Statistical learning as reinforcement learning phenomena

Orpella J[4], Mas-Herrero E[1, 2], Ripollés P[4], Marco-Pallarés J[1,2,3], de Diego-Balaguer R[1, 2, 3, 5].

[1]Cognition and Brain Plasticity Unit, IDIBELL, L'Hospitalet de Llobregat, Spain

[2]Dept of Cognition Development and Educational Psychology, University of Barcelona, Barcelona, Spain

[3]Institute of Neuroscience, University of Barcelona, Barcelona, Spain

[4]Dept of Psychology, New York University, New York, USA

[5]ICREA, Barcelona, Spain

Corresponding author

Ruth de Diego-Balaguer

Faculty of Psychology

Pg. Vall d'Hebron 171

08035 Barcelona

tel. +34 934 034 768

ruth.dediego@ub.edu

1

1   Abstract

2   Statistical learning (SL) is the ability to extract regularities from the environment. In the domain

3   of language, this ability is fundamental in the learning of words and structural rules. In lack of

4   reliable online measures, statistical word and rule learning have been primarily investigated

5   using offline (post-familiarization) tests, which gives limited insights into the dynamics of SL and

6   its neural basis. Here, we capitalize on a novel task that tracks the *online* statistical learning of

7   language rules combined with computational modelling to show that online SL responds to

8   reinforcement learning principles rooted in striatal function. Specifically, we demonstrate - on

9   two different cohorts - that a Temporal Difference model, which relies on prediction errors,

10   accounts for participants' online learning behavior. We then show that the trial-by-trial

11   development of predictions through learning strongly correlates with activity in both ventral and

12   dorsal striatum. Our results thus provide a detailed mechanistic account of language-related SL

13   and an explanation for the oft-cited implication of the striatum in SL tasks. This work,

14   therefore, bridges the longstanding gap between language learning and reinforcement learning

15   phenomena.

Introduction

16

17 Statistical Learning (SL) is the ability to extract regularities from distributional information in

18 the environment. As a concept, SL was most popularized by the work of Saffran and colleagues,

19 who first demonstrated infants' use of the transitional probabilities between syllables to learn

20 both novel words(1) as well as simple grammatical relations(2). The idea of a mechanism for SL

21 has since raised a considerable amount of interest, and much research has been devoted to

22 mapping the scope of this cognitive feat. This work has been crucial in *describing* the SL

23 phenomenon as it occurs across sensory modalities (auditory(3–5), visual(6,7) and haptic(8)),

24 domains(7) (temporal and spatial), age groups(9,10), and even species (non-human

25 primates(11), and rats(12)). After all this research, however, little is yet known about the

26 *mechanisms* by which SL unfolds and their neural substrates.

27 Work on SL has recently moved towards the use of *online* measures of learning, which afford a

28 more detailed representation of the learning dynamics than typically used post-familiarization

29 test scores(13,14). Online measures capitalize on the gradual development of participants'

30 ability to *predict* upcoming sensory information (e.g., an upcoming syllable or word) as the

31 regularities of the input are learned (e.g., a statistical word-form or a grammatical pattern).

32 Indeed, prediction is often understood as the primary *consequence* of SL(15,16). Interestingly,

33 however, the status of prediction as a *mechanism* for SL, rather than the consequence of it, that

34 is, its causal implication in learning, has not been explicitly investigated.

35 In the current study, we examined the online development of predictions as a fundamental

36 computation for SL. In particular, we used an amply validated algorithm of reinforcement

37 learning –Temporal Difference(17,18) (TD)– to model participants' online learning behavior

38 and investigate its neural correlates. Note that, in adopting a model of reinforcement learning, a

39 domain where reward generally plays an important role, we are not assuming (nor discarding)

40 the phenomenological experience of reward (e.g., intrinsic reward(19,20)) during SL. Instead,

41 we assessed whether particular computational principles reflected in TD learning can account

42 for participants' SL behavior and their brain activity during learning.

43   TD models are based on the succession of predictions and prediction errors (the difference

44   between *predicted* and *actual* outcomes) at each time-step, by which predictions are gradually

45   tuned. In contrast to models typically used to explain SL (e.g., (21,22)), a vast body of research

46   supports the neurobiological plausibility of TD learning, with findings of neural correlates of

47   predictions and prediction errors both using cellular-level recordings and functional Magnetic

48   Resonance Imaging (fMRI). Several brain areas, notably the *striatum,* have been implicated in

49   the shaping of predictions over time and the selection of corresponding output behavior(23–

50   28). Interestingly, activity in the striatum has also been documented in SL in relation to

51   rule(29,30) as well as phonological word-form learning(31), but the precise role of these

52   subcortical structures in this domain remains unspecified.

53   With the aim of clarifying the mechanisms for SL and their neural underpinnings, we

54   combined computational (TD) modeling with fMRI of participants' brain activity while

55   performing a language learning task. In particular, participants completed an incidental non-

56   adjacent dependency learning paradigm. In natural languages, non-adjacent dependencies are

57   abundant and reflect important morphological and syntactic probabilistic rules (e.g., the

58   relationship between *un* and *able* in *un*believ*able*). Sensitivity to non-adjacent dependencies is

59   therefore important for grammar learning as well as for the early stages of word learning (i.e.

60   speech segmentation), both in prelexical-development(2) and beyond (i.e., second language

61   acquisition(32)). The main advantage of this particular rule-learning task over similar SL tasks

62   (e.g., (2,33)) is that it provides a reliable measure of *online* learning(34) that we can then model.

63   We used a TD algorithm for its greater sensitivity to temporal structure compared to simpler

64   RL models (e.g. Rescorla-Wagner(35)). Note that this an important prerequisite for non-

65   adjacent dependency learning specifically, since the to-be-associated elements are separated in

66   time. Nonetheless, we additionally investigated the adequacy of these simpler algorithms.

67   We expected the interplay of predictions and prediction errors, as modeled by the TD

68   algorithm, to closely match participants' online rule-learning behavior. In addition, and in line

69   with the aforementioned research on both reinforcement learning and SL, we expected striatal

70   activity to be associated with the computation of predictions.

71   **Results**

72   Two independent cohorts (Behavioral group: N = 19; fMRI group: N = 31) performed the

73   same incidental non-adjacent dependency learning task (see Methods for details). In brief,

74   participants were exposed to an artificial language, which, unbeknownst to them, contained

75   statistical dependencies between the initial (A) and final (C) elements of three-word phrases.

76   Orthogonal to SL, participants' instructions were to detect the presence or absence of a given

77   target word, which was always the final C element of one of the two A_C dependencies

78   presented. Online SL was measured as participants' decrease in reaction times (RTs) over trials,

79   which reflects the gradual learning of the predictive value of the initial element A in respect to

80   the dependent element C of each phrase (i.e., the equivalent of learning that *un* predicts *able* in

81   *un*believ*able*). In line with previous research(34), we expected faster reaction times in a Rule

82   block with such dependencies compared to a No Rule block with no statistical dependencies

83   (i.e. equally probable element combinations). This indicates that participants learned the

84   dependency between A and C elements, and were thus able to use the identity of the initial

85   word A to predict the presence or absence of their target word C.

86   This behavioral paradigm was initially tested in a group of nineteen volunteers (Behavioral

87   group: N = 19; 15 women; mean age = 21 years, SD = 1.47). After ruling out Order effects

88   (Rule block first/No Rule block first; main effect of Order and all its interactions with other

89   factors $p > 0.4$), a repeated-measures ANOVA with Rule (Rule/No Rule) and Target

90   (Target/No Target) as within-participant factors confirmed that learning of the dependencies

91   occurred over the Rule block. In particular, responses to phrases in the Rule block were overall

92   faster compared to the No Rule block ($F(1,18) = 13.6$, $p < 0.002$, Partial $\eta^2 = 0.43$; mean

93   difference = 149.40ms, SE = 40.51; Fig. 1A). A significant effect of Target ($F(1,18) = 24.46$, $p <$

94   0.001, Partial $\eta^2 = 0.58$) further indicated, as reported previously(34), that responses to target C

95   elements were faster than to no target C elements (mean difference = 68.66ms, SE = 13.88).

96   Importantly, we found no interaction between Rule and Target ($F(1,18) = 0.53$, $p > 0.48$),

97   suggesting that both dependencies were learned comparably.

98   **Figure 1. Behavioral group's rule-learning and temporal difference (TD) model results. (A)** Mean reaction times and

99   error bars in the Rule and No Rule blocks. The graph shows faster reaction times for Rule compared to No Rule

100   phrases, indicating successful learning. $**$ $p < 0.01$ **(B)** Plot of participants' mean reaction times (blue) against the TD

101   model's estimates of the development of predictions over learning (red; inverted as 1-P(A) before averaging and z-

102   scoring for display purposes). Vertical bars are the SD. RTs = reaction times. Reaction times were initially normalized

103   between 0 and 1 (Methods) and are plotted with the model estimates in z-score values. P(A) = TD model's predictions

104   from the initial word (A) of the dependencies.

105   We next assessed the extent to which a TD model (see Methods) could predict participants'

106   rule-learning behavior. If participants' reaction times reflect rule-learning in terms of the ability

107   to *predict* the last element (C) of a phrase from the identity of the initial (A) element, their

108   overall development should be mimicked by the model parameter representing the predictive

109   value of the initial (A) element (P(A); see Methods). That is, with time, the predictive value of

110   the initial A element according to the TD model should change (*increase*) in a way resembling

111   participants' reaction times. Note that reaction times can reflect prediction learning as well as

112   fluctuations due to decision processes, motor response preparation and execution, random

113   waxing and waning of attention, and system noise, which are not the object of this investigation.

