

Major evolutionary transitions as Bayesian structure learning

Dániel Czégel^{1,2}, István Zachar^{1,3,4}, and Eörs Szathmáry^{1,2,4}

¹MTA Centre for Ecological Research, Evolutionary Systems Research Group, Hungarian Academy of Sciences, H-8237 Tihany, Hungary

²Dept. of Plant Systematics, Ecology and Theoretical Biology, Eötvös University, H-1117 Budapest, Hungary

³MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, Eötvös University, H-1117, Budapest, Hungary

⁴Parmenides Foundation, Center for the Conceptual Foundations of Science, 82049 Pullach/Munich, Germany

June 29, 2018

Abstract

Complexity of life forms on Earth has increased tremendously, primarily driven by subsequent evolutionary transitions in individuality, a mechanism in which units formerly being capable of independent replication combine to form higher-level evolutionary units. Although this process has been likened to the recursive combination of pre-adapted sub-solutions in the framework of learning theory, no general mathematical formalization of this analogy has been provided yet. Here we show, building on former results connecting replicator dynamics and Bayesian update, that (i) evolution of a hierarchical population under multilevel selection is equivalent to Bayesian inference in hierarchical Bayesian models, and (ii) evolutionary transitions in individuality, driven by synergistic fitness interactions, is equivalent to learning the structure of hierarchical models via Bayesian model comparison. These correspondences support a learning theory oriented narrative of evolutionary complexification: the complexity and depth of the hierarchical structure of individuality mirrors the amount and complexity of data that has been integrated about the environment through the course of evolutionary history.

28 Keywords: Evolution, multilevel selection, major evolutionary transitions, Bayesian
29 models, structure learning, graphical models

30 1 Introduction

31 On Earth, life has undergone immense complexification [1, 2]. The evolutionary path
32 from the first self-replicating molecules to structured societies of multicellular organisms
33 has been paved with exceptional milestones: units that were capable of independent
34 replication have combined to form a higher-level unit of replication [3, 4, 5]. Such *evo-*
35 *lutionary transitions in individuality* opened the door to the vast increase of complexity
36 via hierarchical aggregation of pre-adapted subunits. Paradigmatic examples include the
37 transition of replicating molecules to protocells, the endosymbiosis of mitochondria and
38 plastids by eucaryotic cells and the appearance of multicellular organisms and eusociality.
39 Interestingly, it is possible to identify common evolutionary mechanisms that possibly led
40 to these unique but analogous events [6, 7, 8, 9]. A crucial preliminary condition is the
41 *alignment of interests*: in order to undergo an evolutionary transition in individuality,
42 organisms must exhibit extreme form of cooperation, originating from genetic relatedness
43 and/or synergistic fitness interactions [4]. However, the story does not end here: some-
44 thing must maintain the alignment of interests subsequent to the transition, too. At that
45 point, the fate of the organism depends on selective forces at multiple levels that might be
46 in conflict with each other. Incorporating the effects of multilevel selection is, therefore,
47 a crucial element of understanding evolutionary transitions in individuality [10].

48 These theoretical considerations above delineate conditions under which a transition
49 might occur and a possibly different set of conditions which help to maintain the integrity
50 of units that have already undergone transition. However, these considerations alone can-
51 not offer a predictive theory of complexification as they do not address the question of how
52 necessary these environmental and ecological conditions are. An alternative, supplement-
53 ary approach that circumvents these difficulties is to investigate whether mathematical
54 theories of adaptation and learning can provide further insights about the general scheme
55 of evolutionary transitions in individuality. In this paper, we argue that they do. We
56 first provide a mapping between multilevel selection modeled by discrete-time replica-
57 tor dynamics and Bayesian inference in belief networks (i.e., directed graphical models),
58 which shows that the underlying mathematical structures are isomorphic. The two key
59 ingredients are (i) the already known equivalence between univariate Bayesian update
60 and single-level replicator dynamics [11, 12] and (ii) a possible correspondence between
61 properties of a hierarchical population composition and multivariate probability theory.
62 We then show that this isomorphism allows for a natural interpretation of evolutionary
63 transitions in individuality as *learning the structure* [13, 14] of the belief network. Indeed,
64 following adaptive paths on the fitness landscape over possible hierarchical population
65 compositions is equivalent to a well-known method used for selecting the optimal model
66 structure in the Bayesian paradigm, namely, *Bayesian model comparison*. This suggests
67 that complexification of life via successive evolutionary transitions in individuality is anal-
68 ogous to the complexification of optimal model structure as more (or more complex) data
69 about the environment is available.

70 Relating the dynamics of evolutionary complexification to hierarchical probabilis-

71 tic generative models complements recent efforts of searching for algorithmic analogies
72 between emergent evolutionary phenomena and neural network based learning models
73 [15, 16]. These include correspondences between evolutionary-ecological dynamics and
74 autoassociative networks [17] and also linking the evolution of developmental organiza-
75 tion to learning in artificial neural networks [18]. As such connectionist models account
76 for how global self-organizing learning behavior might emerge from simple local rules
77 (e.g., weight updates), our approach aims at providing a common global framework for
78 modeling both evolutionary and learning dynamics.

