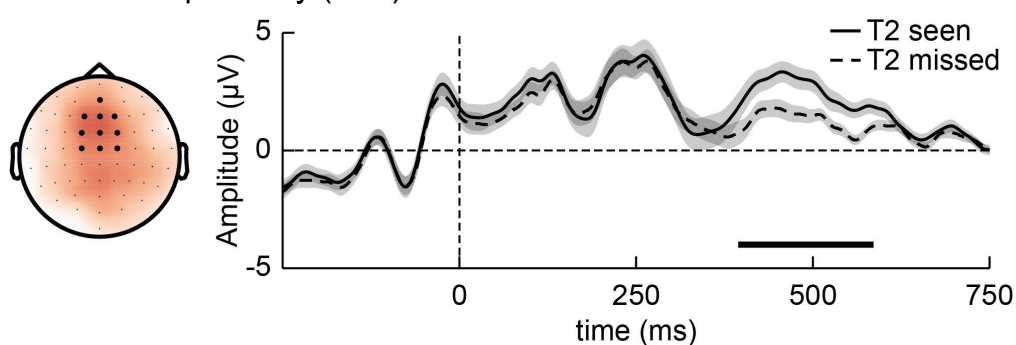0 - 62.5   62.5 - 125   125 - 187.5   187.5 - 250   250 - 312.5   312.5 - 375