114   Indeed, we used modelling to strip these off and so derive a purer measure of prediction

115   learning. Figure 1B shows the development of participants' reaction times over trials plotted

116   against the development of the predictive value of the initial word A (P(A)) as computed by the

117   TD model (inverted as 1-P(A) and z-scored for display purposes). Model fit was evaluated at

118   the individual level by the Model Fit Index (see Methods), calculated as 1 minus the Log-

119   Likelihood Ratio (LLR) between the Log-Likelihood Estimate (LLE) for the TD model and

120   the LLE for a model predicting at chance. Model Fit Index values of 1 would indicate an exact

121   model fit. Our results show a group average Model Fit Index of 0.74 ($std = 0.05$), indicating that

122   the TD model was 3 to 4 times better than the chance model at adjusting to participants'

123   reaction times. We additionally compared the performance of the TD model against that of a

124 Rescorla-Wagner (RW) model(35) (Supplementary Information, Fig. S1). In contrast to the

125 TD model, the RW model treats each AX_ combination as a single event, therefore combining

126 the predictive values of the two (A plus X) elements (35–37), and so does not take into account

127 non-adjacent relations (which are captured by the TD model through the devaluation of the

128 prediction parameter; see Methods). A paired-samples t-test indicated that Model Fit Index

129 values produced by the TD model were significantly better than those produced by the RW

130 Model (mean difference = 0.098, SE = 0.013; $t(18)$ = 7.65, $p < 0.001$, d = 0.83), which only

131 achieved an average Model Fit Index of 0.64 (2 to 3 times better than the chance model; std =

132 0.1). The TD model was, therefore, superior to the RW in adjusting to participants' reaction

133 time data.

134 We then replicated these behavioral results on a new cohort of participants from whom we

135 additionally acquired fMRI data while performing the incidental non-adjacent dependency

136 learning task (fMRI group; N = 31; 20 women; mean age = 23 years, SD = 3.62). We used the

137 same analytical procedure to evaluate rule learning at the behavioral level and model adequacy

138 thereafter. Having discarded block order effects (main effect of Order and all interactions: $p >$

139 0.1), a repeated-measures ANOVA with factors Rule (Rule/No Rule) and Target (Target/No

140 Target) indicated that rule-learning occurred in the Rule block ($F(1,30)$ = 4.96, $p < 0.034$,

141 Partial η^2 = 0.14), again with faster mean reaction times to rule compared to no rule phrases

142 (mean difference = 42.67ms, SE = 19.16; Fig. 2A). As expected, reaction times to target C

143 elements were faster than to no target C elements (mean difference = 57.2ms, SE = 43.28;

144 $F(1,30)$ = 54.15, $p < 0.001$, Partial η^2 = 0.64). As with the behavioral group's data, the null

145 interaction between the factors Rule and Target ($F(1,30)$ = 0.168, $p > 0.68$) indicated that both

146 target and no target dependencies were similarly learned.

147 **Figure 2. fMRI group's rule-learning and TD model results. (A)** Mean reaction times and error bars in the Rule and

148 No Rule blocks. Faster reaction times for Rule compared to No Rule phrases indicate successful learning. * $p < 0.05$

149 **(B)** Plot of participants' mean reaction times (blue) against the TD model's estimates of the development of predictions

150 over learning (red; inverted as 1-P(A) before averaging and z-scoring for display purposes). Vertical bars are the SD.

151 RTs = reaction times. Reaction times were initially normalized between 0 and 1 (Methods) and are plotted with the

152 model estimates in z-score values. $P(A)$ = TD model's predictions from the initial word (A) of the dependencies.

153 We next fitted the TD model to the fMRI group behavioral dataset. The development of

154 participants' reaction times is plotted in Figure 2B with the development of the predictive value

155 of the initial element (A) according to the TD model. At a group level, the mean Model Fit

156 Index was 0.71 (std = 0.04), again indicating a fit between 3 and 4 times better than that of a

157 model predicting at chance. This was also significantly better than the average Model Fit Index

158 produced by the RW Model (mean difference = 0.11, SE = 0.01; $t(30)$ = 12.20, $p < 0.001$, d =

159 1.26), which only reached a benchmark of 0.6 (again, 2 to 3 times better than the chance

160 model; std = 0.08; Fig. S1).

161 These results, therefore, represent a replication of our previous findings from the Behavioral

162 group, both in terms of the participants' overall statistical learning behavior and the adequacy of

163 the TD model in providing a mechanistic account of its dynamics.

164 To investigate the brain areas or networks sensitive to the trial-wise computations related to

165 statistical rule-learning from speech, we used a measure of the trial-by-trial development of

166 predictions from the initial word A of rule phrases ($P(A)$; see Methods) as estimated by the TD

167 model for each participant. Specifically, we correlated this proxy for prediction learning with

168 participants' trial-wise BOLD signal measures for the Rule block, time-locked to the onset of

169 the A element of each phrase. The contrast between $P(A)$-modulated Rule block activity against

170 an implicit baseline (Methods) yielded a large cluster covering most of the striatum (i.e.,

171 bilateral caudate nuclei, putamen, and ventral striatum; Fig. 3 and Table S1). Also, noteworthy,

172 there were two additional clusters, one in the left superior posterior temporal gyrus extending

173 medially to Rolandic opercular regions and another including right inferior and middle occipital

174 areas. While formalized as prediction learning, we note that activity in these regions could also

175 reflect the gradual increase in prediction error on the initial element A of each phrase. This is

176 because predictions and prediction errors on the A element should be commensurate with

177 each other, since it can never be anticipated (see Methods). An investigation of prediction error

178   responses on the C (target/non-target) elements was not possible due to the presence of button

179   presses on these elements as required by the task.

180   **Figure 3. Brain regions related to prediction learning (Rule P(A) – Baseline).** Activity in the Basal Ganglia (bilateral

181   caudate nuclei, putamen, and ventral striatum) and in the left posterior superior temporal gyrus (STG) was modulated

182   the trial-by-trial development of predictions (P(A)) as estimated by the TD model (contrast: Rule P(A) - Baseline).

183   Results are reported for clusters FWE-corrected at $p < 0.001$ at the cluster level (minimum cluster size = 20), with an

184   additional $p < 0.001$ uncorrected threshold at the voxel level. Neurological convention is used with MNI coordinates

185   shown at the bottom right of each slice.

186   In order to further support the specificity of these results, we completed a series of control

187   analyses. First, while, by definition, no rule can be derived from the No Rule block, this does

188   not preclude the engagement of particular brain regions in the attempt to capture the

189   relationship between specific phrase elements. In other words, we cannot ascertain that similar

190   type-computations are not taking place in the No Rule block, even when these will accrue no

191   substantial knowledge. Therefore, to assess the specificity of the reported clusters in prediction

192   *learning*, we next contrasted P(A)-modulated activity for Rule and No Rule blocks directly (see

193   Methods). That is, we compared the brain activity related to the trial-by-trial predictive value of

194   stimuli A during the Rule block to its counterpart during the No Rule block. Significant

195   differences centered on the same two relatively large clusters (Figure S2) observed in the main

196   analysis, namely bilateral caudate, putamen and ventral striatum and right middle occipital

197   cortex (not shown in the figure), with previous left temporal areas not reaching significance by

198   our selected threshold (Table S2). The converse contrast (P(A)-modulated No Rule vs P(A)-

199   modulated Rule) did not produce any significant results.

200   It is generally understood that the final goal of (TD) learning is to inform behavior(37). Even if

201   we consider predictions themselves as some form of *covert* behavior(38) used to optimize

202   online learning and processing, our paradigm also required participants to make an *overt*

203   response (a button press) to the presence of their target word. Reaction times are often used as

204   modulators of a condition's related BOLD signal to extract the variability pertaining to such

205   motor responses. However, as previously illustrated (Fig. 1 and Fig. 2), reaction times in the

206   Rule block will tend to show a close relationship to online learning. Hence, a more suitable

207   baseline to remove response related brain activity is the reaction times to the No Rule block,

208   that is, where no specific rule-learning can occur. We therefore contrasted P(A)-modulated

209   Rule block activity with the reaction time (RT)-modulated No Rule block activity (activity

210   estimates for the contrast between P(A)-modulated Rule and RT-modulated Rule activity are

211   also reported in Fig. S4 and Table S4). Significant prediction-related Rule activity remained in

212   the dorsal striatum, particularly in bilateral caudate nuclei and right putamen (Figure S3 and

213   Table S3). Altogether, therefore, our analyses (main and control) demonstrate that activity

214   within the striatum was related to the computations that facilitate statistical rule-learning from

215   speech as predicted by the TD model.

216   **Discussion**

217   In this study, we provide evidence for the SL of non-adjacent dependencies as an instance of

218   reinforcement learning. A TD model of reinforcement learning, which capitalizes on the

219   iteration of predictions and prediction errors, was able to mimic participants' reaction time data

220   reflecting gradual SL over trials. This was replicated on two independent cohorts, producing

221   similar model fits that were also clearly superior to those of simpler learning models. Functional

222   neuroimaging data of participants' online learning behavior also allowed us to examine the

223   neural correlates of prediction-based SL. In line with neuro-computational models of TD

224   learning, the trial-by–trial development of predictions from the initial word of the dependencies

225   was strongly related to activity in bilateral striatum. Importantly, striatal activity was unrelated to

226   the overt motor responses required by the task (i.e., button presses) or more general

227   computations, supporting the implication of the striatum *specifically* in prediction-based SL.

