City Research Online

City, University of London Institutional Repository

Citation: Endress, A. (2019). A simple, biologically plausible feature detector for language acquisition. Journal of Cognitive Neuroscience,

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/id/eprint/23031/

Link to published version:

City Research Online: http://openaccess.city.ac.uk/   publications@city.ac.uk

1

2

3

4

5 **A simple, biologically plausible feature detector for language**

6 **acquisition**

7

8

9 Ansgar D. Endress

10 City, University of London

11

12

13

14

15

16 Running head: A feature detector for language acquisition

17

18

19

20   **Abstract**

21   Language has a complex grammatical system we still have to understand

22   computationally and biologically (Hauser et al., 2002; Yang, 2013). However,

23   some evolutionarily ancient mechanisms have been repurposed for grammar

24   (Dehaene & Cohen, 2007; Endress, Cahill, et al., 2009; Endress, Nespor, et al.,

25   2009; Fitch, 2017) so that we can use insight from other taxa into possible circuit-

26   level mechanisms of grammar. Drawing upon recent evidence for the importance

27   of disinhibitory circuits across taxa and brain regions (Chevalier & Deniau, 1990;

28   Letzkus et al., 2015; Hangya et al., 2014; Xu et al., 2013; Goddard et al., 2014;

29   Mysore & Knudsen, 2012; Koyama et al., 2016; Koyama & Pujala, 2018), I

30   suggest a simple circuit that explains the acquisition of core grammatical rules

31   used in 85% of the world's languages (Rubino, 2013): grammatical rules based on

32   sameness/difference relations. This circuit acts as a sameness-detector. *Different*

33   items are suppressed through inhibition, but presenting two *identical* items leads

34   to inhibition of inhibition. The items are thus propagated for further processing.

35   This sameness-detector thus acts as a feature detector for a grammatical rule. I

36   suggest that having a set of feature detectors for elementary grammatical rules

37   might make language acquisition feasible based on relatively simple

38   computational mechanisms.

39

40   Keywords: Language Acquisition; Rule Learning; Perceptual or Memory

41   Primitives; Disinhibition; Circuit Motifs; Reduplication

42

A simple, biologically plausible feature detector for language acquisition

Language acquisition is fast, largely based on positive evidence (or sometimes no evidence at all; Goldin-Meadow & Mylander, 1998; Senghas et al., 2004), goes far beyond what learners hear or see in their environment (Chomsky, 1959; Pinker, 1984) and results in a uniquely complex grammatical system that stands out in the animal kingdom (Hauser et al., 2002; Yang, 2013). Even seemingly straightforward "memory" problems such as learning the meanings of words hide complexities that call for human-specific grammatical adaptations (Medina, Snedeker, Trueswell, & Gleitman, 2011; Pinker & Jackendoff, 2005). Unsurprisingly, we know very little about the underlying computational mechanisms at the circuit level.

However, some linguistic mechanisms are evolutionarily ancient and have been repurposed for linguistic use (Dehaene & Cohen, 2007; Endress, Cahill, et al., 2009; Endress, Nespor, et al., 2009; Fitch, 2017). In such cases, it might be possible to identify core linguistic mechanism whose systems-level implementation might be tractable due to its evolutionary history.

Here, I use sameness/difference relations as a case in point. I will first show that many grammatical rules are based on such relations, especially in morphology and phonology, but that similar relations are critical in many other domains and animals, suggesting that they reflect a linguistic core mechanism with evolutionarily ancient roots. I will then suggest that such relations can be computed using an ubiquitous processing motif: disinhibition among neurons or neural populations.

67   <u>Sameness/difference relations in language and other domains and animals</u>

68   Sameness/difference relations are critical for many aspects of linguistic

69   structure, especially in phonology and morphology. For example, some 85% of

70   the world's languages use some form of reduplication (Rubino, 2013). Among

71   many other uses, reduplications can signal changes in word class (e.g., from noun

72   to verb, as in the Marshallese contrast between "ta*kin* – sock" and "ta*kinkin* – to

73   wear socks"; Moravcsik, 1978), attenuation (as in the Alabama contrast between

74   "ka*s*atka – cold" and "ká*ss*atka – cool"; Hardy & Montler, 1988) or

75   intensification; they can mark differences in number (e.g., singular vs. plural),

76   tense (e.g., past vs. present), aspect (e.g., continued vs. repeated occurrence or

77   temporary vs. permanent), size or case (see Rubino, 2013, and references therein).

78   Phonological processes also often appeal to sameness/difference relations,

79   with some processes requiring some features to be identical within a relevant

80   constituent, and others requiring them to be different. Processes that require

81   identical features include vowel harmony and assimilation. Specifically, in

82   languages with vowel harmony, vowels within words (or smaller domains) need

83   to have one or more features in common (Rose & Walker, 2011). For example,

84   Hungarian words generally have either only back vowels or only front vowels;

85   grammatical suffixes thus come in two varieties, one with back vowels and one

86   with front vowels. Accordingly, the dative suffix is –*nak* for words like "ablak –

87   window" (resulting in forms like "ablaknak") and –*nek* for words like "bíró –

88   judge" (resulting in forms like "bírónek"; Hayes & Londe, 2006). Likewise, in

89   languages with consonant assimilation, consonants must share a feature with other

90   surrounding consonants. For example, in English, "football" might be pronounced

91   as "foopball" because the place of articulation of the [t] at the end of [foot] gets

92   assimilated to the place of articulation of the [b] at the start of "ball"; in contrast,

93   in French, "football" might be pronounced as "foodball" because the voicing

94   feature of the [t] (but not the place feature) gets assimilated to the following [b]

95   (Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009). Both vowel harmony and

96   assimilation thus introduce sameness relations among phonemes. Listeners use

97   these sameness relations not only in word recognition (Darcy et al., 2009; Mitterer

98   & Blomert, 2003; Suomi, McQueen, & Cutler, 1997), but also as cues to learn

99   new words (Vroomen, Tuomainen, & de Gelder, 1998). Further, sameness

100   relations in the form of vowel harmony often interact with other area of grammar,

101   such as stress assignment or morphology (Rose & Walker, 2011).

102   While vowel harmony and assimilation require sameness relations among

103   phonemic features, other phonological processes impose difference relations.

104   Such processes include the Obligatory Contour Principle (Frisch, Pierrehumbert,

105   & Broe, 2004; McCarthy, 1986). Initially, the Obligatory Contour Principle was

106   proposed to account for the observation that, in certain tone languages, tones

107   cannot be repeated within words, but it also applies to other phonological

108   phenomena. For example, in Semitic languages like Arabic and Hebrew, the basic

109   meaning of verbs is given by their consonantal root; roots like /k t b/ are then

110   transformed into surface forms such as "*katab*a – he wrote" and "*kutib*a – it was

111   written" (Frisch et al., 2004). The OCP prevents consonantal roots from having

112   repeated consonants, while other morphological processes can *create* (rather than

113   prevent) sameness relations among consonants (Frisch et al., 2004; McCarthy,

114   1986). Such rules might also interact with other areas of grammar (Yip, 1988) and

115   speakers apply them even when presented with novel non-sense words (e.g.,

116   Berent & Shimron, 1997; Frisch & Zawaydeh, 2001).

117   Sameness relations are also important during language acquisition.

118   Reduplications are prominent in child-directed speech across languages

119   (Ferguson, 1964) and children themselves "invent" forms with reduplicated

120   syllables; these reduplicated forms might be important for acquiring multisyllabic

121   words (Schwartz, Leonard, Wilcox, & Folger, 1980) and syllable-final consonants

122   that would otherwise be lost (Fee & Ingram, 1982).

123   More generally, sameness relations have been critical for defining the

124   computational complexity of phonological rules (Culy, 1985; Manaster-Ramer,

125   1986), and, in developmental psychology, rules based on sameness relations have

126   been the most prominent assay for studying rule-learning in human infants

127   (Marcus et al., 1999), to the extent that in a recent meta-analysis of "rule-

128   learning" in infancy, rule-learning was treated as synonymous with the learning of

129   sameness relations (Rabagliati, Ferguson, & Lew-Williams, 2019).

130   Sameness relations are also important for other forms of language use. Not

131   only are rhymes and alliterations important in poetry (Fabb, 2015), but many

132   language games that spontaneously arise in children also make extensive use of

133   sameness relations in the form of reduplications (Bagemihl, 1995). For example,

134   in the Chinese May-ka language game, syllables are duplicated and then the

135   vowel of the first duplicate is replaced by "ay" and the consonant of the second

136   duplicate by "k"; *ma* (mother) thus becomes *may-ka* (Bao, 1990; Yip, 1982).

137   Despite their simplicity, sameness relations thus appear to be a core part of

138   the language faculty.

