

1 State Dependent Behavioral Theory as a Means
2 for Understanding the Evolution of Play

3 Nicholas Grunloh¹ and Marc Mangel²

4 August 19, 2014

5 ¹ **Program** in Statistics and Applied Mathematics, ² Center for Stock Assessment Re-
6 search, University of California, Santa Cruz, CA, USA **Corresponding author**

7 Marc Mangel, Center for Stock Assessment Research, University of California, Santa
8 Cruz, CA 95064, USA.

9 Email: msmangel@ucsc.edu

10 **Abstract**

11 Understanding the evolutionary origins of play remains an unsolved problem. We show
12 how state dependent behavioral theory, as implement by Stochastic Dynamic Program-
13 ming can illuminate evolutionary biology of play. We first explain in generic terms
14 the components of a state dependent behavioral model. We use this framework to then
15 show how sexual play may be an incidental aspect of behavior during a non-breeding
16 season. We then develop a state dependent life history model on the assumption that
17 social play is an adaptive behavior and thus focus on play's contribution toward the
18 development of skill and how such development affects an individual's fitness. We dis-
19 cuss the interaction of theory and empirical work, and how each can reinforce the other.

20

21 **Keywords** Skill development, social play, stochastic dynamic programming, state de-
22 pendence, life history

23

24 Date received: ; revised;; accepted:

25 Introduction

26 Burdghardt (2005) identifies play according to the following criteria:

27 (i) Play is behavior that is not immediately necessary for survival.

28 (Although, play may coincidentally contribute developmental future fitness.)

29 (ii) Play is a self-motivating behavior; done for its own sake (...because play is “fun”).

30 (iii) Play is often a non-serious version of a similar non-play behavior.

31 (iv) Play is practiced often, yet loosely stereotyped and experimental in nature.

32 (v) Play only occurs in a stress free environment (i.e. the “relaxed field”).

33 These criteria do not define play, but they provide a clear framework for the sorts of
34 behaviors that can, and cannot, be considered play. In addition, they give some sense
35 of just how, and when, play can occur, for the purpose of guiding a model. The evolu-
36 tionary basis for play behavior is a cloudy topic, but if we consider a few fundamental
37 aspects of play, a structure for thinking about the topic emerges and it then becomes
38 clear how to make abstractions in order to formulate a model.

39 To being, consider behaviors that fall within the above criteria (e.g. kittens wrestling).
40 Caro (1995) identifies several specific costs and benefits to playing in cheetah cubs (*Ac-*
41 *nonyx jubatus*); see Figure 1. In short, the benefits of play can be thought of in terms
42 of the acquisition of skill to be used at some time in the future. Whether that skill takes
43 the form of maintenance of physical fitness, improved dexterity, or improved social
44 standing, we summarize these benefits in terms of a single quantity, the player’s skill.
45 In a similar way, the costs associated with play can be loosely grouped into manage-

46 able quantities. There are the costs associated with not playing (e.g. not maintaining
47 physical fitness) and the costs which occur while playing (e.g. injury and mortality).

48 The observation that play occurs in the presence of these costs, suggests that the
49 benefits of play outweigh the costs. Thus, it is reasonable to assume play behavior
50 has adapted in order to allow individuals the benefits of play, in the face of those costs
51 (Burghardt, 2006). Under the premise that play behavior is an evolutionary adaptation,
52 the acquisition of skill through play must be for the sake of the increasing future fitness.

53 If play is adaptive in this way, as opposed to a coincidental non-functional behav-
54 ior, then play decisions should follow some pattern of increasing an organism's fitness
55 through skill (i.e. decisions associated with play should be in some sense optimally
56 tuned to increases fitness). That is, even though individuals are driven to play because
57 it is "fun" the functional interpretation as to why play has become "fun" is that play
58 at a given period of development increases an organism's fitness at some time in the
59 future Burghardt (2006); Caro (1988). We use Burghardt's criteria for recognizing play
60 behavior as the rules of how and when play are allowed to occur, together with the
61 assumption that play occurs on the basis of increasing (or maximizing) fitness as a
62 foundation for modeling.