228   Evidence for the adequacy of a TD algorithm in capturing participants' online learning behavior

229   offers novel insights into the mechanisms for SL. In particular, our results underscore the

230   causal role of predictions for learning, compelling us to reassess the commonly assumed

231   relationship between SL and predictive processing. Indeed, SL not only enables predictions

232   (predictions as *a consequence* of SL), as generally understood (see p.e.(13)), but also capitalizes

233   on predictions (predictions as *a cause* of SL). This new understanding of SL can thus offer

234   interesting reinterpretations of previously reported correlations between SL abilities and

235   predictive processing(15), raising questions about the direction of causality.

236   Moreover, our results make an important contribution to the understanding of the neuro-

237   biological basis of SL. While previous research(13,39) has shown a similar behavioral

238   development of online SL (cf. Fig. 1), brain imaging data and its link to a mechanistic

239   explanation of learning were lacking. Here we used a measure of online SL behavior in

240   combination with computational modeling and fMRI data to unveil the basic mechanism

241   underlying learning and its brain correlates. A complementary approach to describing online

242   SL, which involves the frequency-tagging of participants' neurophysiological responses over

243   learning(40–42), has recently been used to track the emergence of new representations (in time

244   and neuro-anatomical space) as participants learn. We add to these findings by providing a

245   mechanistic account for *how* these representations (i.e., learning) come to be and a plausible

246   neuro-anatomical substrate for its key computations. In particular, we show that the gradual

247   development of predictions for SL is related to robust and widespread activity in bilateral

248   striatum. This finding adds a valuable degree of specificity to the oft-reported implication of

249   these subcortical structures in artificial grammar learning and SL more generally(29–31).

250   Furthermore, both the adequacy of a TD model and the involvement of the striatum in

251   prediction-based SL place this cognitive ability squarely in the terrain of reinforcement learning.

252   Indeed, the link between prediction learning and activity in the striatum is one of the most

253   robust findings in the reinforcement learning literature, from intracranial recordings to fMRI

254   studies(24–28,37,43–45). Activity in the ventral striatum, in particular, has been associated with

255   the delivery and anticipation of rewarding stimuli of different types (i.e. from primary to higher-

256   order rewards)(46). More specifically, the ventral striatum interacts in complex ways with the

257   dopaminergic system (mainly Ventral Tegmental area/Substantial Nigra pars compacta;

258   VTA/SNc) with responses consistent with the computation of reward P.E.(23,47–49). Under

259   this light, our reported pattern of activity in the ventral striatum is consistent with the gradual

260   transfer over learning of prediction error related dopaminergic responses from rewarding to

261   predictive stimuli as found in classic conditioning paradigms(50,51). That is, a gradual increase

262   in response on A elements may be expected as their predictive value is learned, since these

263   elements can never be anticipated. Alternatively, activity in the ventral striatum could reflect

264   inhibitory signals aimed to attenuate dopaminergic inputs from the VTA/SNc(52) in response

265   to C elements as these become more predictable.

266   From a theoretical standpoint, it may be necessary to distinguish between the response of the

267   reward system for learning and the phenomenological experience of reward(20,53). Recent

268   evidence(54–56), nonetheless, supports the notion of language learning as *intrinsically*

269   *rewarding*(19), and suggests quantitative over qualitative differences between endogenous and

270   exogenous sources of reward(20). So far, the adequacy of reinforcement learning algorithms for

271   the learning of intrinsically rewarding tasks has mainly remained theoretical(20,57). Our results

272   now contribute to this literature by showing their suitability in specific instances of SL.

273   Still within the computations of the TD model, activity in the caudate nuclei of the dorsal

274   striatum could respond to the updates at each timestep of the outcome value representations

275   associated with each stimulus in ventromedial and orbitofrontal areas(58). Caudate activity

276   could also promote the (attentional) selection of behaviorally relevant elements in the phrase in

277   frontal cortical areas(59–61). While this attentional selection should initially pertain to C

278   elements (the target of the monitoring task), a shift towards A elements may be expected as

279   their predictive value increases. This interpretation is consistent with the finding of a gradual

280   increase in the P2 ERP component (related to attentional deployment) over the exposure to (A

281   elements of) non-adjacent dependencies but not to similar but unstructured material(62,63).

282   Finally, activity in bilateral putamen could relate to the selection of the specific motor actions in

283   pre/motor regions(64). The fact that the response of this area is reduced but not eliminated

284   when regressing out overt motor responses (button presses; Fig. S3; cf. Fig. 3) raises the

285   possibility that putaminal activity additionally reflects the selection of *covert* motor responses,

286   namely, of (speech) motor programs corresponding to the predicted (C) elements. The

287   selection of these motor-articulatory plans may be used to generate sensory-level

288   predictions(38) ultimately translating into increasingly faster RTs for predicted C elements. In

289   this view, activity in the pSTG (see Fig. 3) would reflect the downstream (i.e. sensory)

290   consequences of this selection(38). We conjecture that prediction-based SL is fundamentally

291   linked to such motor engagement as part of the learning mechanism orchestrated by the

292   striatum. This is consistent with the observation that participants more adept to predicting

293   speech inputs embedded in noise, known to involve the speech motor system(65), are also

294   better statistical learners(15), and agrees with the well-accepted role for these structures in

295   procedural learning(66,67) and the managing of motor routines(49,64,68). Note that this speech

296   motor engagement for learning should become of critical importance when putative alternative

297   learning mechanisms (e.g., purely sensory based) are weakest, for example, when a temporal

298   separation is imposed between the elements to be associated, as in our non-adjacent

299   dependency learning task.

300   In contrast to previous research on grammar learning(31,69,70), the trial-wise development of

301   predictions did not reveal significant activity in the left inferior frontal gyrus (though see Fig.

302   S5). As noted by Karuza and colleagues(14), a possible explanation for this concerns the

303   emphasis that particular models of learning place on specific computations, with TD models

304   being most sensitive to subcortical (rather than neo-cortical) activity(14). Future research should

305   investigate alternative models of learning as a means to relate other neural correlates of SL,

306   such as the left inferior frontal gyrus, to other subroutines of the learning process.

307   In sum, by the combination of an online measure of SL, computational modeling, and

308   functional neuroimaging, we provide evidence for SL as a process of gradual prediction learning

309   strongly related to striatal function. This work, therefore, makes a valuable contribution to our

310   understanding of the mechanisms and neurobiology of this cognitive phenomenon, and

311   introduces the provoking possibility of language-related SL as an instance of reinforcement

312   learning orchestrated by the striatum.

313   **Materials and Methods**

13

Participants

315   Two independent cohorts participated in the study. We first collected data from twenty

316   volunteers from the Facultat de Psicologia of the Universitat de Barcelona as the behavioral

317   group. Data from one participant was not correctly recorded, so the final cohort comprised

318   nineteen participants (15 women, mean age = 21 years, sd = 1.47). We used the partial η^2

319   obtained for the main effect of Rule in the behavioral group to compute a sample size analysis

320   for the fMRI group. To ensure 90% of power to detect a significant effect in a 2 x 2 repeated

321   measures ANOVA at the 5% significance level based on this measure of effect size,

322   MorePower(71) estimated that we would need a sample size of at least 16 participants.

323   However, considering i) that we expected participants to perform worse inside of the fMRI

324   scanner, and ii) the recommendation that at least 30 participants should be included in an

325   experiment in which the expected effect size is medium to large(72), we finally decided to

326   double the recommended sample size for the fMRI experiment. The fMRI group thus

327   consisted of 31 participants (20 women, mean age = 23 years, sd = 3.62) recruited at the

328   Universidad de Granada. All participants were right-handed native Spanish speakers and self-

329   reported no history of neurological or auditory problems. Participants in the fMRI group were

330   cleared for MRI compatibility. The study was approved by the ethics committee of the

331   Universitat de Barcelona and the ERC ethics scientific office and was conducted in accordance

332   with the Declaration of Helsinki. Participation was remunerated and proceeded with the written

333   informed consent of all participants.

Rule Learning Paradigm

335   Two different artificial languages were used in the rule-learning task. Each language comprised

336   twenty-eight bi-syllabic (consonant-vowel-consonant-vowel) pseudo-words (henceforth, words).

337   Words were created using Mbrola speech synthesizer v3.02b (Dutoit et al. 1996) through

338   concatenating diphones from the Spanish male database 'es1' (http://tcts.fpms.ac.be/synthesis/)

339   at a voice frequency of 16 KHz. The duration of each word was 385ms. Words were combined

340   to form three-word phrases with 100ms of silence inserted between words. Phrase stimuli were

14

341 presented using the software Presentation® (Neurobehavioral Systems) via Sennheiser® over-

342 ear headphones (pilot group) and MRI-compatible earphones (Sensimetrics™, Malden, MA,

343 USA; fMRI group).

344 The learning phase consisted of a Rule block and a No Rule block, each employing a different

345 language. The order of blocks was counterbalanced between participants. We also

346 counterbalanced the languages assigned to Rule and No Rule blocks. The Rule block consisted

347 of 72 rule phrases (phrases with dependencies) whereby the initial word (A) was 100%

348 predictive of the last word (C) of the phrase. We used two different dependencies (A1_C1 and

349 A2_C2) presented over 18 different intervening (X) elements to form AXC-type phrases.

350 Twelve of the 18 X elements were common to both dependencies, while the remaining 6 were

351 unique to each dependency. These 36 rule phrases were presented twice over the Rule block,

352 making a total of 72 AXC-type rule phrases issued in pseudo-random order. The probability of

353 transitioning from a given A element to a particular X was therefore 0.056. Phrases in the No

354 Rule block were made out of the combination two X elements and a final C element (either C1

355 or C2, occurring with equal probability). Note that, while C elements could be predicted with

356 100% certainty in the Rule block, these could not be predicted from the previous X elements in

357 the No Rule block. X elements were combined so that each X word had an equal probability to

358 appear in first and second position but never twice within the same phrase. Forty-eight no rule

359 phrases were presented twice over the No Rule block, making a total of 96 pseudo-randomized

360 XXC-type no rule phrases. Each three-word phrase, in both Rule and No Rule blocks, was

361 considered a trial. A recognition test was issued at the end of each block to assess offline

362 learning (see Supplementary Information for further details).