139   However, sameness/difference rules are clearly not specific to language.

140   They are crucial for many other aspects of cognition, including motor learning

141   (Brooks, 1986), any comparison of sensory input to predictions or internal state

142   (e.g., novelty detection in the hippocampus; Kumaran & Maguire, 2007) and

143   short-term memory tasks such as delayed-match to sample tasks (Cope et al.,

144   2018; Engel & Wang, 2011). Accordingly, grammar-like rules based on

145   sameness/difference relations can be learned in many non-linguistic domains in

146   humans (Dawson & Gerken, 2009; Endress, Dehaene-Lambertz, & Mehler, 2007;

147   Marcus, Fernandes, & Johnson, 2007; Saffran, Pollak, Seibel, & Shkolnik, 2007)

148   and by many non-human animals (de la Mora & Toro, 2013; Hauser & Glynn,

149   2009; Martinho & Kacelnik, 2016; Murphy, Mondragon, & Murphy, 2008;

150   Neiworth, 2013; Pepperberg, 1987; Smirnova, Zorina, Obozova, & Wasserman,

151   2015; Versace, Spierings, Caffini, Ten Cate, & Vallortigara, 2017; but see

152   Heijningen, Visser, Zuidema, & Cate, 2009; Hupé, 2017; Langbein & Puppe,

153   2017), possibly through a specialized sameness-detector (Endress, 2013; Endress

154   et al., 2007) that might exist from birth (Antell, Caron, & Myers, 1985; Gervain,

155   Berent, & Werker, 2012; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). The

156   computations underlying sameness/difference relations thus reflect a core

157   linguistic mechanism whose systems-level implementation might be tractable due

158   to its evolutionary history.

159   Disinhibition-based computations

160   Here, drawing upon recent evidence stressing the importance of

161   disinhibitory circuits (neurons that inhibit other inhibitory neurons) across a

162   variety of taxa and brain regions (Chevalier & Deniau, 1990; Goddard et al.,

163   2014; Hangya et al., 2014; Koyama et al., 2016; Mysore & Knudsen, 2012; Xu et

164   al., 2013), I suggest a simple circuit that acts as a sameness-detector. Disinhibition

165   has been observed in a variety of brain areas (Chevalier & Deniau, 1990; Letzkus

166   et al., 2015), and some interneuron populations specifically inhibit other

167   inhibitory interneurons (Hangya et al., 2014; Xu et al., 2013). Critically, some

168   interneuron types receive both local and long-range input; such interneurons have

169   been found to inhibit other inhibitory interneurons in auditory (Pi et al., 2013),

170   visual (Pfeffer, Xue, He, Huang, & Scanziani, 2013), somatosensory (Lee,

171   Kruglikov, Huang, Fishell, & Rudy, 2013) and prefrontal cortex (Pi et al., 2013),

172   from where they can exert spatially remarkably specific disinhibition on other

173   populations (Zhang et al., 2014). Accordingly, Hangya et al. (2014) argued that

174   this disinhibitory circuit might be a cortical circuit motif. Other authors suggested

175   a more local disinhibitory circuit motif with mutual inhibition among inhibitory

176   neurons (Goddard et al., 2014; Koyama et al., 2016; Koyama & Pujala, 2018;

177   Mysore & Knudsen, 2012).

178   Disinhibitory circuits have been proposed to account for a variety of

179   cognitive phenomena, including attentional selection (van Der Velde & de

180   Kamps, 2001; Zhang et al., 2014), gain control (Fu et al., 2014), sequential

181   discriminations of stimulus strength of stimuli (Machens, Romo, & Brody, 2005;

182   Miller & Wang, 2006; but see Barak, Sussillo, Romo, Tsodyks, & Abbott, 2013)

183   categorization of stimuli (Goddard et al., 2014; Kusunoki, Sigala, Nili, Gaffan, &

184   Duncan, 2010; Mysore & Knudsen, 2012), behavioral response selection (Jovanic

185   et al., 2016; Zhao et al., 2019), associative learning (Letzkus et al., 2011),

186   plasticity (Fu, Kaneko, Tang, Alvarez-Buylla, & Stryker, 2015) and social

187 behavior (Marlin, Mitre, D'amour, Chao, & Froemke, 2015; Owen et al., 2013).

188 Here, I suggest that the same biological mechanisms might provide a circuit-level

189 mechanism for a core grammatical computation based on sameness vs. difference

190 computations.

191 <u>Models of sameness/difference relations</u>

192 A number of models of how sameness-relations might be computed have

193 been proposed in the literature (Arena et al., 2013; Carpenter & Grossberg, 1987;

194 Cope et al., 2018; Engel & Wang, 2011; Hasselmo & Wyble, 1997; J. S. Johnson,

195 Spencer, Luck, & Schöner, 2009; Ludueña & Gros, 2013; Wen, Ulloa, Husain,

196 Horwitz, & Contreras-Vidal, 2008). The underlying principles and assumptions

197 vary substantially across models. Some rely on the fact that repeatedly activated

198 representations suffer some form of neural "fatigue" (Grill-Spector, Henson, &

199 Martin, 2006; Kumaran & Maguire, 2007), others on circuitry where the

200 *combined* input from some form of memory and from sensory representations

201 matching (or mismatching) the memory representations must be sufficiently

202 strong (Carpenter & Grossberg, 1987; Hasselmo & Wyble, 1997; Wen et al.,

203 2008) or where the *difference* between input from memory and from sensory

204 representations is the critical variable (Engel & Wang, 2011). Still other models

205 detect reduced levels inhibition for novel compared to previously encountered

206 items (Cope et al., 2018; J. S. Johnson et al., 2009). I discuss these models in

207 more detail in Supplementary Material 1, where I show that they fall short on at

208 least one of two criteria of grammar learning: they either do not generalize to

209 unseen exemplars or they require labeled counter-examples.

210 To better illustrate the computational principles underlying the current

211 dishibition-based circuit, I will first present a version of the model that can detect

212 sameness relations in sequentially presented stimuli. Following this, I will sketch

213 a version of the model that can detect sameness relations in spatially distributed,

214 simultaneously presented stimuli, and finally a model that can detect sameness

215 relations in both simultaneously presented stimuli and sequentially presented

216 stimuli.

217 **Results**

218 <u>Sameness detection for sequential stimuli</u>

219 Figure 1a shows a possible disinhibition-based architecture of how

220 sameness might be detected for sequentially presented items. (Model equations

221 are given in Appendix A; an R implementation is available online). The model

222 comprises two populations of neurons (hereafter "layers") that encode features of

223 items (e.g., frequency, color and so on; in Figure 1, the features are represented as

224 geometric shapes).

225 The *source layer* receives input; input can be sensory or non-sensory,

226 depending on where this circuit is located in the brain. Units in the *copy layer*

227 receive excitatory one-to-one input from units in the source layer that code for the

228 same feature. However, they also receive feature-specific tonic inhibition from an

229 *inhibition layer* (which might consist of interneurons); tonic inhibition has been

230 observed in a variety of brain regions, and might subserve functions such as

231 maintaining an appropriate level of excitability or the suppression of undesirable

232 motor programs (Benjamin, Staras, & Kemenes, 2010; Farrant & Nusser, 2005;

233 Semyanov, Walker, Kullmann, & Silver, 2004).

234 Due to the inhibition from the inhibition layer to the copy layer, input

235   from the source layer is not propagated to the copy layer with a single stimulation.

236   The critical aspect of this circuit is that each feature in the source layer also

237   *inhibits* the corresponding feature in the inhibition layer, which, in turn, reduces

238   inhibitory input to the copy layer for that feature. A similar phenomenon has been

239   observed in auditory fear conditioning, where inhibition of (inhibitory)

240   parvalbumin-positive interneurons allowed for associations between sounds and

241   aversive stimuli to be formed (Letzkus et al., 2011).

242   Accordingly, once the inhibitory input to the copy layer ceases, there will

243   be a time window during which the excitatory input from the source layer can

244   drive the corresponding units in the copy layer. As a result, only repeated items

245   will be propagated to the copy layer. Any readout mechanism for the copy layer

246   (e.g., a population of thresholded neurons) could thus act as a sameness-detector.[1]

247

248

[1] While I model disinhibition across different neural populations, the same computational principles could be implemented using reciprocal inhibition among inhibitory neurons as in earlier models of stimulus selection and categorization (Goddard, Mysore, Bryant, Huguenard, & Knudsen, 2014; Koyama et al., 2016; Koyama & Pujala, 2018; Mysore & Knudsen, 2012). To do so, one would simply replace the inhibitory connections from the source layer to the inhibition layer with inhibition in the source layer that is itself subject to lateral inhibition.
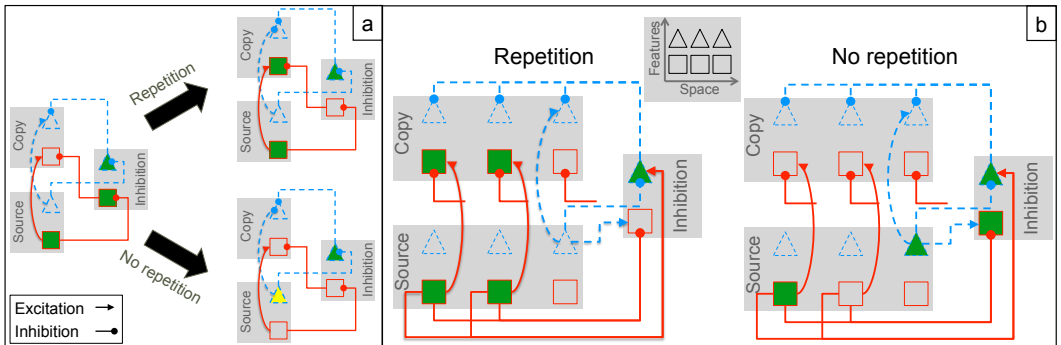
249