63 **Methods**

64 We begin with a description of the components of a state dependent behavioral model.
65 We then turn to the case in which play is incidental to other developmental processes,
66 considering that play may be costless or costly. After that analysis, we model a situa-
67 tion in which play is essential for the development of skills that contribute to subsequent
68 lifetime reproductive success. Neither of our models are intended to be 'realistic', in

69 the sense that we are not modeling a particular play situation by a particular animal.
70 However, we expect that the models have much in common with many different organ-
71 isms in many different situations, and thus may inspire other colleagues to get into the
72 specific details.

73 **The Components of a State Dependent Behavioral Theory**

74 *The Environment*

75 Organisms respond to the environment, so it must be described. In general, we must
76 consider acquisition of food in the environment and the risk of acquiring that food
77 (some older readers may recall the ‘crisis of the common currency’ in behavioral ecol-
78 ogy in the late 1970s and early to mid-1980s). Clearly, such risks are minimized for
79 juveniles whose parents provide food and shelter.

80

81 *States and Their Dynamics*

82 Organisms are interesting to us because they have physiological states (e.g. body mass,
83 length, fat reserves) that change in time, in response to the environment and behavior.
84 When we consider play that has fitness consequences, then another state variable is the
85 level of skill developed through play, and its effect on fitness.

86

87 *The Fitness Increment and Lifetime Fitness*

88 Biology is well-suited for economic thinking because there is a natural pay-off from
89 behavior: the representation of genes in future generations. Often, a proxy is used such
90 as accumulated lifetime reproductive success or the number of grand offspring (Man-
91 gel and Clark 1988, Clark and Mangel 2000, Mangel 2006). In the case of a juvenile

92 organism, the situation is simpler because by definition a juvenile is not accumulating
93 reproductive success. However, at the end of the juvenile period, we may assess future
94 fitness (Figure 1). The question is this: how do we associate play behavior during the
95 juvenile period with future fitness? The equations of Stochastic Dynamic Programming
96 (SDP) allow us to formalize this question mathematically and derive many predictions
97 – both quantitative and qualitative – that both can be tested empirically and provide
98 insight into the biological world.

99

100 *Thinking, Analysis, and Numerical Implementation*

101 As will be seen, deriving the equation of SDP forces one to think deeply about the
102 biology (which is, after all, the goal of this work). Often, some kinds of preliminary
103 analysis can be conducted on the model. However, and especially in the 21st century,
104 numerical solution of the SDP equation can provide exceptional insight – both qualita-
105 tive patterns and detailed numerical predictions. Indeed, as noted in Mangel and Clark
106 (1988) and Clark and Mangel (2000), very often the intuition from a numerical model
107 can be so powerful that one no longer needs the model to understand the phenomenon.
108 And that, of course, is what we are aiming for – understanding of the world.

109 **Play Is An Incidental Outcome of Development**

We begin by considering the case in which play is an incidental outcome of develop-
ment, so that skill is irrelevant to the fitness outcomes of play. We consider an individ-
ual during a non-breeding period, in which the individual may either forage for food, at
some risk, or remain in a central place that is devoid of both food and risk. As a state
variable, we choose energy reserves, denoted by $X(t)$ with specific value x (Table 1).

To characterize the changes in state, we require the costs of foraging a_f , of remaining at home (henceforth burrow) inactive a_0 , and of remaining at home and playing a_p and the energetic benefits of encountering food Y . Thus, if the individual remains in the burrow when $X(t) = x$

$$X(t+1) = x - a_0 \tag{1}$$

while if the individual plays in the burrow

$$X(t+1) = x - a_p. \tag{2}$$

If the individual chooses to forage and survives mortality, then with probability λ

$$X(t+1) = x - a_f + Y \tag{3}$$

and with probability $1 - \lambda$

$$X(t+1) = x - a_f. \tag{4}$$

110 We assume that there is a critical level x_c below which the individual dies from starva-
111 tion.

112 When foraging, the individual is at risk of mortality. We assume that the probability
113 of surviving a single period of foraging given that $X(t) = x$ is $\exp \left\{ - \left(m_0 + \frac{m_1}{x} \right) \right\}$
114 where m_0 and m_1 are the mass-independent and mass-dependent components of mor-
115 tality risk.