363 To obtain an online measure of incidental learning, participants were instructed to detect, as

364 fast as possible via a button press, the presence or absence of a given target word. The target

365 word for each participant remained constant throughout the block and was no other than one

366 of the two C elements of the language (C1 or C2, counterbalanced). A written version of the

367 participant's target word was displayed in the middle of the screen for reference throughout the

368   entire learning phase. Importantly, participants were not informed about the presence of rules,

369   so this word-monitoring task was in essence orthogonal to rule-learning. Yet, if incidental

370   learning of the dependencies occurred over trials in the Rule block, faster mean reaction times

371   should be observed for this block compared to the No Rule block where the appearance or

372   non-appearance of the target word could not be anticipated from any of the preceding

373   elements.

374   Participants in the behavioral group indicated the detection or non-detection of the target word

375   by pressing the left and right arrow keys of the computer keyboard, respectively. They were

376   required to use their left index finger to press the left arrow key, and the right index finger to

377   press the right arrow key. Participants in the fMRI group responded using the buttons

378   corresponding to thumb and index fingers in an MRI compatible device held in their right

379   hand. Response buttons were not counterbalanced for either group. Inter-trial interval was fixed

380   at 1000ms in the pilot study and jittered (with pseudo-random values between 1000 and

381   3000ms) for testing during fMRI acquisition. At the end of a given phrase, a maximum of

382   1000ms was allowed for participants to respond. Reaction times were calculated from onset of

383   the last word in the phrase until button press. Only trials with correct responses were entered

384   into subsequent analyses. Participants' Rule Effects were calculated as the mean reaction time

385   difference between no rule and rule trials. A repeated-measures ANOVA on participants'

386   reaction time data with within-subjects factors Rule (Rule/No Rule) and Target (Target/No

387   Target) and Order as a between-subjects factor was initially performed to discard block order

388   effects. A repeated-measures ANOVA with factors Rule (Rule/No Rule) and Target (Target/No

389   Target) was subsequently performed to assess the statistical significance of rule learning.

390   **Temporal Difference model**

391   We modelled subjects' learning of the dependencies using a Temporal Difference (TD)

392   model(17,18). Drawing from earlier models of associative learning, such as the Rescorla-

393   Wagner (RW) model(35), the main assumption of TD models is that learning is driven by a

16

394 measure of the mismatch between predicted and actual outcome(17,18,37,73) (i.e., prediction

395 error (P.E.). This scalar quantity is computed as:

396 $\partial(t) = o(t) - p(t-1)$ (equation 1)

397 where $\partial(t)$ is the P.E. term at a given time-point t within a trial, which amounts to the

398 discrepancy between the outcome o at that time-point, and the prediction p at the previous

399 time-point (t-1).

400 Computationally, learning through TD is therefore conceptualized (and modelled) as

401 *prediction* learning(37), where predictions p at each time-step are updated according to:

402 $p(t) = p(t-1) + \alpha \cdot \partial$ (equation 2)

403 where α is a free parameter that represents the *learning rate* of the participant and determines

404 the weight attributed to new events and the P.E. they generate(17).

405 One of the advantages of TD models over simpler models of learning, such as the RW, is that

406 they account for the *sequence* of events leading to an outcome, rather than treating each trial as

407 a discrete temporal event. That is, although each trial for the participant (i.e., each three-word

408 phrase) was equivalently treated as a trial for the TD model, model updates occurred at the

409 presentation of each individual element (see below). TD models are thus sensitive to the

410 precise temporal relationship between the succession of predictions and outcomes that take

411 place in a learning trial(17). Note that this is particularly valuable in trying to account for the

412 learning of non-adjacent dependencies as distinct from adjacent ones, making a TD model

413 preferable in such cases. This feature is implemented as a temporal discounting factor, this is an

414 additional free parameter γ that represents the devaluation of predictions that are more distant

415 from the outcome(44,74). Thus, upon 'hearing' the final element of a rule (AXC) phrase, the

416 prediction from the initial element A was updated according to:

417 $p(A,t) = p(A,t-1) + \alpha \cdot \partial \cdot \gamma$ (equation 3)

418 To obtain the free parameters α and γ for each participant, their reaction times were

419 normalized between 0 and 1 by the function 1 ./ (1 + exp(zRTs)). We then selected the α and γ

420 values that produced the minimum Log-Likelihood Estimate (LLE), indicating the best possible

421 fit between the model predictions and the participant's transformed RTs. For this, we used

422 Matlab®'s (Matlab® R2017 by Mathworks®) *fmincon* function, which implements a Nelder-

423 Mead simplex method(75). The model was then run for each block (Rule and No Rule)

424 separately, from which trial-wise prediction (p(A,t) and p(X,t)) values for the different phrase

425 elements A and X (resulting in matrices P(A) and P(X), respectively) were computed. We

426 assigned o a value of 1 for target outcomes at the end of a trial, and a value of -1 for no target

427 outcomes. Note that the sign choice represents a convenient yet arbitrary means to distinguish

428 target and no target outcomes within the same model. This could have been inverted (o (target)

429 = - 1, o (no target) = 1) with no difference in the model's results. Following convention(37), the

430 P.E. at the presentation of the second element of a phrase (X) was always computed as the

431 difference between the values of P(X,t) and P(A,t). Finally, it is important to note that, at the

432 time element A is presented, it is not possible to differentiate the prediction's value from the

433 P.E. the same element A elicits. The reason is that, according to TD models, *all* events in a trial

434 will elicit a P.E.. Thus, the P.E. for the initial (unpredicted) element (A) will be ∂ (A) = p (A) –

435 p (baseline) which, given that p (baseline) = 0, renders ∂ (A) = p (A)(37). This is therefore

436 crucial in the interpretation of the reported brain activity related to this parameter (see main

437 text).

438 To illustrate the consistency between participants' reaction times and model predictions, both

439 which we assume to be proxies for statistical rule-learning, we plotted the development of P(A)

440 computed by the model (inverted as 1-P(A)) averaged across participants against the mean

441 reaction times of the participants over trials in the Rule block (both z-scored; main text Fig. 1

442 and Fig. 2).

443 To assess the fit of the TD model, we computed for each participant the Log-Likelihood Ratio

444 (LLR) between the TD model's LLE and the LLE produced by a model predicting at chance.

445   To make fit assessment more intuitive, a Model Fit Index was then calculated as 1 – LLR,

446   where higher Model Fit Index values equate to a better fit. The overall fit of the TD model was

447   assessed at the group level by averaging across participants Model Fit Index.

448   **fMRI acquisition and apparatus**

449   The rule-learning task comprised a single run with 830 volumes. Functional T2*-weighted

450   images were acquired using a Siemens Magnetom TrioTim syngo MR B17 3T scanner and a

451   gradient echo-planar imaging sequence to measure blood oxygenation level dependent (BOLD)

452   contrast over the whole brain [repetition time (TR) 2000ms, echo time (TE) 25ms; 35 slices

453   acquired in descending order; slice-thickness: 3.5 mm, 68×68 matrix in plane resolution = 3.5

454    $\times$ 3.5 mm; flip angle = 180 º]. We also acquired a high-resolution 3D T1 structural volume

455   using a magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence [TR =

456   2500ms, TE = 3.69ms, inversion time (TI) = 1100ms, flip angle = 90°, FOV = 256 mm, spatial

457   resolution = 1 mm^3/voxel].

458   **fMRI preprocessing and analysis**

459   Data were preprocessed using Statistical Parameter Mapping software (SPM12, Welcome Trust

460   Centre for Neuroimaging, University College, London, UK, www.fil.ion.ucl.ac.uk/spm/).

461   Functional images were realigned, and the mean of the images was co-registered to the T1. The

462   T1 was then segmented into grey and white matter using the Unified Segmentation

463   algorithm(76) and the resulting forward transformation matrix was used to normalize the

464   functional images to standard Montreal Neurological Institute (MNI) space. Functional

465   volumes were re-sampled to 2 mm^3 voxels and spatially smoothed using an 8 mm FWHM

466   kernel.

467   Several event-related design matrices were specified for convolutions with the canonical

468   hemodynamic response function. Trial onsets were always defined as the onset of the first word

469   of the phrase. To identify brain regions related to the trial-by-trial development of subjects'

470   predictions/prediction errors, a model with the conditions Rule Target, Rule No Target, No

471 Rule Target and No Rule No Target, and all offline test conditions was specified at the first

472 level, including a parametric modulator (a vector) corresponding to the trial-wise

473 prediction/prediction error ($p(A)$) for each of the conditions of interest (Rule Target, Rule No

474 Target, No Rule Target and No Rule No Target). For each participant, the contrasts Rule $P(A)$

475 vs implicit and Rule $P(A)$ vs. No Rule $P(A)$ were calculated and entered into corresponding

476 one-sample t-tests. An additional model was specified with the aim of removing response

477 related activity from the $P(A)$-modulated regressor. This included the inverted reaction time per

478 trial as the first parametric modulator for each condition and the $p(A)$ as its second parametric

479 modulator. Note that the reaction times for the Rule block will relate to rule learning as well as

480 motor activity, while the reaction times corresponding to the No Rule block will predominantly

481 relate to motor responses. For this reason, to identify brain regions modulated by the $P(A)$ in

482 the Rule block while extracting response-related motor activity, for each participant we

483 calculated the contrast Rule $P(A)$ vs No Rule RTs in addition to the more conventional contrast

484 Rule $P(A)$ vs Rule RTs, entering these into one-sample t-tests at the second level.

485 In all cases, data were high-pass filtered (to a max. of 1/90 Hz). Serial autocorrelations were also

486 estimated using an autoregressive (AR(1)) model. We additionally included, in all the models

487 described above, the movement parameter estimates for each subject computed during

488 preprocessing to minimize the impact of head movement on the data. We used the Automated

489 Anatomical Labelling Atlas(77) included in the xjView toolbox

490 (http://www.alivelearn.net/xjview8/) to identify anatomical and cytoarchitectonic brain areas.