250 **Figure 1: A disinhibition-based sameness-detector for (a) sequentially (b) simultaneously**

251 **presented identical items. The geometric shapes (squares and triangles) stand for**

252 **populations of neurons that encode features of the items (e.g., frequency, shape etc.); filled**

253 **shapes are currently active while empty shapes are currently inactive. (a) Units in the *source***

254 ***layer* (bottom gray box) receive (sensory or other) input. Units in the *copy layer* (top gray**

255 **box) receive one-to-one excitatory input from the source layer. Critically, units from the**

256 ***inhibition layer* (right gray box) exert tonic inhibition on the copy layer. (a, left) Upon initial**

257 **presentation of a feature (represented here as a square), all units in the inhibition layer are**

258 **active. As a result, excitatory input from the source layer is not propagated to the copy layer.**

259 **(a, right, top) Feature-specific inhibition from the source layer to the corresponding units in**

260 **the inhibition layer shuts down the inhibitory input to the copy layer. If the same item is**

261 **presented again during the time window of reduced inhibition, input from the source layer is**

262 **propagated to the copy layer. (a, right, bottom) If a new, non-identical item is presented, the**

263 **source layer cannot drive the copy layer because the corresponding units in the inhibition**

264 **layer have not been inhibited. Sameness-detection thus proceeds by reading out the copy**

265 **layer, as only repeated items are propagated to the copy layer. (b) Sameness-detection in**

266 **simultaneously presented, spatially arranged items. The source layer consists of populations**

267 **of neurons coding for features (arranged in the y-direction), but these units encode space as**

well (arranged in the x-direction). Tonically active inhibitory (inter-)neurons (small gray box on the right) prevent activation in the copy layer (top gray box). Critically, they receive inhibitory input from those units in the source layer that code for the same feature, and excitatory input from units coding for other features. For example, units representing squares in the input layer inhibit all units representing squares in the inhibition layer, and excite all other units. (b, left) If the stimuli consist of two identical items (squares), the combined inhibitory input from the identical items in the source layer shuts down the corresponding units in the inhibition layer, which lets identical items "pass through" to the copy layer (b, right) In contrast, when the stimuli consist of two different items, these singleton features are insufficient to drive the copy population due to inhibition from the inhibition layer.

280   I simulated this model at various levels of noise; at each noise level, I ran

281   50 simulations, representing 50 virtual participants. Figure 2 (left) shows that, in

282   the copy layer, activation for repeated features is high, while activation for non-

283   repeated features is low. Repeated items are thus highly discriminable from non-

284   repeated items. This result is robust to the simulated noise level. A simple

285   disinhibition-based circuit can thus act as a sameness-detector that discriminates

286   repeated features from not repeated features.

287   While the primary goal of this model is to detect when two temporarily

288   adjacent items are identical, whether or not it can detect the sameness of two

289   objects with intervening material depends on the time constants of the

290   disinhibitory effects. If disinhibition is sufficiently long-lasting, the model will

291   also detect the sameness of two non-adjacent items (e.g., of the two *A*'s in the

292   sequence *ABA*). If so, it would predict that, the further two items are separated (in

293   terms of the amount of intervening time and/or the number of intervening items,

294   which might or might not have separable effects), the harder it should become to

295   detect the sameness of the two items. At least in infants, it might be harder to

296   detect non-adjacent repetitions compared to adjacent repetitions (S. P. Johnson et

297   al., 2009; Kovács & Mehler, 2008, 2009).

298   That being said, the separation of two items is unlikely to be the only

299   determinant of how it easy it is to detect whether they are the same. For example,

300   in a longer sequence like *ABCDEDFGA*, the two *A*'s are further apart than the

301   two *D*'s. Still, it might easier to detect the sameness of the two *A*'s than of the two

302   *D*'s despite their greater distance because initial and final items are more salient

303   than medial items (Benavides-Varela & Mehler, 2015; Endress, Scholl, & Mehler,

304 2005). As a result, the representations of initial items are likely stronger than

305 those of medial items and thus create stronger and longer-lasting disinhibition.

306 However, the goal of the current model is just to show that a simple and

307 ubiquitous mechanism such as disinhibition can serve as the basis of a sameness

308 detector, while more detailed predictions require a biophysically more realistic

309 model.

310
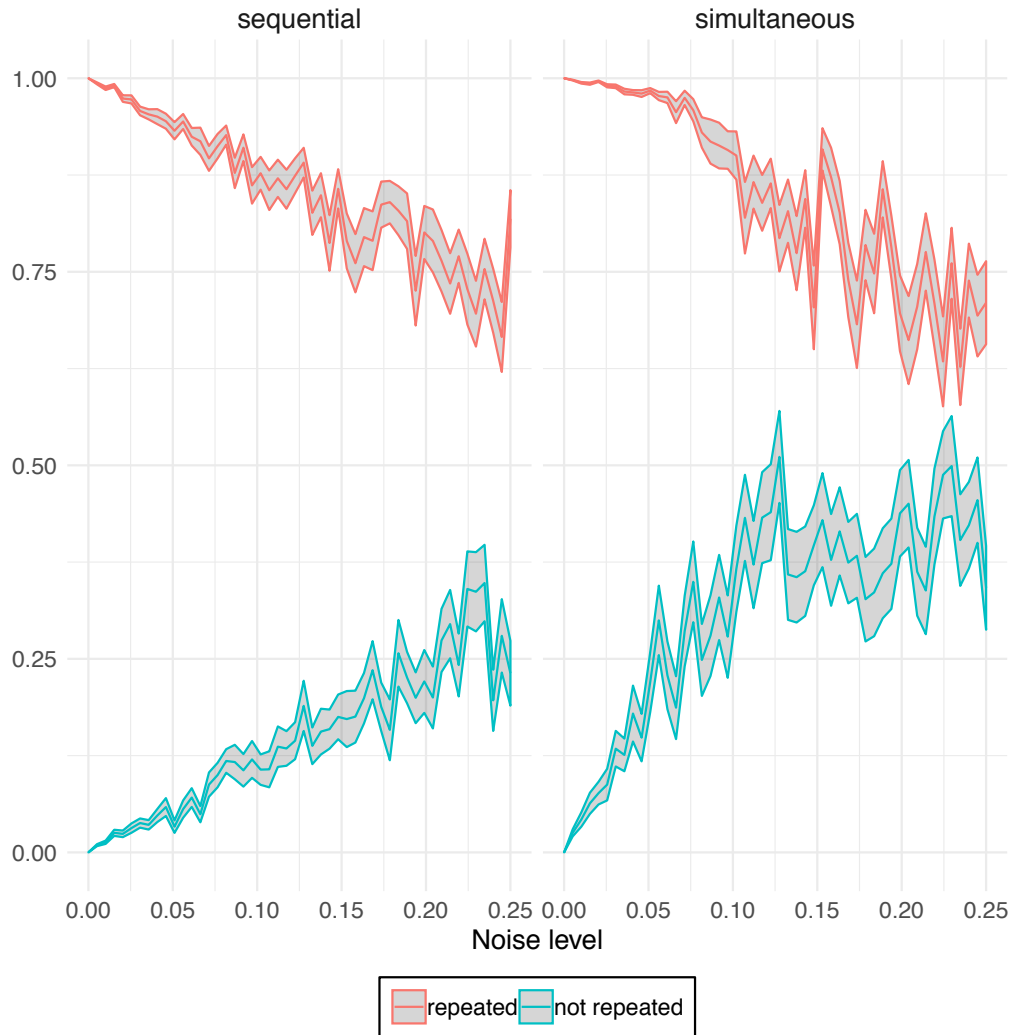
Figure 2: Activation of repeated or non-repeated items in the copy layer. The noise level is the standard deviation of normally distributed noise centered at zero. In each curve, the middle line shows the average activation across 50 simulations, representing 50 participants. The shaded areas represent standard errors from the mean. (Top) Activation in the models shown in Figure 1 that detect either sequentially (Figure 1a) or simultaneously presented (Figure 1b) identical items. (Left). In the *sequential* sameness-detector (Figure 1a), the

319   **activity of repeated items is highly discriminable from that from non-identical items even for**

320   **high noise levels. (Right). In the *simultaneous* sameness-detector (Figure 1b), the activity of**

321   **repeated items is highly discriminable from that of non-repeated items even for high noise**

322   **levels.**

323

324

<u>Sameness detection for simultaneous stimuli</u>

326   In its current stage, the model can detect the sameness of sequentially presented

327   stimuli, but not of spatially distributed, simultaneously presented stimuli, simply

328   because space is not represented. Figure 1b shows a version of the model where

329   items are presented simultaneously rather than sequentially. Again, there is a

330   source layer, a copy layer, and an inhibition layer. The model differs from the

331   sequential model in three critical aspects. First, all layers now represent space. In

332   Figure 1b, the vertical axis represents the features as before, while the horizontal

333   axis represents the spatial locations of the items (though space is presumably

334   represented in some topological order in real neuronal populations). This change

335   is necessary so that two simultaneously presented identical objects can be

336   represented.