79 In the following, we provide a brief introduction to the elementary building blocks
80 of our arguments: Bayesian update and replicator dynamics. Bayesian update [19] fits
81 a probability distribution $P(I)$ of hypotheses $I = I_1, \dots, I_m$ to the data \mathbf{e} . It does so
82 by integrating prior knowledge about the probability $P(I_i)$ of hypothesis I_i with the
83 likelihood that the actual data $\mathbf{e} = \mathbf{e}(t)$ is being generated by hypothesis I_i , given by
84 $P(\mathbf{e}(t)|I_i)$. Mathematically, the fitted distribution $P(I_i|\mathbf{e}(t))$, called the *posterior*, is
85 simply proportional to both the *prior* $P(I_i)$ and the likelihood $P(\mathbf{e}(t)|I_i)$:

$$P(I_i|\mathbf{e}(t)) = \frac{P(\mathbf{e}(t)|I_i)P(I_i)}{\sum_i P(\mathbf{e}(t)|I_i)P(I_i)} \quad (1)$$

86 On the other hand, the discrete replicator equation [20] that accounts for the change
87 in relative abundance $f(I_i)$ of types of replicating individuals I_i in the population driven
88 by their fitness values $w(I_i)$, reads as

$$f(I_i; t + 1) = \frac{w(I_i; t)f(I_i; t)}{\sum_i w(I_i; t)f(I_i; t)}. \quad (2)$$

89 As first noted by Harper [11] and Shalizi [12], equations (1) and (2) are equivalent,
90 with the following identified quantities. The relative abundance $f(I_i; t)$ of type I_i at
91 time t corresponds to the prior probability $P(I_i)$; the relative abundance $f(I_i; t + 1)$ at
92 time $t + 1$ is corresponding to the posterior probability $P(I_i|\mathbf{e}(t))$; the fitness $w(I_i; t)$
93 of type I_i at time t is corresponding to the likelihood $P(\mathbf{e}(t)|I_i)$; and the *average fitness*
94 $\sum_i w(I_i; t)f(I_i; t)$ is corresponding to the normalizing factor $\sum_i P(\mathbf{e}(t)|I_i)P(I_i)$ called the
95 model *evidence*.

96 Building on this observation, a natural question to ask is if this mathematical equiv-
97 alence is only an apparent similarity due to the simplicity of both models, or it is a
98 consequence of a deeper structural analogy between evolutionary and learning dynamics.
99 We propose two conceptually new avenues along which this equivalence can be gener-
100 alized. First, we identify concepts of hierarchical evolutionary processes with concepts
101 of (i) multivariate probability theory, (ii) Bayesian inference in hierarchical models and
102 (iii) conditional independence relations between variables in such models. Building on
103 this theoretical bridge, we then investigate the dynamics of learning the structure (as
104 opposed to parameter fitting in a fixed model) of hierarchical Bayesian models and the
105 Darwinian evolution of multilevel populations, concluding that following adaptive evolu-
106 tionary paths on the landscape of hierarchical populations naturally maps to optimizing
107 the structure of hierarchical Bayesian models via Bayesian model comparison.

108 2 Results

109 In order to generalize the algebraic equivalence between discrete-time replicator dynamics
110 (2) and Bayesian update (1) to multilevel selection scenarios, multivariate distributions
111 have to be involved. In general, a multivariate distribution $P(x_1, \dots, x_k)$ over k variables,
112 each taking m possible values, can be encoded by $m^k - 1$ independent parameters, which
113 is exponential in the number of variables. Apart from practical considerations such as
114 the possible infeasibility of computing marginal and conditional distributions, sampling
115 and storing such general distributions, a crucial theoretical limitation is that fitting data
116 by a model with such a sizable parameter space would result in overfitting, unless the
117 training dataset is itself comparably large [21].

118 A way to overcome such obstacles is to explicitly abandon indirect dependencies be-
119 tween variables by using structured probabilistic models, such as belief networks (called
120 also Bayesian networks or directed graphical models) [22, 23]. Indeed, belief networks
121 simplify joint distribution over multiple variables by specifying *conditional independence*
122 *relations* corresponding to indirect (as opposed to direct) dependencies between variables.

123 In the following, we build up an algebraic isomorphism between discrete-time multi-
124 level replicator dynamics and iterated Bayesian inference in belief networks on a step-by-
125 step basis. The key identified quantities are summarized in Table 1.

126 **Composition: mapping properties of multilevel populations to multivariate**
127 **probability theory.** A multilevel population is regarded as a hierarchical containment
128 structure of types: Individual types I_i might be part of collectives C_j^1 which themselves
129 might be part of higher-level collectives C_k^2 , and so on, as illustrated in Figure 1. Note
130 that collectives at any level might possess heritable information (henceforth referred to as
131 their identity); collectives of the same (hierarchical) composition might very well have dif-
132 ferent identities. This makes this framework flexible enough to incorporate qualitatively
133 different stages of evolutionary interdependence between organisms, leading eventually to
134 a transition in individuality: (i) selection in which individuals enjoy the synergistic effect
135 of belonging to a collective, but the collectives themselves do not possess any heritable
136 information; (ii) selection in which collectives possess their own heritable information but
137 also the individuals in them might replicate at different rates; (iii) and selection in which
138 individuals have already lost their ability to replicate independently, therefore, their fit-
139 ness is totally determined by the collective they belong to. As Michod and Nedelcu
140 write on p. 61 of Ref. [24], "group fitness is, initially, taken to be the average of the
141 lower-level individual fitnesses; but as the evolutionary transition proceeds, group fitness
142 becomes decoupled from the fitness of its lower-level components". This, as we shall see,
143 is exactly what our model accounts for mathematically, incorporating also the effect of
144 stochastically varying environment.