491 Group results are reported for clusters at a $p < 0.001$ FWE-corrected threshold at the cluster

492 level with a minimum cluster extent of 20 voxels and an additional $p < 0.001$ uncorrected

493 threshold at the voxel level.

494 **Acknowledgements**

497 **Author Contributions**

498   R.d.D.-B., J.O. and J.M.-P. designed the research. J.O. collected the data. J.O., E.M.-H. and

499   P.R. analyzed the data. J.O., P.R., E.M.-H., J.M.-P. and R. d. D.-B. wrote the manuscript.

500   **Competing interests**

501   The authors declare no competing interests.

502   **Supplementary Information Captions**

503   **Figure S1.** Plot of **(A)** Behavioral group and **(B)** fMRI group participants' mean reaction times (blue) against the

504   Rescorla-Wagner model's estimates of the development of predictions over learning (red; inverted as 1-P(A) before

505   averaging and z-scoring for display purposes). Vertical bars are the SD. RTs = reaction times. Reaction times were

506   initially normalized between 0 and 1 (Methods) and are plotted with the model prediction estimates in z-score values.

507   P(A) = Rescorla-Wagner model's predictions from the initial word (A) of the dependencies.

508   **Figure S2. Brain regions related to prediction learning (Rule P(A) – No Rule P(A)).** Activity in the Basal Ganglia

509   (bilateral caudate nuclei, putamen, and ventral striatum; see Table S2) was modulated the trial-by-trial development of

510   predictions (P(A)) as estimated by the TD model (contrast: Rule – No Rule). Results are reported for clusters FWE-

511   corrected at $p < 0.001$ at the cluster level (minimum cluster size = 20), with an additional $p < 0.001$ uncorrected

512   threshold at the voxel level. Neurological convention is used with MNI coordinates shown at the bottom right of each

513   slice.

514   **Figure S3. Brain regions related to prediction learning (Rule P(A) – No Rule RTs).** Subtracting the activity for the No

515   Rule block modulated by participants' reaction times from the P(A)-modulated Rule block activity had virtually no

516   effect on Basal Ganglia activity estimates (see **Table S3**). Significant centered on the caudate nuclei and the right

517   putamen. Results are reported at a $p < 0.001$ FWE-corrected threshold at the cluster level with 20 voxels of minimum

518   cluster extent, with an additional uncorrected $p < 0.001$ threshold at the voxel level. Neurological convention is used

519   with MNI coordinates shown at the bottom right of each slice.

520   **Figure S4. Brain regions related to prediction learning (Rule P(A) – Rule RTs).** Significant activations by the contrast

521   between P(A)-modulated Rule and RT-modulated Rule activity (see also **Table S4**) were found in a widespread left-

522   lateralized network of areas, including a large portion of the inferior frontal gyrus, parts of the pre- and post-central

523   gyri, and of the superior temporal gyrus in and around the left auditory cortex. Interestingly, bilateral caudate nuclei

524   were also statistically significant along with a small portion of the thalamus. Results are reported for clusters FWE-

525   corrected at $p < 0.001$ at the cluster level (minimum cluster size = 20), with an additional $p < 0.001$ uncorrected

526   threshold at the voxel level. Neurological convention is used with MNI coordinates shown at the bottom right of each

527   slice.

528   **Table S1. Whole brain fMRI Rule P(A)-modulated activity vs. implicit baseline.** Group-level fMRI local maxima for

529   the P(A)-modulated Rule against implicit baseline contrast (see also red-yellow regions in **Fig. 3**, main text). Results are

530   reported for clusters FWE-corrected at p < 0.001 at the cluster level (minimum cluster size = 20), with an additional p

531   < 0.001 uncorrected threshold at the voxel level. MNI coordinates were used. BA, Brodmann Area.

532   **Table S2. Whole brain fMRI P(A)-modulated activity for Rule vs. No Rule.** Group-level fMRI local maxima for the

533   P(A)-modulated Rule minus P(A)-modulated No Rule contrast (see also red-yellow regions in **Fig. S2**). Results are

534   reported for clusters FWE-corrected at p < 0.001 at the cluster level (minimum cluster size = 20), with an additional p

535   < 0.001 uncorrected threshold at the voxel level. MNI coordinates were used. BA, Brodmann Area.

536   **Table S3. Whole brain fMRI activity for the P(A)-modulated Rule vs. RT-modulated No Rule contrast.** Group-level

537   fMRI local maxima for the P(A)-modulated Rule minus RT-modulated No Rule contrast (see also red-yellow regions

538   in **Fig. S3**). Results are reported at a p < 0.001 FWE-corrected threshold at the cluster level with 20 voxels of minimum

539   cluster extent, with an additional uncorrected p < 0.001 threshold at the voxel level. MNI coordinates were used. BA,

540   Brodmann Area.

541   **Table S4. Whole brain fMRI Rule P(A)-modulated activity vs. Rule RT-modulated activity.** Group-level fMRI local

542   maxima for the P(A)-modulated Rule minus RT –modulated Rule contrast (see also red-yellow regions in **Fig. S4**).

543   Results are reported for clusters FWE-corrected at p < 0.001 at the cluster level (minimum cluster size = 20), with an

544   additional p < 0.001 uncorrected threshold at the voxel level. MNI coordinates were used. BA, Brodmann Area.

545   **Text S1. Offline Recognition Test.** Following each block, participants' knowledge of the rules was assessed via a

546   recognition test. Participants were presented with correct sentences (phrases that conformed the rules) and incorrect

547   sentences (phrases that violated the rules). In half the trials, incorrect sentences consisted of violations of the A_C

548   dependencies where A and C elements maintained their correct order within the phrase but belonged to different rule

549   structures (i.e., A1xC2, A2xC1). In the other half, incorrect sentences contained order violations, where A and C

550   elements from a dependency swapped positions (i.e., C1xA1 and C2xA2). The complete offline test consisted in a

551   total of 48 test phrases (24 per dependency). Participants were required to discriminate phrases that could belong to

552   the previously heard language from phrases that could not by pressing the appropriate button. A maximum of 1500

553   ms was allowed to respond, after which there was a jittered interval (1-3 secs.) before the next trial began. Participants'

554   ability to discriminate rule items from violations was assessed by computing d prime scores (d′) from their responses.

555   For each participant, the proportion of hits (i.e., yes responses to correct phrases) and false alarms (i.e., yes responses

556   to incorrect phrases) was used to calculate the d′ score. Hit and false alarm rates of zero or one were corrected

557   according to Macmillan and Kaplan (1985). We computed two distinct d' scores by using false alarms to 1) order

558   violations (d' Cat) and 2) to dependency violations (d' Dep). These scores were then submitted to one-sample t-tests

559   against 0 to determine statistical significance. After a Rule block, the test for the corresponding language's Rule block

560　was administered. After a No Rule block, the test for the corresponding language's Rule block was administered. We

561　thus expected learning of the language's specific dependencies (significant d' Dep) only in the Rule block but learning

562　of positional information (significant d' Cat) after both Rule and No Rule blocks. Participants from both groups

563　exhibited a similar pattern of results suggesting that they were able to significantly discriminate correct phrases from

564　dependency violations only after the Rule block (Behavioral group Rule block: mean d' Dep = 0.92, std = 1.46, t(21)

565　= 2.97, p < 0.01; Behavioral group No Rule block: mean d' Dep = 0.14, std = 0.52, t(21) = 1.3, p > 0.2; fMRI group

566　Rule block: mean d' Dep = 0.36, std = 0.96, t(30) = 2.1, p < 0.04; fMRI group No Rule block: mean d' Dep = 0.12,

567　std = 0.49, t(30) = 1.3, p > 0.19) and correct phrases from category violations after both Rule and No Rule blocks

568　(Behavioral group Rule block: mean d' Cat = 1.74, std = 1.19, t(21) = 6.87, p < 0.001; Behavioral group No Rule block:

569　mean d' Cat = 1.82, std = 1.18, t(21) = 7.23, p < 0.001; fMRI group Rule block: mean d' Cat = 1.36, std = 1.12, t(30) =