337   Second, the connectivity between the source layer and the inhibition layer

338   has been changed. Units in the source layer send (i) inhibitory input to all units in

339   the inhibition layer that code for the same feature across all locations and (ii)

340   excitatory input to all units in the inhibition layer that code for different features;

341   in other words, there is center-surround disinhibition among features. This ensures

342   that, in the copy layer, different-feature input from the source layer stays

343   inhibited, while same-feature input is disinhibited.

344   Third, the sequential model needs to update the activation of the copy

345   layer before that of the inhibition layer; if the inhibition layer were updated first, a

346   single presentation of a feature would be sufficient to produce disinhibition. In

347   contrast, the simultaneous model needs to update the inhibition layer before the

348   copy layer; if the copy layer were updated first, there would be no disinhibition

349   for identical features.

350   I simulated this architecture using 50 virtual participants. As shown in

351   Figure 2, identical items are highly discriminable from non-identical items even at

352   high levels of noise. A simple, disinhibition-based circuit can thus detect

353   sameness relations among simultaneously presented identical objects.

354   A combined model of sameness detection for simultaneous and sequential stimuli

355   While the main differences between the sequential and the simultaneous

356   circuit are simply due to how stimuli are presented (i.e., spatial representations

357   and lateral inhibition among features could be added to the sequential model, but

358   are not necessary), the different update orders raise the question of whether a

359   combined model can be developed that detects both sequential and simultaneous

360   sameness relations. Practically speaking, sequential and simultaneous presentation

361   might not be as different as they seem. For example, if observers attend

362   simultaneously presented items one after the other (Liu & Becker, 2013; Vogel,

363   Woodman, & Luck, 2006; but see Mance, Becker, & Liu, 2012), we need a

364   sequential model to account for *simultaneous* sameness-detection; conversely, if

365   sequential items are placed in some kind of (short-term) memory before being

366   compared, we need a simultaneous model for sameness-detection in sequentially

367   presented items. As such, a combined sequential/simultaneous model might be

368   neither necessary nor desirable.

369   Be that as it might, such a combined model is shown in Figure 3.

Excitation ⟶
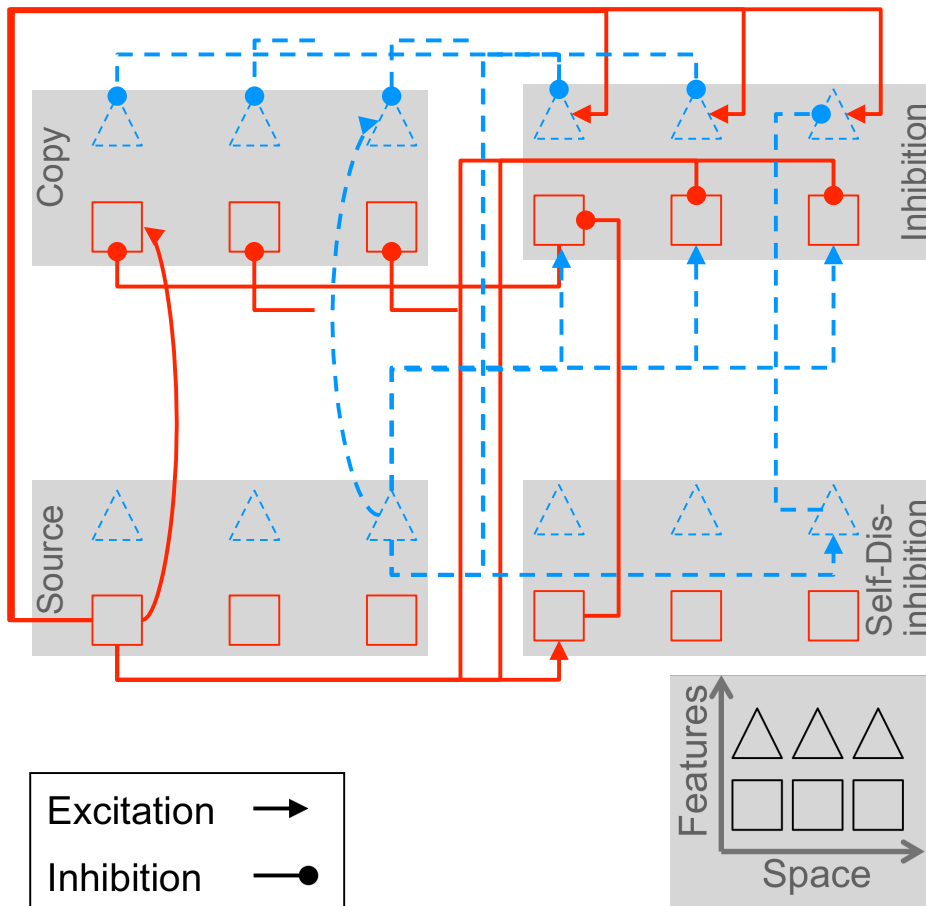
Inhibition ⟶●

Figure 3: Combined disinhibition-based sameness-detector for both sequential and simultaneous sameness relations. As in the simultaneous circuit from Figure 1b, the source layer (bottom left gray box) consists of populations of neurons coding for features (arranged in the y-direction) and spatial locations (arranged in the x-direction). Tonically active units in the inhibition layer (top right gray box) prevent activation in the copy layer (top left gray box). Units in the inhibition layer receive (i) inhibitory input from the source layer for units coding for the same feature and (ii) excitatory input for units coding for other features, leading to center-surround disinhibition among features and, in the copy layer, to inhibition

380   **for different-feature input and disinhibition for same-feature input. Critically, and in**

381   **contrast to the simultaneous model from Figure 1b, units in the source layer do not inhibit**

382   **units in the inhibition layer that code for features at their own spatial location; they**

383   **disinhibit features only at other locations. To obtain disinhibition at the spatial location of a**

384   **given unit, a self-inhibition layer (bottom right gray box) was added that receives one-to-one**

385   **input from the source layer, and that specifically inhibits units in the inhibition layer that**

386   **code for the same feature *at the same spatial location*. This delays same-feature/same-location**

387   **disinhibition to prevent a single sequential presentation of a feature from disinhibiting that**

388   **feature.**

389

390   This "combined" sameness-detector is similar to the simultaneous

391   sameness-detector in that it comprises a source layer, a copy layer and an

392   inhibition layer, and that the copy layer receives excitatory input from the source

393   layer. However, (dis-)inhibition is organized differently. The copy layer still

394   receives tonic inhibition from those units in the inhibition layer that code for the

395   same feature and spatial position. Further, each feature of the input layer inhibits

396   the corresponding feature in the inhibition layer across spatial positions (i.e., it

397   disinhibits this feature in the copy layer), and excites all other features.

398   The critical difference is that disinhibition of features at the same location

399   is delayed. To do so, I removed direct connections between the source layer and

400   the inhibition layer that coded for the same feature at the same location (while

401   keeping the center-surround disinhibition at other locations). Instead, I added a

402   *self-disinhibition layer* where each unit (i) receives excitatory input from the

403   corresponding feature and location in the source layer and (ii) sends inhibitory

404   input to all units coding for the same feature (across locations) in the inhibition

405   layer. (While these modifications might seem to some extent *ad-hoc*, as

406   mentioned above, it is not clear if a combined sequential/simultaneous model is

407   necessary or desirable in the first place.)

408   As shown in Figure 4, identical items were highly discriminable from non-

409   identical items in the simultaneous situation across noise levels; in contrast, in the