145 A key assumption that enables the machinery of multivariate probability theory
146 to work is that abundance of collectives is measured in terms of abundance of indi-
147 viduals they contain. Indeed, by identifying the abundance of individuals of type I_i ,
148 $f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots)$, that are part of collectives of type C_j^1 that are themselves part
149 of collectives of type C_k^2 , etc., with the joint probabilities $P(I_i, C_j^1, C_k^2, \dots)$, two important
150 additional identification follows:

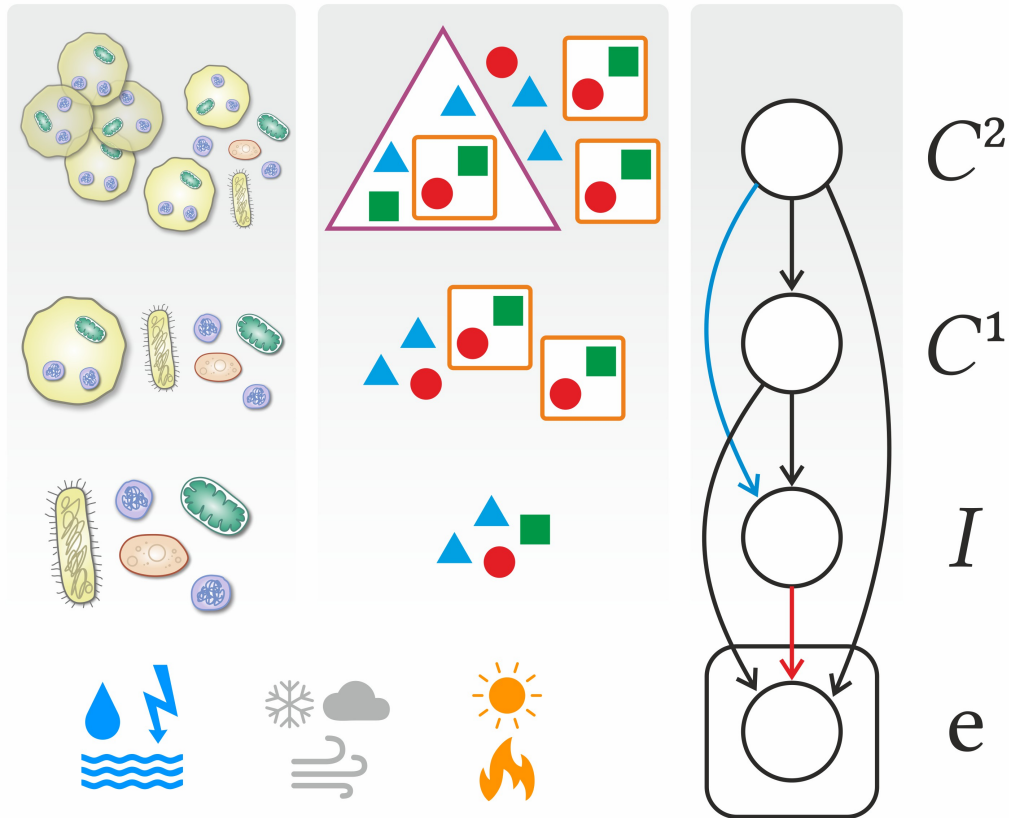


Figure 1: Evolution of multilevel population as inference in Bayesian belief network. The stochastic environment \mathbf{e} governs the evolutionary dynamics of multilevel population composition $f(I_i \text{ in } C_j^1 \text{ in } C_k^2)$. This is, in turn, equivalent to successive Bayesian inference of hidden variables I , C^1 and C^2 based on the observation of current the environmental parameters \mathbf{e} . Since these environmental parameters are sampled and observed multiple times (i.e., at every timestep $t = 1, 2, 3 \dots$), the corresponding node of the belief network is conventionally placed on a plate. Also note that the deletion of links between nodes of the belief network is corresponding to conditional independence relations between variables in the Bayesian setting and to specific structural properties of selection and population composition in the evolutionary setting; see text for details.

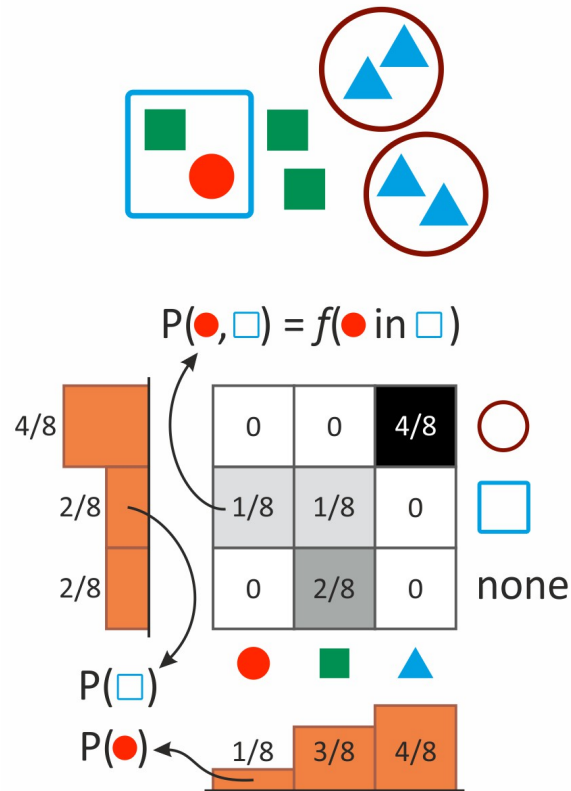


Figure 2: Two-level population encoded as a bivariate probability distribution. Joint probabilities represent the relative abundance of different individuals in different collectives. Conditional distributions depict the composition of collectives (rows) or the membership distribution of individuals (columns). Marginals, illustrated by the one-dimensional histograms, represent the abundance distribution of types at the individual level (horizontal) or at the level of collectives (vertical histogram).