570　6.75, p < 0.001; fMRI group No Rule block: mean d' Cat = 1.37, std = 1.01, t(30) = 7.55, p < 0.001).
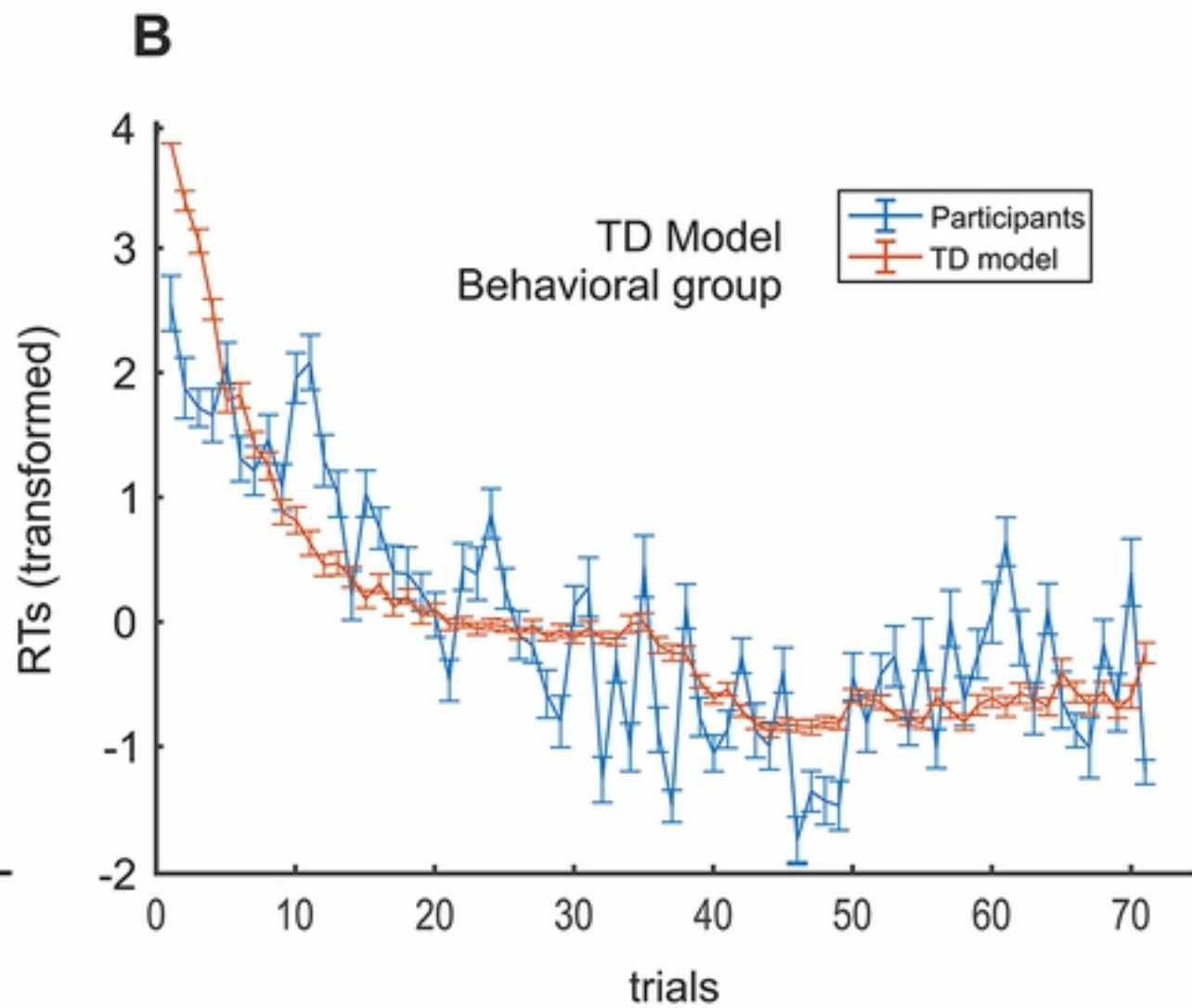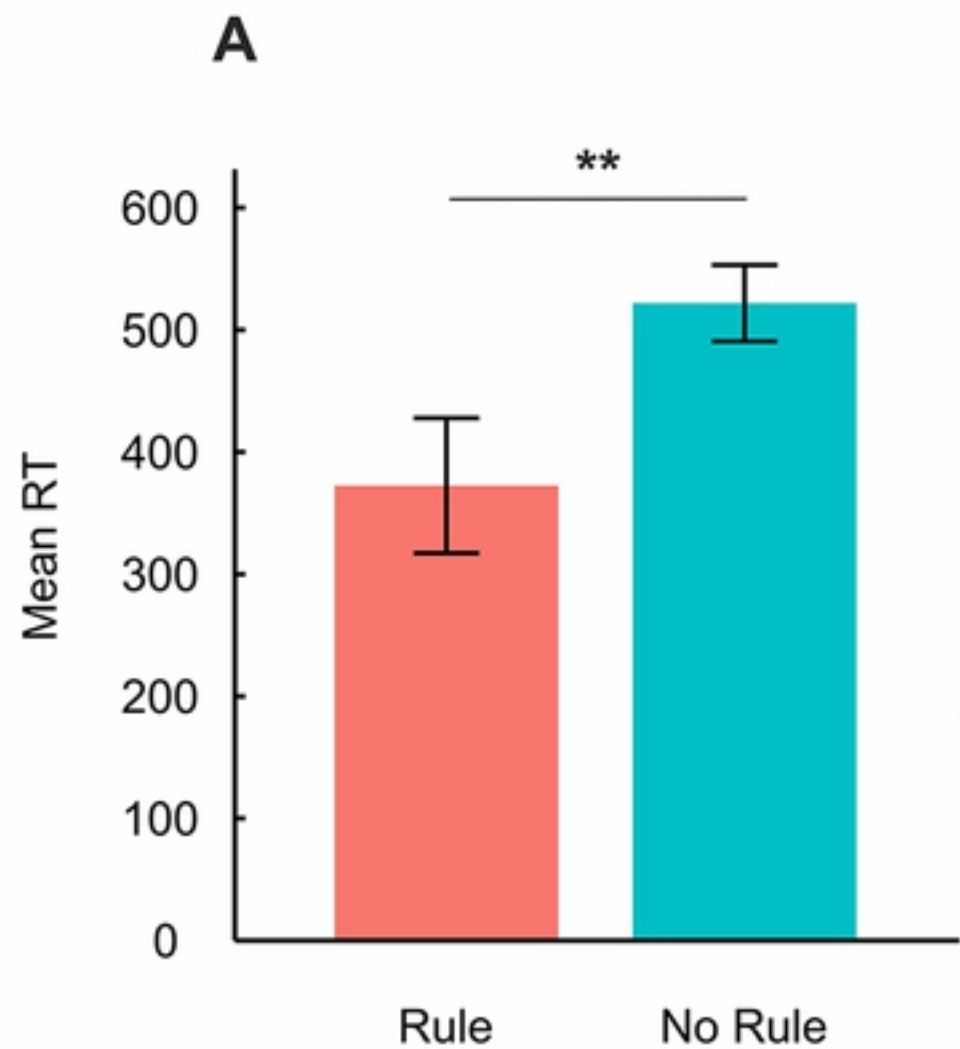
References

1. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science (80-). 1996;247(5294):1926–8.

2. Saffran JR, Wilson DP. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. Infancy. 2003;4(2):273–84.

3. Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999;70:27–52.

4. Gebhart AL, Newport EL, Aslin RN. Statistical Learning of Adjacent and Non-Adjacent Dependencies among Non-Linguistic Sounds. Psychon Bull Rev. 2009;16(3):486–90.

5. Creel SC, Newport EL, Aslin RN. Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. J Exp Psychol Learn Mem Cogn. 2004;30(5):1119–30.

6. Conway CM, Christiansen MH. Statistical Learning Within and Across Modalities: 2005;488–93.

7. Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. J Exp Psychol Learn Mem Cogn [Internet]. 2002;28(3):458–67. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.28.3.458

8. Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. J Exp Psychol Learn Mem Cogn. 2005;31(1):24–39.

9. Gervain J, Macagno F, Cogoi S, Pena M, Mehler J. The neonate brain detects speech structure. Proc Natl Acad Sci [Internet]. 2008;105(37):14222–7. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0806530105

10. Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. Cognition. 2011;121(1):127–32.

11. Meyer T, Olson CR. Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci [Internet]. 2011;108(48):19401–6. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1112895108

12. Toro JM, Trobalón JB. Statistical computations over a speech stream in a rodent. Percept Psychophys. 2005;67(5):867–75.

13. Siegelman N, Bogaerts L, Kronenfeld O, Frost R. Redefining "Learning" in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities? Cogn Sci [Internet]. 2018 Jun;42(1996):692–727. Available from: http://doi.wiley.com/10.1111/cogs.12556

14. Karuza EA, Emberson LL, Aslin RN. Combining fMRI and Behavioral Measures to Examine the Process of Human Learning. Neurobiol Learn Mem. 2014;1–31.

15. Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: Word predictability is the key. Cognition [Internet]. [cited 2019 May 23];114:356–71. Available from: www.tigerspeech.com

16. Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: The paradox of statistical learning. Trends Cogn Sci [Internet]. 2015;19(3):117–25. Available from: http://dx.doi.org/10.1016/j.tics.2014.12.010

17. Sutton RS, Barto AG. Time-Derivative Models of Pavlovian Reinforcement. In: Gabriel M, Moore J, editors. Learning and Computational Neuroscience: Foundations of Adaptive Networks [Internet]. MIT Press; 1990. p. 497–537. Available from: https://www.worldcat.org/title/learning-and-computational-neuroscience-foundations-of-adaptive-networks/oclc/21520873

18. Sutton RS, Barto AG. Reinforcement learning: an introduction. [Internet]. MIT Press. MIT Press; 2018. 1054 p. Available from: http://incompleteideas.net/book/bookdraft2018jan1.pdf%0Ahttp://incompleteideas.net/sutton/book/bookdraft2017june.pdf

19.   White, Robert W. Motivation reconsidered: The concept of competence. Psychol Rev. 1959;66(5):297–333.

20.   Barto AG. Intrinsic Motivation and Reinforcement Learning. In: Baldassarre G, Mirolli M, editors. Intrinsically Motivated Learning in Natural and Artificial Systems. Berlin: Springer-Verlag; 2013. p. 1–458.

21.   Perruchet P, Vinter A. PARSER: A Model for Word Segmentation. J Mem Lang. 1998;39(2):246–63.

22.   Misyak JB, Christiansen MH, Tomblin JB. Sequential expectations: The role of prediction-based learning in language. Top Cogn Sci. 2010;2(1):138–53.

23.   Montague PR, Dayan P, Sejnowski J. A Framework for Mesencephalic Predictive Hebbian Learning. J Neurosci. 1996;76(5):1936–47.