410   sequential situation, discriminability suffered as noise increased.
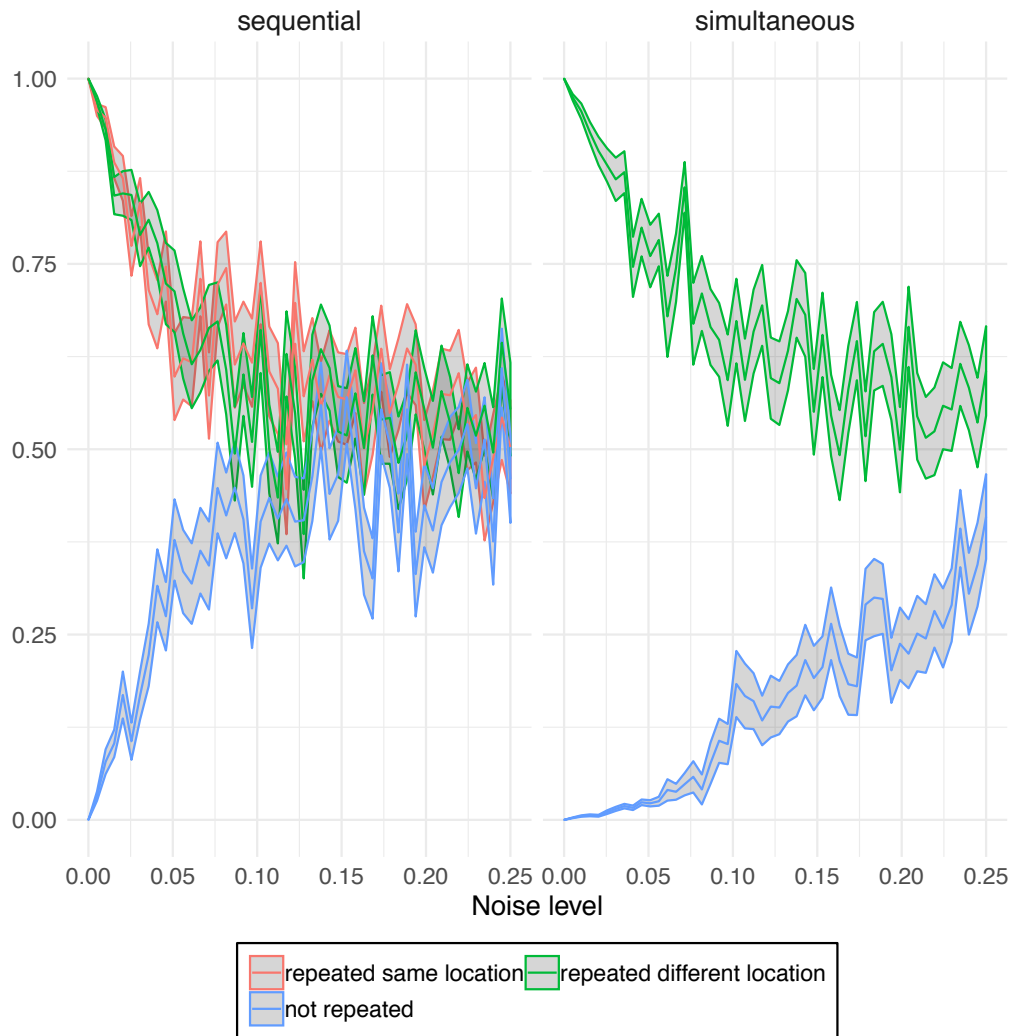
Figure 4: Activation in the copy layer of the combined sequential/simultaneous sameness-detector (Figure 3). (Left) In the combined sequential/simultaneous sameness-detector, repeated features can be repeated either at the same location or at a different location. While activation of (same or different location) repeated items is highly discriminable from activation for non-repeated items for moderate noise levels, discriminability becomes much poorer at high noise levels, when the standard deviation of the noise reaches about 15% of the activation level of active neurons. (Right) The combined sequential/simultaneous

419   **sameness-detector (Figure 3) shows that the activation in the copy layer is highly**

420   **discriminable between simultaneously repeated items and non-repeated items, even for high**

421   **noise levels.**

422   ## Discussion

423   The current results thus show that a simple and biologically realistic

424   circuit can support a core grammatical computation that is used in more than 80%

425   of the world's languages: grammatical rules based on sameness/difference

426   relationships. In this circuit, non-identical items are filtered out through tonic

427   inhibition as well as center-surround inhibition. In contrast, when identical items

428   are presented sequentially or simultaneously, inhibition is inhibited; this

429   disinhibition of identical items then allows them to be propagated for further

430   processing.

431   Unlike previous models of sameness-detection (Arena et al., 2013;

432   Carpenter & Grossberg, 1987; Cope et al., 2018; Engel & Wang, 2011; Hasselmo

433   & Wyble, 1997; Johnson, Spencer, Luck, & Schöner, 2009; Ludueña & Gros,

434   2013; Wen, Ulloa, Husain, Horwitz, & Contreras-Vidal, 2008; see Supplementary

435   Material 1), the model satisfies critical criteria of grammar acquisition: (1) It

436   generalizes to unseen stimuli and (2) does not require any labeled

437   counterexamples for learning, simply because this circuit architecture does not

438   require any learning at all.

439   Once such a sameness-detector is available, it can be used for building

440   more complex grammatical rules. For example, after exposure to syllable

441   sequences such as *dubaba*, seven-month-olds notice that the last two syllables are

442   identical, and generalize this sameness-relation to new items (Marcus et al.,

443 1999). Critically, they do not only have to detect the sameness relation between

444 the last two syllables, but also have to associate it with the *correct serial* position

445 (Endress et al., 2007; Gervain et al., 2012). Once a sameness-detector is available,

446 it can form associations with representations of sequential positions or other

447 stimuli (Kabdebon & Dehaene-Lambertz, 2019), allowing learners to acquire

448 more complex, composite rules, which is one of the hallmarks of complex

449 cognition (Corballis, 2014; Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015;

450 Fitch & Martins, 2014; Hauser & Watumull, 2017).

451   This, in turn, suggests a fundamentally new view on language acquisition.

452 Learners might be equipped with a potentially large number of potentially

453 complex detectors for a variety of rules that act as feature detectors for

454 grammatical rules (Endress, Nespor, et al., 2009). Learning then involves

455 combining these features, potentially through the use of associative mechanisms.

456 This would be consistent with results from formal language theory, where suitable

457 pre-processing (e.g., through feature detectors) can reduce the complexity of the

458 required computational mechanism. For example, a finite state automaton

459 operating on trees can recognize context-free languages (Morgan, 1986) and even

460 humble rules based on sameness relations can be shown to be beyond the reach of

461 even context-free grammars (Culy, 1985; Manaster-Ramer, 1986).

462   Feature detectors for elementary grammatical rules might thus expand the

463 range of grammars that even simple learning mechanisms (such as associative

464 mechanisms) can learn, which, in turn might make language acquisition feasible

465 using relatively simple computational machinery.

466

467   **References**

468   Antell, S. E., Caron, A. J., & Myers, R. S. (1985). Perception of relational

469   invariants by newborns. *Developmental Psychology*, *21*(6), 942–948.

470   Arena, P., Patané, L., Stornanti, V., Termini, P. S., Zäpf, B., & Strauss, R. (2013).

471   Modeling the insect mushroom bodies: Application to a delayed match-to-

472   sample task. *Neural Networks*, *41*, 202–211.

473   https://doi.org/10.1016/j.neunet.2012.11.013

474   Bagemihl, B. (1995). Language games and related areas. In J. A. Goldsmith (Ed.),

475   *Handbook of Phonological Theory* (1st ed., pp. 697–712). Cambridge,

476   MA: Blackwell.

477   Bao, Z. (1990). Fanqie Languages and Reduplication. *Linguistic Inquiry*, *21*(3),

478   317–350.

479   Barak, O., Sussillo, D., Romo, R., Tsodyks, M., & Abbott, L. F. (2013). From

480   fixed points to chaos: Three models of delayed discrimination. *Progress in*

481   *Neurobiology*, *103*, 214–222.

482   https://doi.org/10.1016/j.pneurobio.2013.02.002

483   Benavides-Varela, S., & Mehler, J. (2015). Verbal Positional Memory in 7-

484   Month-Olds. *Child Development*, *86*(1), 209–223.

485   https://doi.org/10.1111/cdev.12291

486   Benjamin, P. R., Staras, K., & Kemenes, G. (2010). What roles do tonic inhibition

487   and disinhibition play in the control of motor programs? *Frontiers in*

488   *Behavioral Neuroscience*, *4*, 30. https://doi.org/10.3389/fnbeh.2010.00030

489   Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence

490   from the obligatory contour principle. *Cognition*, *64*(1), 39–72.

491　Brooks, V. B. (1986). How does the limbic system assist motor learning? A

492　　　limbic comparator hypothesis. *Brain, Behavior and Evolution*, *29*(1–2),

493　　　29–53.

494　Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a

495　　　self-organizing neural pattern recognition machine. *Computer Vision,*

496　　　*Graphics, and Image Processing*, *37*(1), 54–115.

497　　　http://dx.doi.org/10.1016/S0734-189X(87)80014-2

498　Chevalier, G., & Deniau, J. M. (1990). Disinhibition as a basic process in the

499　　　expression of striatal functions. *Trends in Neurosciences*, *13*(7), 277–280.

500　Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*,

501　　　*35*(1), 26–58.

502　Cope, A. J., Vasilaki, E., Minors, D., Sabo, C., Marshall, J. A. R., & Barron, A. B.

503　　　(2018). Abstract concept learning in a simple neural network inspired by

504　　　the insect brain. *PLoS Computational Biology*, *14*(9), e1006435.

505　　　https://doi.org/10.1371/journal.pcbi.1006435

506　Corballis, M. C. (2014). *The Recursive Mind: The Origins of Human Language,*

507　　　*Thought, and Civilization*. Princeton, NJ: Princeton University Press.

508　Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and*

509　　　*Philosophy*, *8*(3), 345–351. https://doi.org/10.1007/BF00630918

510　Darcy, I., Ramus, F., Christophe, A., Kinzler, K. D., & Dupoux, E. (2009).

511　　　Phonological knowledge in compensation for native and non-native

512　　　assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation*

513　　　*and Gradience in Phonetics and Phonology* (pp. 265–309). Berlin:

514　　　Mouton De Gruyter.

515 Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity:

516   4-month-olds learn an abstract repetition rule in music that 7-month-olds

517   do not. *Cognition*, *111*(3), 378–382.

518   https://doi.org/10.1016/j.cognition.2009.02.010

519 de la Mora, D. M., & Toro, J. M. (2013). Rule learning over consonants and

520   vowels in a non-human animal. *Cognition*, *126*(2), 307–312.