At the end of the non-breeding interval, the focal individual has future expected

reproductive success $\Phi(x)$ determined by its state $X(T)$. For computations, we choose the saturating function

$$\Phi(x) = \frac{x}{x + 0.25X_{max}} \quad (5)$$

We ask: given that $X(t) = x$ at time t within the non-breeding interval, do we predict that the individual will remain in the burrow (and play) or forage? To answer this question, we introduce the fitness function

$$F(x, t) = \max \mathbb{E} \left[\Phi(X(T)) \mid X(t) = x \right] \quad (6)$$

116 where ‘max’ indicates that we are to take the maximum over behavioral choices (here
 117 remain in the burrow or forage) and ‘ \mathbb{E} ’ denotes the expectation over the stochastic
 118 events of finding food or not and surviving predation. In words, $F(x, t)$ is the max-
 119 imum average value of future expected reproductive success at the end of the non-
 120 breeding interval given that energy reserves at time t are x .

At any time previous to T , the individual has three options. First, the individual may remain in its burrow and be inactive. In that case, given $X(t) = x$, the new state is $x - a_0$ and thus the fitness value of remaining in the burrow and resting is $V_r(x, t) = F(x - a_0, t + 1)$. Second, the individual may remain in the burrow and play, in which case the new state is $x - a_p$ and the fitness value of playing is $V_p(x, t) = F(x - a_p, t + 1)$. Third, the individual may forage. With probability λ it finds food so that the new state is $x - a_f + Y$ and with probability $1 - \lambda$ it does not find food so that its new state is $x - a_f$. In either case, it needs to survive predation. Thus,

the fitness value of foraging is

$$V_f(x, t) = \exp \left\{ - \left(m_0 + \frac{m_1}{x} \right) \right\} [\lambda F(x - a_f + Y, t + 1) + \lambda F(x - a_f, t + 1)]$$

By its definition, we then choose the largest of these three to determine fitness at time t

$$F(x, t) = \max[V_r(x, t), V_p(x, t), V_f(x, t)] \quad (7)$$

Eqn 7 is solved backwards in time, since we know its value at $t = T$ (see Mangel and Clark 1988, Clark and Mangel 2000 for details about how to do this). As this equation is solved we determine both the fitness function, and the optimal decision (to either remain in the burrow inactive or playing or to go foraging), $D^*(x, t)$. Intuition suggests, and computations will confirm, that in this case there is a threshold level of energy reserves $x_{th}(t)$ such that if $X(t) > x_{th}$ the optimal behavior is to remain in the burrow and that otherwise it is to play. Thus, animals do not need to solve an equation of dynamic programming; rather natural selection needs to act on this boundary.

129

130 *Costless Play and Forward Iteration*

When play is costless (i.e. $a_p = a_0$), the fitness values of remaining inactive in the burrow and playing in the burrow are the same (they would be different, for example, if there were mortality with play but not with resting; see the next section for ideas about that). In general, we cannot easily observe the internal states of animals and it is virtually impossible to observe the threshold $X_{th}(t)$. Rather, we observe whether animals are in the burrow or out foraging. To be able to make predictions about such observations, we use forward Monte Carlo Iteration (Mangel and Clark 1988, Clark

138 and Mangel 2000).