multivariate probability theory	multilevel population
joint probabilities $P(I_i, C_j^1, C_k^2, \dots)$	relative abundances of individuals $f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots)$
marginals, e.g., $P(C_j^1) = \sum_{i,k,\dots} P(I_i, C_j^1, C_k^2, \dots)$	relative abundances of units at a given level, e.g., of collectives at level C^1 , $f(C_j^1) = \sum_{i,k,\dots} f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots) = f(\text{any } I \text{ in } C_j^1 \text{ in any } C^2 \text{ in } \dots)$
conditional probabilities, e.g., $P(I_i C_j^1) = P(I_i, C_j^1)/P(C_j^1)$ OR $P(C_j^1 I_i) = P(I_i, C_j^1)/P(I_i)$	composition of collectives $f(I_i \text{ in } C_j^1)/f(\text{any } I \text{ in } C_j^1)$ OR membership distribution of individuals $f(I_i \text{ in } C_j^1)/f(I_i \text{ in any } C^1)$
Bayesian inference in hierarchical models	multilevel replicator dynamics
prior, $P(I_i, C_j^1, C_k^2, \dots; t)$	relative abundance $f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$
likelihood, $P(\mathbf{e}(t) I_i, C_j^1, C_k^2, \dots; t)$	fitness $w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$
posterior, $P(I_i, C_j^1, C_k^2, \dots; t+1)$	relative abundance $f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t+1)$
model evidence, $\sum_{i,j,k,\dots} P(\mathbf{e}(t) I_i, C_j^1, C_k^2, \dots; t) \times P(I_i, C_j^1, C_k^2, \dots; t)$	average fitness $\sum_{i,j,k,\dots} w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t) \times f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$
conditional independence relations	properties of multilevel selection
conditional independence of the observed variable \mathbf{e} and a latent variable, e.g., I , $P(\mathbf{e} I, C^1, C^2, \dots) = P(\mathbf{e} C^1, C^2, \dots)$	units at a given level, e.g., individuals, "freeze": their fitness is completely determined by the collective(s) they belong to: $w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots)$ is the same for all i
conditional independence between two latent variables, e.g., I and C^2 , $P(I C^1, C^2, \dots) = P(I C^1, \dots)$	the composition of units at level C^1 is independent of what units they belong to at level C^2 .
Bayesian structure learning	evolutionary transitions in individuality
evidence of model \mathcal{M}_a , $E(\mathcal{M}_a) = P(\mathbf{e} \mathcal{M}_a) = \sum_{i,j,k,\dots} P(\mathbf{e} I_i, C_j^1, C_k^2, \dots, \mathcal{M}_a) \times P(I_i, C_j^1, C_k^2, \dots \mathcal{M}_a)$	average fitness given population composition \mathcal{M}_a , $\bar{w}(\mathcal{M}_a) = \sum_{i,j,k,\dots} w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots) \times f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots)$
difference of evidence, $E(\mathcal{M}_b) - E(\mathcal{M}_a)$	difference of average fitness of those units that are participating in the transition in individuality, causing the $\mathcal{M}_a \rightarrow \mathcal{M}_b$ change in population structure

Table 1: Identified quantities of evolution and learning

- 151 • *marginal distributions*, such as $P(C_j^1) = \sum_{i,k,\dots} P(I_i, C_j^1, C_k^2, \dots)$ translate to the
152 *abundance distribution of types at the corresponding level* (here, C^1), $f(C_j^1) =$
153 $\sum_{i,k,\dots} f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots) = f(\text{any } I \text{ in } C_j^1 \text{ in any } C^2 \text{ in } \dots)$
- 154 • *conditional distributions*, e.g., $P(I_i|C_j^1) = P(I_i, C_j^1)/P(C_j^1)$ or $P(C_j^1|I_i) = P(I_i, C_j^1)/P(I_i)$
155 translate either to *composition of collectives* $f(I_i \text{ in } C_j^1)/f(\text{any } I \text{ in } C_j^1)$ or *member-*
156 *ship distribution of individuals* (or lower level collectives), $f(I_i \text{ in } C_j^1)/f(I_i \text{ in any } C^1)$.

157 These computations are illustrated by a toy example in Figure 2.

158 **Dynamics: multilevel replicator dynamics as inference in Bayesian belief**
159 **networks.** Just like in the single-level case, the environmental parameters $\mathbf{e}(t)$, $t =$
160 $1, 2, 3, \dots$ are assumed to be sampled from an unknown generative process; the succes-
161 sive observation of them drives the successive update of population composition. As
162 discussed earlier, however, multilevel population structures can be mapped to multivariate
163 probability distributions, forming multiple *latent* variables to be updated upon the
164 observation of \mathbf{e} .

165 Formally, just as prior probabilities over multiple hypotheses $P(I_i, C_j^1, C_k^2, \dots; t)$ are
166 updated to posterior probabilities $P(I_i, C_j^1, C_k^2, \dots; t+1)$ based on the likelihood,
167 $P(\mathbf{e}(t)|I_i, C_j^1, C_k^2, \dots; t)$, in the same way, multilevel population composition at time t ,
168 $f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$ is updated to the composition at $t+1$ based on fitnesses
169 $w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$. The critical conceptual identification here is therefore of (i)
170 the likelihood of the hypothesis parametrized by $(I_i, C_j^1, C_k^2, \dots)$ and of (ii) the fitness
171 of those individuals I_i that belong to those collectives C_j^1 that belong to C_k^2 , etc. The
172 normalization factor that ensures that (i) the multivariate distribution is normalized (the
173 model evidence $\sum_{i,j,k,\dots} P(\mathbf{e}(t)|I_i, C_j^1, C_k^2, \dots; t) \times P(I_i, C_j^1, C_k^2, \dots; t)$) or that (ii) abun-
174 dances are always measured relative to the total abundance of individuals (the average
175 fitness $\sum_{i,j,k,\dots} w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t) \times f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots; t)$), is conceptually
176 irrelevant here as they do not change the ratio of probabilities or abundances. Their
177 equivalence will, however, play a critical role in relating evolution of individuality and
178 structure learning of belief networks.

179 In order to demonstrate how simple calculations are performed in this framework and
180 also to elucidate how fitnesses are determined, here we calculate the fitness of collective
181 C_j^1 , $w(C_j^1)$, which has been identified with $P(\mathbf{e}|C_j^1)$. Using simple relations of probability
182 theory, $P(\mathbf{e}|C_j^1) = \sum_{I_i} P(\mathbf{e}, I_i|C_j^1) = \sum_{I_i} P(\mathbf{e}|I_i, C_j^1)P(I_i|C_j^1)$. Translating this back to
183 the language of evolution tells us that the fitness of C_j^1 is simply the average fitness of
184 individuals it contains, as anticipated earlier.