521   https://doi.org/10.1016/j.cognition.2012.09.015

522 Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*,

523   *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

524 Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The

525   Neural Representation of Sequences: From Transition Probabilities to

526   Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2–19.

527   https://doi.org/10.1016/j.neuron.2015.09.019

528 Endress, A. D. (2013). Bayesian learning and the psychology of rule induction.

529   *Cognition*, *127*(2), 159–176.

530   https://doi.org/10.1016/j.cognition.2012.11.014

531 Endress, A. D., Cahill, D., Block, S., Watumull, J., & Hauser, M. D. (2009).

532   Evidence of an evolutionary precursor to human language affixation in a

533   nonhuman primate. *Biol Lett*, *5*(6), 749–751.

534 Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual

535   constraints and the learnability of simple grammars. *Cognition*, *105*(3),

536   577–614.

537 Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory

538   constraints on language acquisition. *Trends Cogn Sci*, *13*(8), 348–353.

539 Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the

540   extraction of algebraic rules. *J Exp Psychol Gen*, *134*(3), 406–419.

541 Engel, T. A., & Wang, X.-J. (2011). Same or different? A neural circuit

542   mechanism of similarity-based pattern match decision making. *Journal of*

543   *Neuroscience*, *31*(19), 6982–6996.

544   https://doi.org/10.1523/JNEUROSCI.6150-10.2011

545 Fabb, N. (2015). *What is Poetry? Language and Memory in the Poems of the*

546   *World*. https://doi.org/10.1017/CBO9780511736575

547 Farrant, M., & Nusser, Z. (2005). Variations on an inhibitory theme: Phasic and

548   tonic activation of GABA(A) receptors. *Nature Reviews. Neuroscience*,

549   *6*(3), 215–229. https://doi.org/10.1038/nrn1625

550 Fee, J., & Ingram, D. (1982). Reduplication as a strategy of phonological

551   development. *J Child Lang*, *9*(1), 41–54.

552 Ferguson, C. (1964). Baby talk in six languages. *American Anthropologist*, *66*,

553   103–14.

554 Fitch, W. T. (2017). Empirical approaches to the study of language evolution.

555   *Psychonomic Bulletin & Review*, *24*(1), 3–33.

556   https://doi.org/10.3758/s13423-017-1236-5

557 Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music,

558   language, and action: Lashley revisited. *Annals of the New York Academy*

559   *of Sciences*, *1316*, 87–104. https://doi.org/10.1111/nyas.12406

560 Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity Avoidance

561   and the OCP. *Nat Lang Ling Theory*, *22*(1), 179–228.

562 Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place

563   in Arabic. *Language*, *77*, 91–106.

564 Fu, Y., Kaneko, M., Tang, Y., Alvarez-Buylla, A., & Stryker, M. P. (2015). A

565   cortical disinhibitory circuit for enhancing adult plasticity. *ELife*, *4*,

566   e05558. https://doi.org/10.7554/eLife.05558

567 Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A.,

568   … Stryker, M. P. (2014). A cortical circuit for gain control by behavioral

569   state. *Cell*, *156*(6), 1139–1152. https://doi.org/10.1016/j.cell.2014.01.050

570 Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn

571   brain detects identity relations and sequential position in speech. *Journal*

572   *of Cognitive Neuroscience*, (Early Access), 1–11.

573 Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate

574   brain detects speech structure. *Proc Natl Acad Sci U S A*, *105*(37), 14222–

575   14227. https://doi.org/10.1073/pnas.0806530105

576 Goddard, C. A., Mysore, S. P., Bryant, A. S., Huguenard, J. R., & Knudsen, E. I.

577   (2014). Spatially reciprocal inhibition of inhibition within a stimulus

578   selection network in the avian midbrain. *PloS One*, *9*(1), e85865.

579   https://doi.org/10.1371/journal.pone.0085865

580 Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created

581   by deaf children in two cultures. *Nature*, *391*(6664), 279–281.

582   https://doi.org/10.1038/34646

583 Grill-Spector, K., Henson, R. N., & Martin, A. (2006). Repetition and the brain:

584   Neural models of stimulus-specific effects. *Trends Cogn Sci*, *10*(1), 14–

585   23. https://doi.org/10.1016/j.tics.2005.11.006

586 Hangya, B., Pi, H.-J., Kvitsiani, D., Ranade, S. P., & Kepecs, A. (2014). From

587   circuit motifs to computations: Mapping the behavioral repertoire of

588   cortical interneurons. *Current Opinion in Neurobiology*, *26*, 117–124.

589   https://doi.org/10.1016/j.conb.2014.01.007

590   Hardy, H. K., & Montler, T. (1988). Imperfective Gemination in Alabama.

591   *International Journal of American Linguistics*, *54*, 399–475.

592   https://doi.org/10.1086/466094

593   Hasselmo, M. E., & Wyble, B. P. (1997). Free recall and recognition in a network

594   model of the hippocampus: Simulating effects of scopolamine on human

595   memory function. *Behavioural Brain Research*, *89*(1–2), 1–34.

596   Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language:

597   What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–

598   1579. https://doi.org/10.1126/science.298.5598.1569

599   Hauser, M. D., & Glynn, D. (2009). Can free-ranging rhesus monkeys (Macaca

600   mulatta) extract artificially created rules comprised of natural

601   vocalizations? *J Comp Psychol*, *123*(2), 161–167.

602   https://doi.org/10.1037/a0015584

603   Hauser, M. D., & Watumull, J. (2017). The Universal Generative Faculty: The

604   source of our expressive power in language, mathematics, morality, and

605   music. *Journal of Neurolinguistics*, *43*, 78–94.

606   https://doi.org/10.1016/j.jneuroling.2016.10.005

607   Hayes, B., & Londe, Z. C. (2006). Stochastic phonological knowledge: The case

608   of Hungarian vowel harmony. *Phonology*, *23*, 59–104.

609   Heijningen, C. A. A. van, Visser, J. de, Zuidema, W., & Cate, C. ten. (2009).

610   Simple rules can explain discrimination of putative recursive syntactic

611   structures by a songbird species. *Proc Natl Acad Sci U S A*, *106*(48),

612   20538–20543. https://doi.org/10.1073/pnas.0908113106

613 Hupé, J.-M. (2017). Comment on "Ducklings imprint on the relational concept of

614   'same or different.'" *Science*, *355*(6327), 806.

615   https://doi.org/10.1126/science.aah6047

616 Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic

617   neural field model of visual working memory and change detection.

618   *Psychological Science*, *20*(5), 568–577. https://doi.org/10.1111/j.1467-

619   9280.2009.02329.x

620 Johnson, S. P., Fernandas, K. J., Frank, M. C., Kirkham, N., Marcus, G.,

621   Rabagliati, H., & Slemmer, J. A. (2009). Abstract Rule Learning for

622   Visual Sequences in 8- and 11-Month-Olds. *Infancy*, *14*(1), 2–18.

623   https://doi.org/10.1080/15250000802569611

624 Jovanic, T., Schneider-Mizell, C. M., Shao, M., Masson, J.-B., Denisov, G.,

625   Fetter, R. D., … Zlatic, M. (2016). Competitive Disinhibition Mediates

626   Behavioral Choice and Sequences in Drosophila. *Cell*, *167*(3), 858-

627   870.e19. https://doi.org/10.1016/j.cell.2016.09.009

628 Kabdebon, C., & Dehaene-Lambertz, G. (2019). Symbolic labeling in 5-month-

629   old human infants. *Proceedings of the National Academy of Sciences of*

630   *the United States of America*, *116*(12), 5805–5810.

631   https://doi.org/10.1073/pnas.1809144116

632 Kovács, Á. M., & Mehler, J. (2008). *Regularity learning in 7-month-old infants*

633   *under `noisy' conditions: Adjacent repetitions vs. Non-adjacent*

634   *repetitions*.

635 Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech

636   structures in bilingual infants. *Science*, *325*(5940), 611–612.

637   https://doi.org/10.1126/science.1173947

638 Koyama, M., Minale, F., Shum, J., Nishimura, N., Schaffer, C. B., & Fetcho, J. R.

639   (2016). A circuit motif in the zebrafish hindbrain for a two alternative

640   behavioral choice to turn left or right. *ELife*, *5*.

641   https://doi.org/10.7554/eLife.16808

642 Koyama, M., & Pujala, A. (2018). Mutual inhibition of lateral inhibition: A

643   network motif for an elementary computation in the brain. *Current*

644   *Opinion in Neurobiology*, *49*, 69–74.

645   https://doi.org/10.1016/j.conb.2017.12.019

646 Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms

647   operate in the hippocampus during novelty detection? *Hippocampus*,

648   *17*(9), 735–748. https://doi.org/10.1002/hipo.20326

649 Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2010). Target

650   detection by opponent coding in monkey prefrontal cortex. *Journal of*

651   *Cognitive Neuroscience*, *22*(4), 751–760.

652   https://doi.org/10.1162/jocn.2009.21216

653 Langbein, J., & Puppe, B. (2017). Comment on "Ducklings imprint on the

654   relational concept of 'same or different.'" *Science*, *355*(6327), 806.

655   https://doi.org/10.1126/science.aai7431

656 Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., & Rudy, B. (2013). A

657   disinhibitory circuit mediates motor integration in the somatosensory

658   cortex. *Nature Neuroscience*, *16*(11), 1662–1670.