24.   O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable Role of Ventral and Dorsal Striatum in Instrumental Conditioning. Science (80-) [Internet]. 2004;304(5669):452–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15087550

25.   Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport. 2001;12(17):3683–7.

26.   Lauwereyns J, Watanabe K, Coe B, Hikosaka O. A neural correlate of response bias in monkey caudate nucleus. Nat Publ Gr [Internet]. 2002;418:413–7. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4601820

27.   Samejima K, Ueda Y, Doya K, Kimura M. Neuroscience: Representation of action-specific reward values in the striatum. Science (80-). 2005;310(5752):1337–40.

28.   Schönberg T, Daw ND, Joel D, Doherty JPO. Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-. J Neuropsychiatry Clin Neurosci. 2007;27(47):12860–7.

29.   Conway CM, Pisoni DB. Neurocognitive Basis of Implicit Learning of Sequential Structure and Its Relation to Language Processing. Ann N Y Acad Sci. 2008;131(1145):113–31.

30.   De Diego-Balaguer R, Couette M, Dolbeau G, Dürr A, Youssov K, Bachoud-Lévi AC. Striatal degeneration impairs language learning: Evidence from Huntington's disease. Brain. 2008;131(11):2870–81.

31.   Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. The neural correlates of statistical learning in a word segmentation task: An fMRI study. BRAIN Lang. 2013;127(1):1–21.

32.   Plante E, Patterson D, Gómez R, Almryde KR, White MG, Asbjørnsen AE. The nature of the language input affects brain activation during learning from a natural language. J Neurolinguistics. 2015;36:17–34.

33.   Gomez R, Maye J. The Developmental Trajectory of Nonadjacent Dependency Learning The Developmental Trajectory of Nonadjacent Dependency Learning. Infancy. 2005;7(2):183–206.

34.   López-Barroso D, Cucurell D, Rodríguez-Fornells A, de Diego-Balaguer R. Attentional effects on rule extraction and consolidation from speech. Cognition. 2016;152:61–9.

35.   Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non reinforcement. In: Black AH, Prokasy WF, editors. Classical conditioning II: current research and theory [Internet]. 1972. p. 64–99. Available from: papers2://publication/uuid/51EED98C-39D3-4ECA-9CC8-F7E445CCB145

36.   Niv Y, Schoenbaum G. Dialogues on prediction errors. Trends Cogn Sci. 2008;12(7):265–72.

37.   Niv Y, Montague PR. Theoretical and Empirical Studies of Learning. 2008;329–50.

38.   Colder B. The basal ganglia select the expected sensory input used for predictive coding. Front Comput Neurosci [Internet]. 2015;9(September):1–10. Available from:

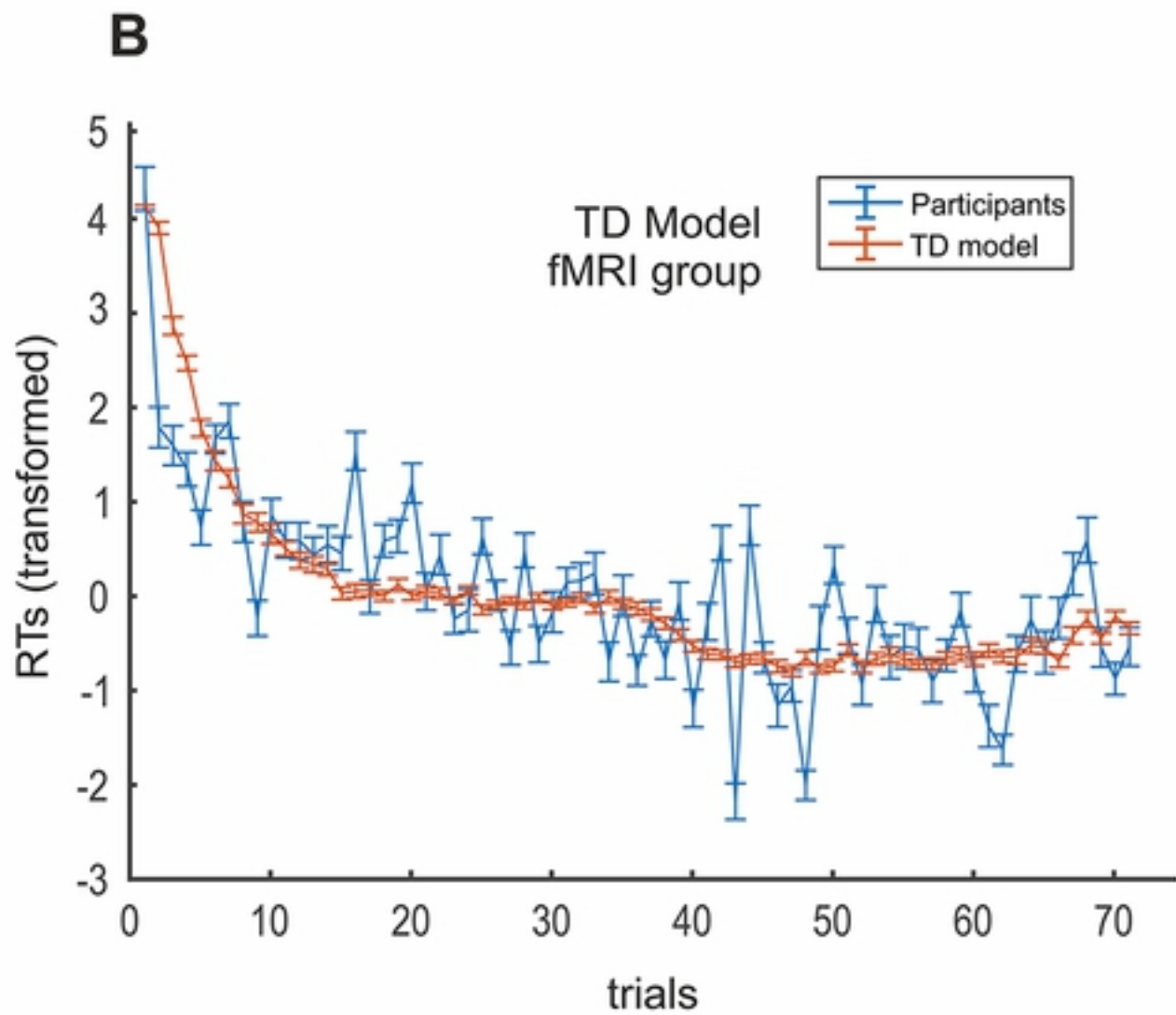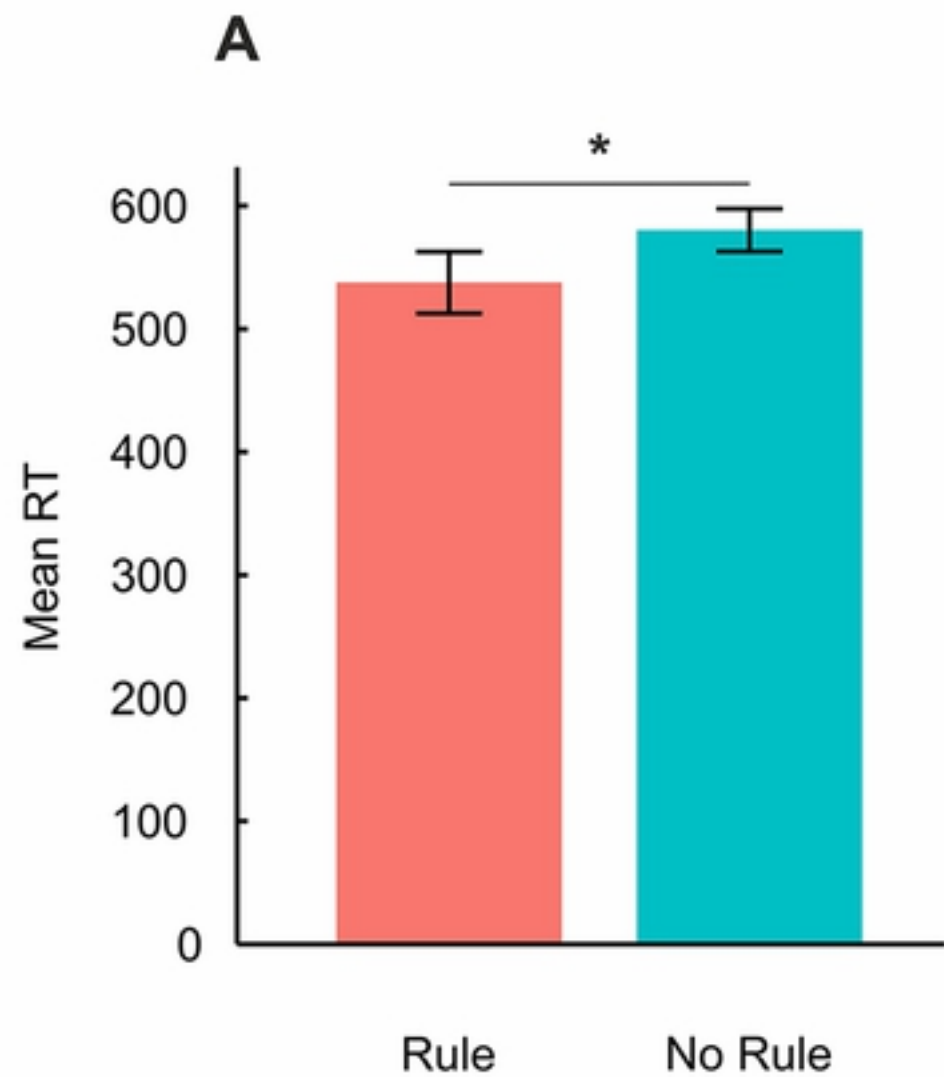http://journal.frontiersin.org/Article/10.3389/fncom.2015.00119/abstract

39. Misyak JB, Christiansen MH, Tomblin JB. On-line individual differences in statistical learning predict language processing. 2010;1(September):1–9.