185 **Structure: mapping structural properties of multilevel selection to the struc-**
186 **ture of Bayesian belief network.** Structured probabilistic models are useful be-
187 cause they concisely summarize direct and indirect dependencies between multiple vari-
188 ables. Specifically, Bayesian belief networks depict multivariate distributions, such as
189 $P(\mathbf{e}, I, C^1, C^2)$, as a directed network, with the variables corresponding to the nodes and
190 conditioning one variable on another corresponds to a directed link between the two.
191 Since $P(\mathbf{e}, I, C^1, C^2)$ can *always* be written as $P(\mathbf{e}|I, C^1, C^2)P(I|C^1, C^2)P(C^1|C^2)P(C^2)$

192 in terms of conditional probabilities, the corresponding belief network is the one illus-
193 trated in Figure 1. The route to simplify the structure of the distribution and corre-
194 spondingly, the structure (i.e., connectivity) of the belief network is through *conditional*
195 *independence relations*. Conditional independence relations, such as

$$P(\mathbf{e}|I, C^1, C^2) = P(\mathbf{e}|C^1, C^2) \quad (3)$$

196 correspond to the deletion of connections; (3), for example, corresponds to the deletion
197 of the connection between variables \mathbf{e} and I , shown in red in Figure 1, and it describes the
198 conditional independence of the observed variable \mathbf{e} and a latent variable, I . What does
199 this independence relation mean in evolutionary terms? As it logically follows from the
200 previous identifications, it specifies that the units at level I are *frozen* in an evolutionary
201 sense: their fitness is completely determined by the collective they belong to. There is
202 a second, qualitatively different type of conditional independence relations: those be-
203 tween two latent variables, corresponding to two levels of the population. For example,
204 $P(I|C^1, C^2) = P(I|C^1)$, corresponding to the deletion of the blue link in Figure 1, is
205 interpreted as the following: the composition of any collective at level C^1 is independent
206 of what higher-level collective (at level C^2) it belongs to. Such simplifications in hierar-
207 chical population composition allows for the step-by-step modular combination of units
208 to higher-level units, re-using existing sub-solutions over and over again.

209 **Structural dynamics: evolutionary transitions in individuality as Bayesian**
210 **structure learning.** It has been shown above that Bayesian inference in belief net-
211 works can be interpreted as Darwinian evolutionary dynamics of multilevel populations,
212 driven by the "observation" of the actual environment $\mathbf{e}(t)$. What fits the environment
213 is the hierarchical distribution of individuals (i.e., lowest level replicators) to collectives.
214 However, the number of levels and the existing types within each level, along with the
215 assumptions of hierarchical containment dependencies (i.e., conditional independence re-
216 lations) has to be *a priori* specified. In this sense, fitting the environment by such a
217 pre-defined structure via successive Bayesian updates has limited adaptation abilities. In
218 particular, it is unable to adjust the complexity of the model to be in accordance with
219 that of the environment, an inevitable property to avoid under- or overfitting.

220 In order to enlarge the space of possible models and therefore fit the environment
221 better, one might allow the model structure to adapt as well. More complex models,
222 however, will *always* fit any data better, and accordingly, adapting the model structure
223 naively might result in overfitting, i.e., the inability of the model to account for never-seen
224 data, corresponding to possible future environments. Organisms with too complicated
225 hierarchical containment structures (and other adaptive parameters that are not modeled
226 explicitly here) would go extinct in any varying environment. In order to remedy this
227 situation, one has to take into consideration not only how good the best parameter
228 combination fits the data, but also how hard it is to find such a parameter-combination.
229 A systematic way of doing so is known as *Bayesian model comparison*, a well-known
230 method in machine learning and Bayesian modeling. Mathematically, Bayesian model
231 comparison simply ranks models (here, belief networks) according to their average ability
232 to fit the data, referred to as the *evidence* $E(\mathcal{M})$ of model \mathcal{M} :

$$E(\mathcal{M}) = P(\mathbf{e}|\mathcal{M}) = \sum_{i,j,k,\dots} P(\mathbf{e}|I_i, C_j^1, C_k^2, \dots, \mathcal{M}) \times P(I_i, C_j^1, C_k^2, \dots | \mathcal{M}) \quad (4)$$

233 The first term in the sum describes the likelihood of the current parameters (i.e., their
234 ability to fit the data), whereas the second term weights these likelihoods according to
235 the prior probabilities of the parameters.

236 How evolution, on the other hand, limits the number of to-be-fitted parameters in any
237 organism to reinforce evolvability is an intriguing phenomenon. Here we show that in our
238 minimal framework, selection naturally accounts for model complexity: model evidence
239 corresponds to the average fitness \bar{w} of individuals, determined by their hierarchical
240 grouping to higher-level replicators. Indeed, interpreting 4 in evolutionary terms gives

$$\sum_{i,j,k,\dots} w(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots) \times f(I_i \text{ in } C_j^1 \text{ in } C_k^2 \text{ in } \dots) = \bar{w}(\mathcal{M}) \quad (5)$$

241 in which the first term in the sum corresponds to fitnesses of individuals according to
242 what collectives they belong to, and the second terms weights these fitnesses according
243 to the abundance of such hierarchical arrangements. It implies that not only the evolu-
244 tion of the *composition* of multilevel population, but also the evolution of the *structure*
245 of the multilevel population can be interpreted both in Darwinian and Bayesian terms:
246 adaptive trajectories in the fitness landscape over population structures translate to adap-
247 tive trajectories of model evidence over belief networks. Note that the word structure
248 here is borrowed from learning theory for consistency and it does not refer to structured
249 populations in population ecology.