659   https://doi.org/10.1038/nn.3544

660   Letzkus, J. J., Wolff, S. B. E., & Lüthi, A. (2015). Disinhibition, a Circuit

661   Mechanism for Associative Learning and Memory. *Neuron*, *88*(2), 264–

662   276. https://doi.org/10.1016/j.neuron.2015.09.024

663   Letzkus, J. J., Wolff, S. B. E., Meyer, E. M. M., Tovote, P., Courtin, J., Herry, C.,

664   & Lüthi, A. (2011). A disinhibitory microcircuit for associative fear

665   learning in the auditory cortex. *Nature*, *480*(7377), 331–335.

666   https://doi.org/10.1038/nature10674

667   Liu, T., & Becker, M. W. (2013). Serial Consolidation of Orientation Information

668   Into Visual Short-Term Memory. *Psychol Sci*.

669   https://doi.org/10.1177/0956797612464381

670   Ludueña, G. A., & Gros, C. (2013). A self-organized neural comparator. *Neural*

671   *Computation*, *25*(4), 1006–1028. https://doi.org/10.1162/NECO_a_00424

672   Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual

673   inhibition: A neural model of two-interval discrimination. *Science*,

674   *307*(5712), 1121–1124. https://doi.org/10.1126/science.1104171

675   Manaster-Ramer, A. (1986). Copying in Natural Languages, Context-freeness,

676   and Queue Grammars. *Proceedings of the 24th Annual Meeting on*

677   *Association for Computational Linguistics*, 85–89.

678   https://doi.org/10.3115/981131.981145

679   Mance, I., Becker, M. W., & Liu, T. (2012). Parallel consolidation of simple

680   features into visual short-term memory. *J Exp Psychol Hum Percept*

681   *Perform*, *38*(2), 429–438. https://doi.org/10.1037/a0023925

682   Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning

683 facilitated by speech. *Psychol Sci*, *18*(5), 387–391.

684 https://doi.org/10.1111/j.1467-9280.2007.01910.x

685 Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. (1999). Rule learning by

686 seven-month-old infants. *Science*, *283*(5398), 77–80.

687 Marlin, B. J., Mitre, M., D'amour, J. A., Chao, M. V., & Froemke, R. C. (2015).

688 Oxytocin enables maternal behaviour by balancing cortical inhibition.

689 *Nature*, *520*(7548), 499–504. https://doi.org/10.1038/nature14402

690 Martinho, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept

691 of "same or different." *Science*, *353*(6296), 286–288.

692 https://doi.org/10.1126/science.aaf4247

693 McCarthy, J. J. (1986). OCP Effects: Gemination and Antigemination. *Linguistic*

694 *Inquiry*, *17*(2), 207–263.

695 Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How

696 words can and cannot be learned by observation. *Proc Natl Acad Sci U S*

697 *A*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

698 Miller, P., & Wang, X.-J. (2006). Inhibitory control by an integral feedback signal

699 in prefrontal cortex: A model of discrimination between sequential stimuli.

700 *Proceedings of the National Academy of Sciences of the United States of*

701 *America*, *103*(1), 201–206. https://doi.org/10.1073/pnas.0508072103

702 Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in

703 speech perception: Evidence for early compensation. *Perception &*

704 *Psychophysics*, *65*(6), 956–969.

705 Moravcsik, E. (1978). Reduplicative constructions. In J. H. Greenberg (Ed.),

706 *Universals of human language: Word structure* (Vol. 3, pp. 297–334).

707   Stanford, CA: Stanford University Press.

708 Morgan, J. L. (1986). *From Simple Input to Complex Grammar*. Cambridge, MA:

709   MIT Press.

710 Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule Learning by Rats.

711   *Science*, *319*(5871), 1849–1851. https://doi.org/10.1126/science.1151564

712 Mysore, S. P., & Knudsen, E. I. (2012). Reciprocal inhibition of inhibition: A

713   circuit motif for flexible categorization in stimulus selection. *Neuron*,

714   *73*(1), 193–205. https://doi.org/10.1016/j.neuron.2011.10.037

715 Neiworth, J. J. (2013). Chasing sounds. *Behav Processes*, *93*, 111–115.

716   https://doi.org/10.1016/j.beproc.2012.11.009

717 Owen, S. F., Tuncdemir, S. N., Bader, P. L., Tirko, N. N., Fishell, G., & Tsien, R.

718   W. (2013). Oxytocin enhances hippocampal spike transmission by

719   modulating fast-spiking interneurons. *Nature*, *500*(7463), 458–462.

720   https://doi.org/10.1038/nature12330

721 Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African

722   Grey parrot (Psittacus erithacus): Learning with respect to categories of

723   color, shape, and material. *Anim Learn Behav*, *15*, 421–32.

724 Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. (2013). Inhibition

725   of inhibition in visual cortex: The logic of connections between

726   molecularly distinct interneurons. *Nature Neuroscience*, *16*(8), 1068–

727   1076. https://doi.org/10.1038/nn.3446

728 Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., & Kepecs, A.

729   (2013). Cortical interneurons that specialize in disinhibitory control.

730   *Nature*, *503*(7477), 521–524. https://doi.org/10.1038/nature12676

731 Pinker, S. (1984). *Language Learnability and Language Development*.

732   Cambridge, MA: MIT Press.

733 Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special

734   about it? *Cognition*, *95*(2), 201–236.

735   https://doi.org/10.1016/j.cognition.2004.08.004

736 Rabagliati, H., Ferguson, B., & Lew-Williams, C. (2019). The profile of abstract

737   rule learning in infancy: Meta-analytic and experimental evidence.

738   *Developmental Science*, *22*(1), e12704. https://doi.org/10.1111/desc.12704

739 Rose, S., & Walker, R. (2011). Harmony Systems. In J. Goldsmith, J. Riggle, &

740   A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., pp.

741   240–290). https://doi.org/10.1002/9781444343069.ch8

742 Rubino, C. (2013). Reduplication. In M. S. Dryer & M. Haspelmath (Eds.), *The*

743   *World Atlas of Language Structures Online*. Retrieved from

744   http://wals.info/chapter/27

745 Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is

746   a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3),

747   669–680. https://doi.org/10.1016/j.cognition.2006.11.004

748 Schwartz, R. G., Leonard, L. B., Wilcox, M. J., & Folger, M. K. (1980). Again

749   and again: Reduplication in child phonology. *Journal of Child Language*,

750   *7*(1), 75–87. https://doi.org/10.1017/S0305000900007030

751 Semyanov, A., Walker, M. C., Kullmann, D. M., & Silver, R. A. (2004).

752   Tonically active GABA A receptors: Modulating gain and maintaining the

753   tone. *Trends in Neurosciences*, *27*(5), 262–269.

754   https://doi.org/10.1016/j.tins.2004.03.005

755 Senghas, A., Kita, S., & Ozyürek, A. (2004). Children creating core properties of

756   language: Evidence from an emerging sign language in Nicaragua.

757   *Science*, *305*(5691), 1779–1782. https://doi.org/10.1126/science.1100199

758 Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows

759   spontaneously exhibit analogical reasoning. *Current Biology*, *25*(2), 256–

760   260. https://doi.org/10.1016/j.cub.2014.11.063

761 Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel Harmony and Speech

762   Segmentation in Finnish. *J Mem Lang*, *36*(3), 422–444.

763 van Der Velde, F., & de Kamps, M. (2001). From knowing what to knowing

764   where: Modeling object-based attention with feedback disinhibition of

765   activation. *Journal of Cognitive Neuroscience*, *13*(4), 479–491.

766 Versace, E., Spierings, M. J., Caffini, M., Ten Cate, C., & Vallortigara, G. (2017).

767   Spontaneous generalization of abstract multimodal patterns in young

768   domestic chicks. *Animal Cognition*, *20*(3), 521–529.

769   https://doi.org/10.1007/s10071-017-1079-5

770 Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of

771   consolidation in visual working memory. *J Exp Psychol Hum Percept*

772   *Perform*, *32*(6), 1436–1451. https://doi.org/10.1037/0096-1523.32.6.1436

773 Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The Roles of Word Stress

774   and Vowel Harmony in Speech Segmentation. *J Mem Lang*, *38*(2), 133–

775   149.

776 Wen, S., Ulloa, A., Husain, F., Horwitz, B., & Contreras-Vidal, J. L. (2008).

777   Simulated neural dynamics of decision-making in an auditory delayed

778   match-to-sample task. *Biological Cybernetics*, *99*(1), 15–27.

779   https://doi.org/10.1007/s00422-008-0234-0

780 Xu, H., Jeong, H.-Y., Tremblay, R., & Rudy, B. (2013). Neocortical somatostatin-

781   expressing GABAergic interneurons disinhibit the thalamorecipient layer

782   4. *Neuron*, *77*(1), 155–167. https://doi.org/10.1016/j.neuron.2012.11.004

783 Yang, C. (2013). Ontogeny and phylogeny of language. *Proc Natl Acad Sci U S*

784   *A*, *110*(16), 6324–6327. https://doi.org/10.1073/pnas.1216803110

785 Yip, M. (1982). Reduplication and C-V Skeleta in Chinese Secret Languages.

786   *Linguistic Inquiry*, *13*(4), 637–661.