375 - 437.5   437.5 - 500   500 - 562.5   562.5 - 625   625 - 687.5   687.5 - 750
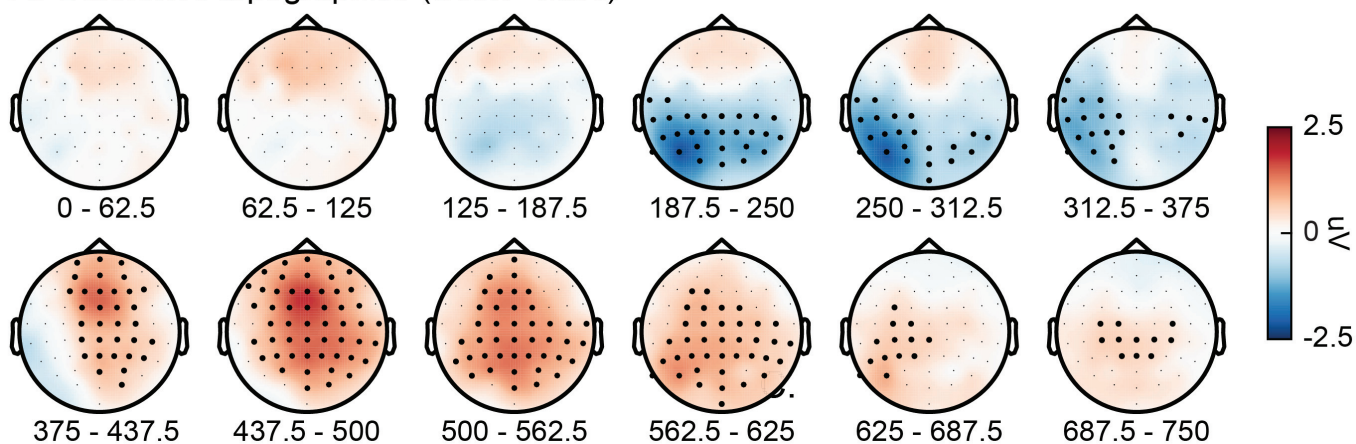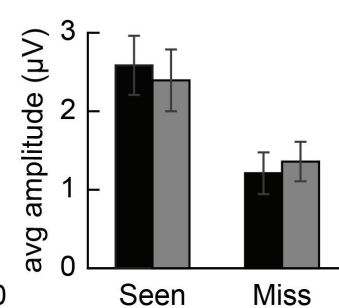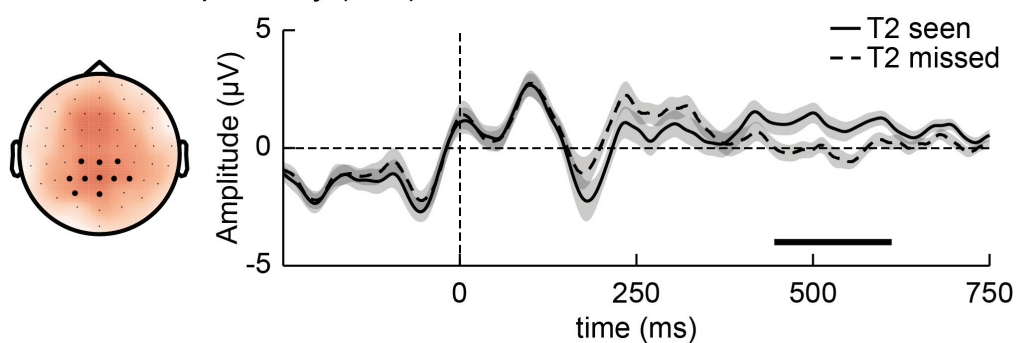
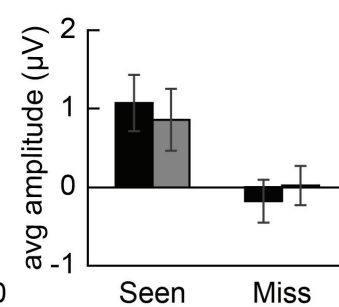B. Early left-posterior negativity
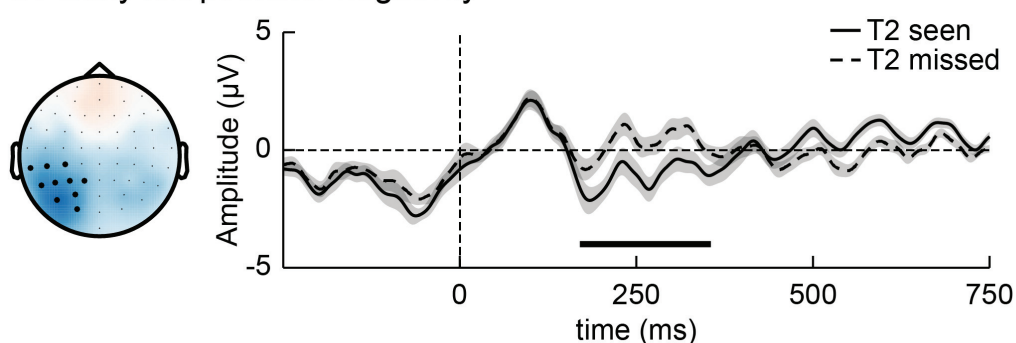
C.
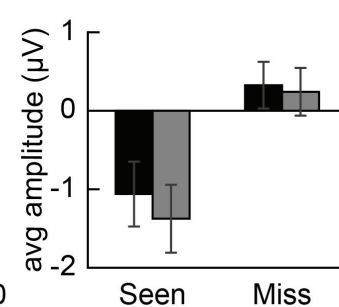
D. Anterior positivity (P3a)
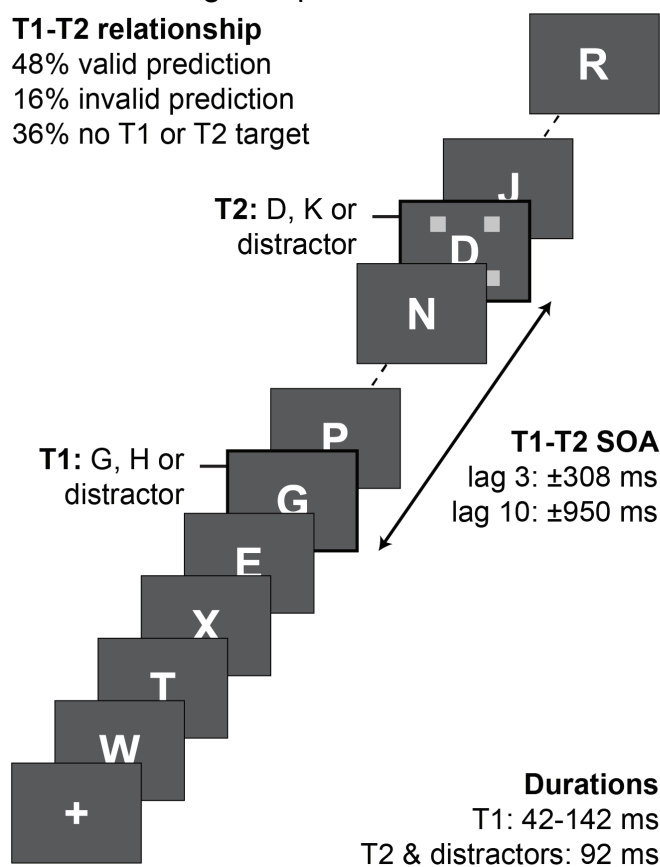
E.

F. Posterior positivity (P3b)

G.

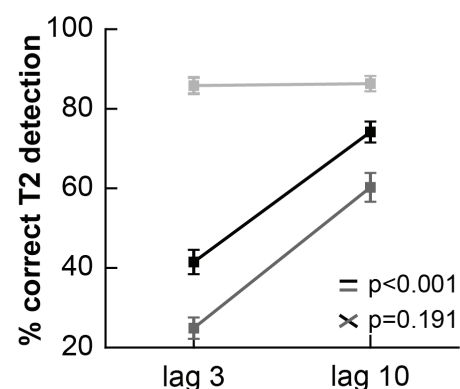## A. Task Design: Exp 3

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

R
J
D
N
P
G
E
X
T
W
+

B. Exp 3: T1 seen

% correct T2 detection

lag 3   lag 10

$=$ p<0.001
$\times$ p=0.191

C. Exp 3: T1 missed / absent

% correct T2 detection

lag 3   lag 10

$=$ p=0.554
$\times$ p=0.104

Condition
— Valid
— Invalid
— T2 absent
···· T1 absent