250 Let us now turn specifically to the Bayesian interpretation of the evolution of indi-
251 viduality. Transitions in individuality, an evolutionary process in which lower-level units
252 that were previously capable of independent replication form a higher-level evolutionary
253 unit, correspond to a specific type transitions in the Bayesian model structure: either a
254 new node is added to the top of the network (in case there was no such population level
255 at all earlier), or a new value is added to any of the existing variables (in case the new
256 evolutionary unit is formed at an already existing level). In each case, most of the belief
257 network, including its parameters, remains the same, except the part that is participating
258 in the transition. This part, however, always involves only those values (corresponding to
259 types) of those variables (corresponding to levels) that are participating in the transition.
260 If average fitness of these types is larger by grouping them together, they undergo a tran-
261 sition in individuality. Although this is a general description of transitions disregarding
262 many details, the correspondence with Bayesian model comparison is remarkable.

263 Having defined our model framework mathematically, we now review its relation to
264 multilevel selection and transition theory in more detail. Multilevel selection is concep-
265 tually characterized into two types, dubbed multilevel selection 1 and 2, both assuming
266 that collectives form in a population of replicators, which themselves affect selection of
267 lower level units [25, 10, 6]. In case of multilevel selection 1 (MLS1), only temporary
268 collectives form that periodically disappear to revert to an unstructured population of
269 lower level units (transient compartmentation) [26, 27]. Multilevel selection 2 (MLS2) on
270 the other hand involves collectives that last and reproduce indefinitely, hence being bona
271 fide evolutionary units [28], see also [29]). Only if collectives are evolutionary units can
272 they inherit information stably (i.e., being informational replicators, [30]), thus the step

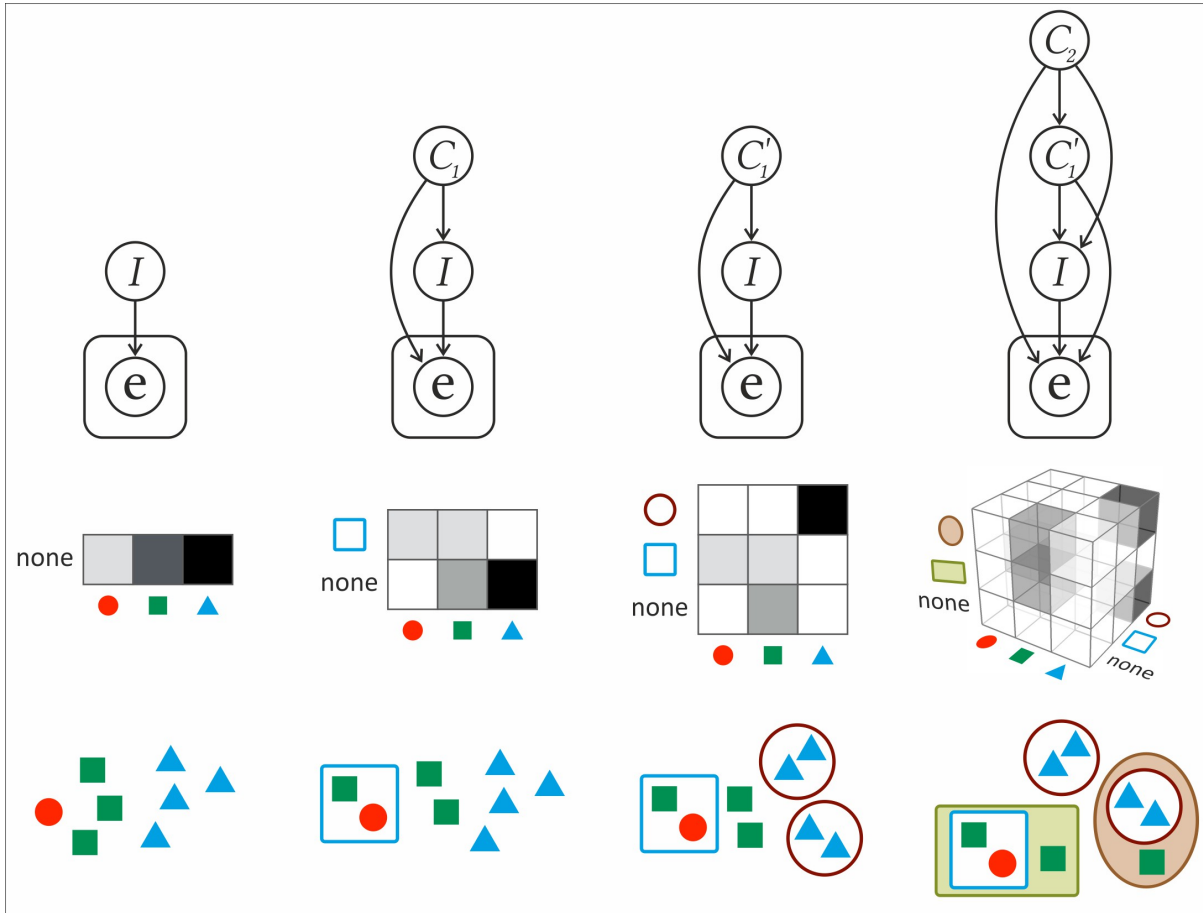


Figure 3: Evolutionary transitions as Bayesian structure learning. Initially, a single-level population I fits the environment e via replicator dynamics, or equivalently, via successive Bayesian update. Then, a new collective (the square) emerges at a new level C^1 , represented as a new node in the Bayesian belief network. Then, another new collective emerges at level C^1 (the circles), therefore, the variable C^1 is renamed to $C^{1'}$ as its possible values now include the circle as well. Finally, new collectives emerge at an even higher level (the rectangle and the ellipse at level C^2), and correspondingly, a new node is added to the network again. Note that the evolution of parameters (i.e., population composition in a fixed structure) is not illustrated here for simplicity.