787 Yip, M. (1988). The Obligatory Contour Principle and Phonological Rules: A

788   Loss of Identity. *Linguistic Inquiry*, *19*(1), 65–100.

789 Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., …

790   Dan, Y. (2014). Selective attention. Long-range and local circuits for top-

791   down modulation of visual cortex processing. *Science*, *345*(6197), 660–

792   665. https://doi.org/10.1126/science.1254126

793 Zhao, W., Zhou, P., Gong, C., Ouyang, Z., Wang, J., Zheng, N., & Gong, Z.

794   (2019). A disinhibitory mechanism biases Drosophila innate light

795   preference. *Nature Communications*, *10*(1), 124.

796   https://doi.org/10.1038/s41467-018-07929-w

797

798

Appendix A: Model equations

800 <u>A.1 Sequential model</u>

801 The feature f is encoded in the source layer, the inhibition layer and the

802 copy layer; the corresponding activations, are $S_f(t)$ for a unit encoding feature f in

803 the source layer, $I_f(t)$ for such a unit in the inhibition layer and (3) $C_f(t)$ such for a

804 unit in the copy layer. $E_f(t)$ is the external input, $\mathcal{N}(\mu,\sigma)$ is a random value drawn

805 from a normal distribution with mean μ and standard deviation σ.

806 Before stimulation, the activation in the source layer and in the copy layer

807 are initialized to zero (plus noise), while the activation in the inhibition layer is

808 initialized to some value a_I (here arbitrarily set to 1):

809

$$
\begin{aligned}
S_f(t=0) &\sim \mathcal{N}(0,\sigma_{activation}) \\
(1) \quad C_f(t=0) &\sim \mathcal{N}(0,\sigma_{activation}) \\
I_f(t=0) &\sim \mathcal{N}(a_I,\sigma_{activation})
\end{aligned}
$$

810

811

812 The connection weights between units in the different layers are indicated

813 by w: $w^{I,S}$ from the source layer to the inhibition layer, $w^{C,S}$ from the source layer

814 to the copy layer and $w^{C,I}$ from the inhibition layer to the copy layer. A connection

815 between a source layer unit coding for feature f and a copy layer unit coding for

816 feature f' is indicated by $w^{C,S}_{f'f}$. The weights are given as follows:

817

818   (2)
$$w_{f',f}^{C,S} \sim \begin{cases} \mathcal{N}(1,\sigma_{weight}) & f = f' \\ 0 & f \neq f' \end{cases}$$

$$w_{f',f}^{C,I} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f = f' \\ 0 & f \neq f' \end{cases}$$

$$w_{f',f}^{I,S} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f = f' \\ 0 & f \neq f' \end{cases}$$

819

820   At each time step, the activations in the different layers are then updated

821   as follows; as mentioned in the main text, the update order is critical.

822

823   (3)
$$S_f(t) = E_f(t) + \mathcal{N}(0,\sigma_{activation})$$
$$C_f(t) = w_f^{C,S} S_f(t) + w_f^{C,I} I_f(t) + \mathcal{N}(0,\sigma_{activation})$$
$$I_f(t) = \mathcal{N}(a_I,\sigma_{activation}) + w_f^{I,S} S_f(t)$$

824

825   At the end of each update cycle, the activations are curtailed to be between

826   zero and one.

827   <u>A.2. Simultaneous model</u>

828   In the simultaneous model, units represent both features and spatial

829   locations. $S_{f,l}(t)$ is thus the activation of a unit in the source layer that encodes

830   feature f at location l, $I_{f,l}(t)$ is the corresponding activation in the inhibition layer

831   and (3) $C_{f,l}(t)$ is the corresponding activation in the copy layer. $E_{f,l}(t)$ is the

832   external input.

833   Before stimulation, the activation in the source layer and in the copy layer

834   are initialized to zero (plus noise), while the activation in the inhibition layer is

835   initialized to some value a_I (here arbitrarily set to 1):

836

$$S_{f,l}(t=0) \sim \mathcal{N}(0,\sigma_{activation})$$
837   (4) $$C_{f,l}(t=0) \sim \mathcal{N}(0,\sigma_{activation})$$
$$I_{f,l}(t=0) \sim \mathcal{N}(a_I,\sigma_{activation})$$

838

839   Connection weights now carry indices for both features and spatial

840   locations. For example, a connection between a source layer unit coding for

841   feature f at location l and a copy layer unit coding for feature f' at location l' is

842   indicated by $w^{C,S}_{f',f,l',l}$. The weights are given as follows:

843

$$w^{C,S}_{f',f,l',l} \sim \begin{cases} \mathcal{N}(1,\sigma_{weight}) & f=f', l=l' \\ 0 & \text{otherwise} \end{cases}$$

844   (5) $$w^{C,I}_{f',f,l',l} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f=f' \\ 0 & f \neq f' \end{cases}$$

$$w^{I,S}_{f',f,l',l} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f=f' \\ \mathcal{N}(1,\sigma_{weight}) & f \neq f' \end{cases}$$

845

846   At each time step, the activations in the different layers are then updated

847   as follows; as mentioned in the main text, the update order is critical.

848

$$S_{f,l}(t) = E_{f,l}(t)+N(0,\sigma_{activation})$$
849   (6) $$I_{f,l}(t) = N(a_I,\sigma_{activation})+ \sum_{f_S,l_S} w^{I,S}_{f,l,f_S,l_S} S_{f_S,l_S}(t)$$

$$C_{f,l}(t) = \sum_{f_S,l_S} w^{C,S}_{f,l,f_S,l_S} S_{f_S,l_S}(t)+ \sum_{f_I,l_I} w^{C,I}_{f,l,f_I,l_I} I_{f_I,l_I}(t)+N(0,\sigma_{activation})$$

850

851   At the end of each update cycle, the activations are curtailed to be between

852   zero and one.

A.3. Combined model

854   The combined sequential/simultaneous model is similar to the

855   simultaneous model in that it comprises a source layer, a copy layer and an

856   inhibition layer and that the copy layer receives excitatory input from the source

857   layer as well as tonic inhibition from those units in the inhibition layer that code

858   for the same feature and spatial position. Further, each feature of the input layer

859   inhibits the corresponding feature in the inhibition layer across spatial positions

860   and excites all other features. The critical difference between the simultaneous

861   and the combined model is that there are no connections between the source layer

862   and the inhibition layer that code for the same feature *at the same location* (while

863   disinhibition occurs for other locations), and that same-location disinhibition of

864   features proceeds through a *self-disinhibition layer* where each unit (1) receives

865   excitatory input from the corresponding feature and location in the source layer

866   (2) sends inhibitory input to all units coding for the same feature (across

867   locations) in the inhibition layer.

868   The symbols for the activation in the source, inhibition and copy layers are

869   the same as in the simultaneous model; activation in the self-disinhibition layer

870   for a unit coding for feature f at location l is designated as $D_{f,l}(t)$ and is initialized

871   using random values around zero.

872   The symbols for the connection weights are similar to those in the

873   simultaneous model, but the weights reflect the changes above:

874

$$
w_{f',f,l',l}^{C,S} \sim \begin{cases} \mathcal{N}(1,\sigma_{weight}) & f = f', l = l' \\ 0 & \text{otherwise} \end{cases}
$$

$$
w_{f',f,l',l}^{D,S} \sim \begin{cases} \mathcal{N}(1,\sigma_{weight}) & f = f', l = l' \\ 0 & \text{otherwise} \end{cases}
$$

875   (7)

$$
w_{f',f,l',l}^{C,I} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f = f' \\ 0 & f \neq f' \end{cases}
$$

$$
w_{f',f,l',l}^{I,S} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f = f', l \neq l' \\ 0 & f = f', l = l' \\ \mathcal{N}(1,\sigma_{weight}) & f \neq f' \end{cases}
$$

$$
w_{f',f,l',l}^{I,D} \sim \begin{cases} \mathcal{N}(-1,\sigma_{weight}) & f = f' \\ 0 & f \neq f' \end{cases}
$$

876

877   At each time step, the activations in the different layers are then updated

878   as follows; again, the update order is critical.

879

$$
S_{f,l}(t) = E_{f,l}(t) + N(0,\sigma_{activation})
$$

$$
I_{f,l}(t) = N(a_I,\sigma_{activation}) + \sum_{f_S,l_S} w_{f,l,f_S,l_S}^{I,S} S_{f_S,l_S}(t) + \sum_{f_D,l_D} w_{f,l,f_D,l_D}^{I,D} D_{f_D,l_D}(t)
$$

880   (8)

$$
C_{f,l}(t) = \sum_{f_S,l_S} w_{f,l,f_S,l_S}^{C,S} S_{f_S,l_S}(t) + \sum_{f_I,l_I} w_{f,l,f_I,l_I}^{C,I} I_{f_I,l_I}(t) + N(0,\sigma_{activation})
$$

$$
D_{f,l}(t) = \sum_{f_S,l_S} w_{f,l,f_S,l_S}^{D,S} S_{f_S,l_S}(t) + N(0,\sigma_{activation})
$$

881

882   At the end of each update cycle, the activations are curtailed to be between

883   zero and one.