40. Buiatti M, Peña M, Dehaene-Lambertz G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage [Internet]. 2009;44(2):509–19. Available from: http://dx.doi.org/10.1016/j.neuroimage.2008.09.015

41. Farthouat J, Franco A, Mary A, Delpouve J, Wens V, Op de Beeck M, et al. Auditory Magnetoencephalographic Frequency-Tagged Responses Mirror the Ongoing Segmentation Processes Underlying Statistical Learning. Brain Topogr [Internet]. 2017 Mar 9 [cited 2019 Apr 21];30(2):220–32. Available from: http://link.springer.com/10.1007/s10548-016-0518-y

42. Henin S, Turk-Browne N, Friedman D, Liu A, Dugan P, Flinker A, et al. Statistical learning shapes neural sequence representations. bioRxiv [Internet]. 2019 [cited 2019 Apr 21]; Available from: http://dx.doi.org/10.1101/583856

43. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. Nat Neurosci. 2006;9(8):1057–63.

44. Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci. 2007;10(12):1615–24.

45. Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron. 2004;43(1):133–43.

46. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron [Internet]. 2010;68(5):815–34. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=1993159104&lang=nl&site=ehost-live&scope=site

47. Barto AG. Adaptive critics and the basal ganglia. In: Models of information processing in the basal ganglia. Cambridge, MA, US: The MIT Press; 1995. p. 215–32. (Computational neuroscience.).

48. Doya K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Vol. 12, Neural Networks. 1999. p. 961–74.

49. Frank MJ. Computational models of motivated action selection in corticostriatal circuits. Curr Opin Neurobiol. 2011;21:381–6.

50. Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. 1997;275(MARCH):1593–600.

51. Fiorillo CD, Tobler PN, Schultz W. Discrete Coding of Reward Dopamine Neurons. Science (80-) [Internet]. 2003;299(March):1898–902. Available from: http://www.sciencemag.org/content/299/5614/1898.short

52. O'Reilly RC, Frank MJ. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. 2006;18(2):283–328.

53. Berridge KC, Kringelbach ML. Affective neuroscience of pleasure: reward in humans and animals. [cited 2019 May 23]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3004012/pdf/nihms257672.pdf

54. Ripollés P, Marco-Pallarés J, Hielscher U, Mestres-Missé A, Tempelmann C, Heinze HJ, et al. The role of reward in word learning and its implications for language acquisition. Curr Biol. 2014;24(21):2606–11.

55. Ripollés P, Marco-Pallarés J, Alicart H, Tempelmann C, Rodríguez-Fornells A, Noesselt T. Intrinsic monitoring of learning success facilitates memory encoding via the activation of the SN/VTA-hippocampal loop. Elife. 2016;5(September):1–35.

56. Ripollés P, Ferreri L, Mas-Herrero E, Alicart H, Gómez-Andrés A, Marco-Pallares J, et al. Intrinsically regulated learning is modulated by synaptic dopamine signaling. Elife. 2018;7:1–23.

57.   Singh S, Barto AG, Chentanez N. Intrinsically Motivated Reinforcement Learning. Adv Neural Inf Process Syst. 2004;1–8.

58.   O'Doherty JP. Reward representations and reward-related learning in the human brain: Insights from neuroimaging. Curr Opin Neurobiol. 2004;14(6):769–76.

59.   Zink CF, Pagnoni G, Martin ME, Dhamala M, Berns GS. Human striatal response to salient nonrewarding stimuli. J Neurosci [Internet]. 2003;23(22):8092–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12954871

60.   Shulman GL, Astafiev S V, Franke D, Pope DLW, Snyder AZ, Mcavoy MP, et al. Interaction of Stimulus-Driven Reorienting and Expectation in Ventral and Dorsal Frontoparietal and Basal Ganglia- Cortical Networks. 2009;29(14):4392–407.

61.   Frank MJ, Loughry B, O'Reilly RC. Interactions between frontal cortex and basal ganglia in working memory: A computational model. Cogn Affect Behav Sci [Internet]. 2001;1(2):137–60. Available from: http://www.colorado.edu/ics/sites/default/files/attached-files/00-01.pdf

62.   De Diego Balaguer R, Toro JM, Rodriguez-Fornells A, Bachoud-Lévi AC. Different neurophysiological mechanisms underlying word and rule extraction from speech. PLoS One. 2007;2(11).

63.   De Diego-Balaguer R, Rodríguez-Fornells A, Bachoud-Lévi A-C. Prosodic cues enhance rule learning by changing speech segmentation mechanisms. Front Psychol [Internet]. 2015;6(September):1–11. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.01478/abstract

64.   Alexander G. Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. Annu Rev Neurosci [Internet]. 1986;9(1):357–81. Available from: http://neuro.annualreviews.org/cgi/doi/10.1146/annurev.neuro.9.1.357

65.   Du Y, Buchsbaum BR, Grady CL, Alain C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci [Internet]. 2014;111(19):7126–31. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1318738111

66.   Ullman MT. Contributions of memory circuits to language: The declarative/procedural model. Cognition. 2004;92(1–2):231–70.

67.   Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, et al. The Neural Correlates of Motor Skill Automaticity. 2005;25(22):5356–64.

68.   Cools R, Ivry RB, D'Esposito M. The human striatum is necessary for responding to changes in stimulus relevance. J Cogn Neurosci. 2006;18(12):1973–83.

69.   Opitz B, Friederici AD. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. Neuroimage. 2003;19(4):1730–7.

70.   Tettamanti M, Alkadhi H, Moro A, Perani D, Kollias S, Weniger D. Neural correlates for the acquisition of natural language syntax. Neuroimage. 2002;17(2):700–9.

71.   Campbell JID, Thompson VA. MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. Behav Res Methods. 2012 Dec 22;44(4):1255–65.

72.   Cohen J. Statistical Power Analysis for the Behavioral Sciences. Statistical Power Analysis for the Behavioral Sciences. New York: Lawrence Erlbaum Associates; 1988.

73.   Bush RR, Mosteller F. A mathematical model for simple learning. Psychol Rev. 1951;58(5):313–23.

74.   Marco-Pallarés J, Mohammadi B, Samii A, Münte TF. Brain activations reflect individual discount rates in intertemporal choice. Brain Res [Internet]. 2010;1320:123–9. Available from: http://dx.doi.org/10.1016/j.brainres.2010.01.025

75.   Cohen MX, Ranganath C. Reinforcement Learning Signals Predict Future Decisions. J Neurosci. 2007;27(2):371–8.

76.   Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005;26(3):839–51.
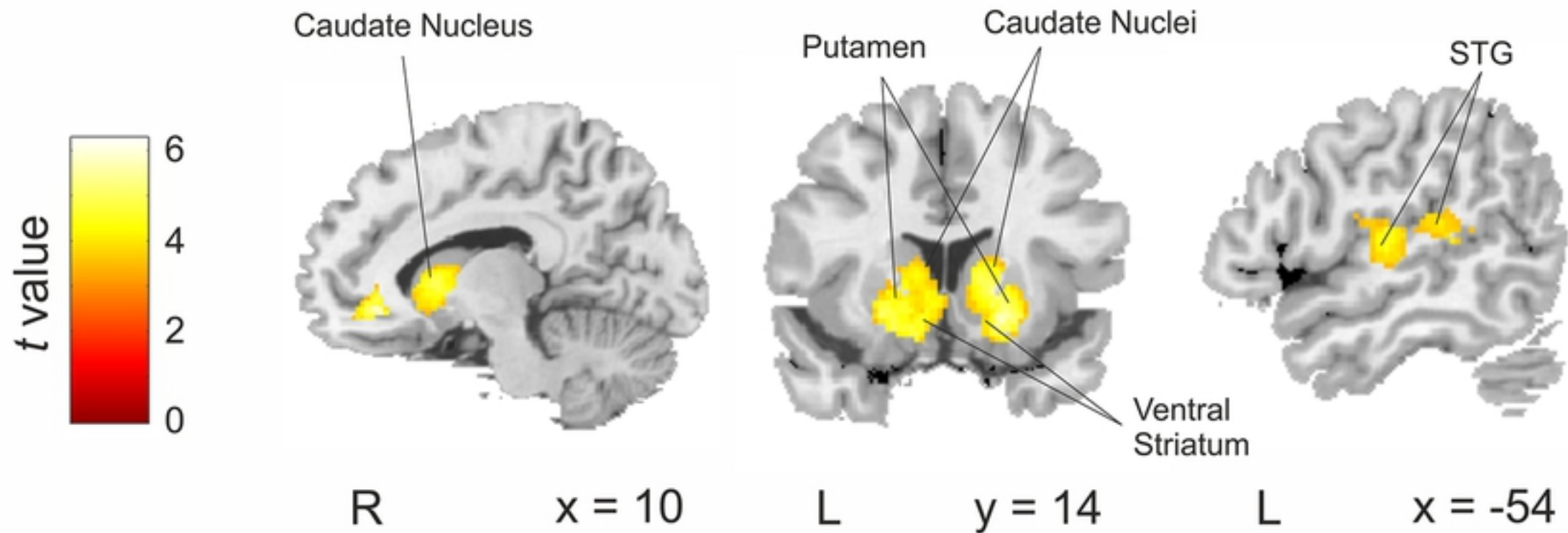
77.   Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;15(1):273–89.

A

B

TD Model
Behavioral group

Figure

Figure

Figure