273 toward a major evolutionary transition is MLS2. Note, that MLS1 can be understood as
274 kin selection for most of the cases (cf. [28]), and might not even be a necessary prerequi-
275 site for MLS2 to evolve. In general, compartmentalization itself (transient or not) is not
276 a sufficient property for a system to be a true evolutionary unit (cf. [31, 32]).

277 Our framework allows for parameterization of collective fitnesses such that they only
278 depend on the collective's composition, therefore corresponding to multilevel selection 2.
279 The model is capable of handling MLS1 if, at each timestep, individuals are randomly
280 reassorted among higher level collectives; incorporating this in the presented framework
281 here is left for future work. Here we focus on the step from MLS2 toward a major
282 transition: when collectives evolve to inherit information *above* their own composition. In
283 our model this corresponds to the case when a property of the collective appears, possibly
284 assigning different identities to collectives having identical composition. Such an identity-
285 providing piece of information is understood as an emergent property of the collective that
286 does not depend on the composition of lower level particles. If this is granted, higher level
287 units can evolve on their own, somewhat independent of their compositions. In biological
288 context, any such property corresponds to epigenetically inherited information that is not
289 coded by genes.

290 Let us conclude this section with some general remarks. First, in order to perform
291 explicit calculations, the fitness of each type at each level, i.e., $P(\mathbf{e}|I_i, C_j^1, C_k^2, \dots)$, has
292 to be specified. A natural way to do so is to pre-define a family of basis functions (e.g.,
293 Gaussians) on the space of possible environments \mathbf{e} , parametrized by a set of parameters
294 (e.g., the mean and covariance of the Gaussian). Then, each type at each level is assigned
295 one member of the family through its parameters. What determines the fitness of a given
296 type at time t then is the value of the basis function assigned to that type at $\mathbf{e}(t)$. The
297 advantage of such parametrization is threefold. First, it opens the possibility of model-
298 ing inter-type (i.e., microevolutionary) adaptation by making the parameters adaptive.
299 Second, genetic relatedness, a crucial determining factor of evolutionary transitions, can
300 be incorporated by coupling the parameters of types that have the similar containment
301 structure. Third, normalization of such basis functions over the space of possible envi-
302 ronments provides a natural way of accounting for adaptive trade-offs (i.e., the inability
303 of a single organism to adapt to multiple substantially different environments at the same
304 time). Here we do not enter into further details; investigating the relation between basis
305 function types, adaptation algorithms and generative models of the environment $P(\mathbf{e})$ is
306 the subject of future work.

307 3 Discussion

308 In this paper we introduced a mapping between concepts of hierarchical Bayesian models
309 and concepts of Darwinian evolution, providing a learning theory based interpretation of
310 complexification of life through evolutionary transitions of individuality. The backbone of
311 this interpretation is the fact that measuring the abundance and the composition of any
312 type at any level can be naturally mapped to performing marginalization and computing
313 conditional probabilities, respectively, of multivariate discrete probability distributions.
314 Another key ingredient is that the stochastic environment determines the fitness of both
315 individuals and collectives in a multilevel selection process. These two pillars are united by

316 the already known algebraic equivalence between Bayesian update and discrete replicator
317 dynamics. Accordingly, the learning theory narrative of multilevel selection is as follows:
318 as the environment e is successively observed, the distribution over the latent variables
319 I, C^1, C^2, \dots , corresponding to the hierarchical population composition, is successively
320 updated according to Bayes' rule.

321 Having identified this analogy, one might ask how the structure of the belief network
322 (i.e., not just the parameters of a fixed network) itself evolves. In learning theory, differ-
323 ent structures can be scored according to their model evidence, giving rise to Bayesian
324 model comparison, which accounts not only for how good a given solution is, but also for
325 how unlikely it is to find such a good solution in the parameter space. Consequently, this
326 procedure optimizes the trade-off between complexity and goodness of fit, hence dubbed
327 as automatic Occam's razor. The evolution of belief network structure, in the context
328 of Bayesian learning theory, is therefore driven by comparing model evidences of differ-
329 ent structures. Interestingly, Bayesian model comparison fits neatly to our multilevel
330 evolutionary dynamics interpretation: model evidence turns out to be equivalent to the
331 average fitness of individuals, i.e., of the lowest level replicating units. This allows for
332 a learning-theory based view of evolutionary transitions in individuality: units aggre-
333 gate to form a higher-level replicating unit if their average fitness increases by doing so;
334 this is mathematically equivalent to performing Bayesian model comparison between the
335 different belief network structures.

336 This procedure of simultaneous data acquisition, fitting, and structure learning is far
337 from unique to our proposed model framework; apart from its extensive use in machine
338 learning algorithms, it is conjectured to govern classified-as-intelligent systems such as
339 the conceptual development in children and also our collective understanding of the world
340 in terms of scientific concepts, both relying on the extraordinary generalization abilities
341 from sparse and noisy data [33, 34]. We argue, based on the mathematical equivalence
342 presented in this paper, that in order to devise seemingly-engineered complex organisms,
343 evolution, on Earth or anywhere, utilized comparable hierarchical learning mechanisms
344 as we humans do to make sense of the world around us.

345 Acknowledgements

346 The authors thank Ádám Radványi, András Szilágyi, András Hubai and Szabolcs Szá-
347 madó for their insightful comments on the manuscript. This research was funded by the
348 grant 'Theory and solutions in the light of evolution' (GINOP-2.3.2-15-2016-00057).

349 References

- 350 [1] John Tyler Bonner. *The evolution of complexity by means of natural selection*.
351 Princeton University Press, 1988.
- 352 [2] Peter A Corning and Eörs Szathmáry. "synergistic selection": a darwinian frame for
353 the evolution of complexity. *Journal of theoretical biology*, 371:45–58, 2015.
- 354 [3] John Maynard Smith and Eörs Szathmáry. *The major transitions in evolution*.
355 Oxford University Press, 1997.

- 356 [4] Stuart A West, Roberta M Fisher, Andy Gardner, and E Toby Kiers. Major evolu-
357 tionary transitions in individuality. *Proceedings of the National Academy of Sciences*,
358 112(33):10112–10119, 2015.
- 359 [5] Jordi van Gestel and Corina E Tarnita. On the origin of biological construction,
360 with a focus on multicellularity. *Proceedings of the National Academy of Sciences*,
361 114(42):11018–11026, 2017.
- 362 [6] Eörs Szathmáry. Toward major evolutionary transitions theory 2.0. *Proceedings of*
363 *the National Academy of Sciences*, 112(33):10104–10111, 2015.
- 364 [7] Corina E Tarnita, Clifford H Taubes, and Martin A Nowak. Evolutionary con-
365 struction by staying together and coming together. *Journal of theoretical biology*,
366 320:10–22, 2013.
- 367 [8] William C Ratcliff, Matthew Herron, Peter L Conlin, and Eric Libby. Nascent
368 life cycles and the emergence of higher-level individuality. *Phil. Trans. R. Soc. B*,
369 372(1735):20160420, 2017.
- 370 [9] Samuel R Levin, Thomas W Scott, Helen S Cooper, and Stuart A West. Darwin’s
371 aliens. *International Journal of Astrobiology*, pages 1–9, 2017.
- 372 [10] Samir Okasha. Multilevel selection and the major transitions in evolution. *Philosophy*
373 *of science*, 72(5):1013–1025, 2005.
- 374 [11] Marc Harper. The replicator equation as an inference dynamic. *arXiv preprint*
375 *arXiv:0911.1763*, 2009.
- 376 [12] Cosma Rohilla Shalizi et al. Dynamics of bayesian updating with dependent data
377 and misspecified models. *Electronic Journal of Statistics*, 3:1039–1074, 2009.
- 378 [13] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian net-
379 works: The combination of knowledge and statistical data. *Machine learning*,
380 20(3):197–243, 1995.
- 381 [14] Richard E Neapolitan et al. *Learning bayesian networks*, volume 38. Pearson Prentice
382 Hall Upper Saddle River, NJ, 2004.
- 383 [15] Richard A Watson and Eörs Szathmáry. How can evolution learn? *Trends in ecology*
384 *& evolution*, 31(2):147–157, 2016.
- 385 [16] Richard A Watson, Rob Mills, CL Buckley, Kostas Kouvaris, Adam Jackson, Si-
386 mon T Powers, Chris Cox, Simon Tudge, Adam Davies, Loizos Kounios, et al. Evo-
387 lutionary connectionism: algorithmic principles underlying the evolution of biological
388 organisation in evo-devo, evo-eco and evolutionary transitions. *Evolutionary biology*,
389 43(4):553–581, 2016.
- 390 [17] Daniel A Power, Richard A Watson, Eörs Szathmáry, Rob Mills, Simon T Powers,
391 C Patrick Doncaster, and Blažej Czapp. What can ecosystems learn? expanding
392 evolutionary ecology with learning theory. *Biology direct*, 10(1):69, 2015.

- 393 [18] Kostas Kouvaris, Jeff Clune, Loizos Kounios, Markus Brede, and Richard A Watson.
394 How evolution learns to generalise: Using the principles of learning theory to un-
395 derstand the evolution of developmental organisation. *PLoS computational biology*,
396 13(4):e1005358, 2017.
- 397 [19] James V Stone. *Bayes' rule: A tutorial introduction to Bayesian analysis*. Sebtel
398 Press, 2013.
- 399 [20] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.
- 400 [21] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*,
401 volume 1. MIT press Cambridge, 2016.
- 402 [22] Christopher M Bishop. *Pattern recognition and machine learning*. 2006.
- 403 [23] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and*
404 *techniques*. MIT press, 2009.
- 405 [24] Richard E Michod and Aurora M Nedelcu. On the reorganization of fitness dur-
406 ing evolutionary transitions in individuality. *Integrative and Comparative Biology*,
407 43(1):64–73, 2003.
- 408 [25] John Damuth and I Lorraine Heisler. Alternative formulations of multilevel selection.
409 *Biology and Philosophy*, 3(4):407–430, 1988.
- 410 [26] David Sloan Wilson. A theory of group selection. *Proceedings of the national academy*
411 *of sciences*, 72(1):143–146, 1975.
- 412 [27] David Sloan Wilson. Structured demes and trait-group variation. *The American*
413 *Naturalist*, 113(4):606–610, 1979.
- 414 [28] J Maynard Smith. Group selection. *THE QUARTERLY REVIEW OF BIOLOGY*,
415 51(2):277–283, 1976.
- 416 [29] András Szilágyi, István Zachar, István Scheuring, Ádám Kun, Balázs Könnyű, and
417 Tamás Czárán. Ecology and evolution in the rna world dynamics and stability of
418 prebiotic replicator systems. *Life*, 7(4):48, 2017.
- 419 [30] István Zachar and Eörs Szathmáry. A new replicator: a theoretical framework for
420 analysing replication. *BMC biology*, 8(1):21, 2010.
- 421 [31] Vera Vasas, Eörs Szathmáry, and Mauro Santos. Lack of evolvability in self-
422 sustaining autocatalytic networks constraints metabolism-first scenarios for the ori-
423 gin of life. *Proceedings of the National Academy of Sciences*, 107(4):1470–1475, 2010.
- 424 [32] Vera Vasas, Chrisantha Fernando, András Szilágyi, István Zachár, Mauro Santos,
425 and Eörs Szathmáry. Primordial evolvability: Impasses and challenges. *Journal of*
426 *theoretical biology*, 381:29–38, 2015.
- 427 [33] Charles Kemp and Joshua B Tenenbaum. The discovery of structural form. *Pro-*
428 *ceedings of the National Academy of Sciences*, 105(31):10687–10692, 2008.

- 429 [34] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman.
430 How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–
431 1285, 2011.