139 That is, imagine that we simulate the behavior of K individuals and let $X_k(t)$ de-
140 note the state of the k^{th} individual at time t . To begin, we specify the initial state,
141 $X_k(1)$ for each individual. We then go forward in time. At any time t , if an individual
142 is still alive we use the decision matrix $D^*(X_k(t), t)$ to determine whether the indi-
143 vidual remains in the burrow or not. If the decision is to remain in the burrow, then
144 $X_k(t+1) = X_k(t) - a_0$. If the decision is to forage, then we draw a random variable
145 uniformly distributed between 0 and 1, U , and compare it with λ . If $U \leq \lambda$ then this
146 individual finds food and $X_k(t+1) = X_k(t) - a_f + Y$; otherwise this individual does
147 not find food and $X_k(t+1) = X_k(t) - a_f$. If this value falls below the critical level
148 then the individual is dead. We then draw another random variable, still denoted by
149 U , to determine if the individual survives mortality by comparing the value of U and
150 $\exp \left\{ - \left(m_0 + \frac{m_1}{x} \right) \right\}$. In the next section, we summarize this kind of forward simulation
151 in a formal algorithm.

152 In this manner, we construct the state $X_k(t)$ for $k = 1, \dots, K$ and $t = 1, \dots, T$ (under-
153 standing that when an individual dies either through starvation or mortality, its state is
154 set at x_c for the entire rest of the forward iteration. Once this is done, we can construct
155 various statistics such as the number of individuals still alive at each time, the number
156 of individuals in the burrow (key for our purposes here), the mean and variance of the
157 state (something harder to observe in nature).

158

159 *The Fitness Effect of Costly Play*

160 When $a_p > a_0$ it is easy to see that play can never be optimal if it does not provide a
161 benefit to skill. We may ask, however, how much fitness is lost if the animal plays. John

162 McNamara and Alasdair Houston, in their seminal paper on state variable modeling
 163 of behavior, called this the ‘canonical cost’ of non-optimal behavior (McNamara and
 164 Houston 1986). We illustrate the idea here with a simple case. Imagine another value
 165 of the state $x_p > x_{th}$ such that if $x > x_p$ then the animal, which is already in the burrow,
 166 plays rather than rests (so that its state the next period is $x - x_p$ rather than $x - x_0$).

Now let $F_p(x, t)$ denote the value of $\mathbb{E} \left[\Phi(X(T) | x) \right]$ given that the animal plays
 whenever $x \geq x_p$. As before, $F(x, T) = \Phi(x)$. For previous times, we proceed as
 follows. If the optimal decision when $X(t) = x$ is to forage then

$$F_p(x, t) = \exp \left\{ - \left(m_0 + \frac{m_1}{x} \right) \right\} [\lambda F_p(x - a_f + Y, t + 1) + \lambda F_p(x - a_f, t + 1)] \quad (8)$$

If the optimal decision when $X(t) = x$ is to remain in the burrow and $x < x_p$ then

$$F_p(x, t) = F_p(x - a_0, t + 1) \quad (9)$$

while if $x \geq x_p$

$$F_p(x, t) = F_p(x - a_p, t + 1) \quad (10)$$

We solve Eqns 8-10 backwards in time as we did for $F(x, t)$ and define the fitness cost
 $C(x, t)$ of playing when $X(t) = x$ as

$$C(x, t) = \frac{F(x, t) - F_p(x, t)}{F(x, t)} \quad (11)$$

167 Play Develops Fitness Enhancing Skills

168 In order to simplify the dynamics of social play in the model, we consider a focal
169 individual separately from all of the other potential play partners in the environment.
170 Individuals can have skill levels ranging from a minimum skill, S_L , to a maximum
171 skill, S_U . At time t , an individual has level of skill $S(t)$, with particular value noted by
172 i . Similarly, potential play partners have particular skill levels denoted by j . Each time
173 period of the model, the skill of the focal individual decrements by, α , to capture the
174 idea that skill requires maintenance through repeated practice. At each time period, the
175 focal individual may encounter a play partner, or not; if a partner is encountered, then
176 the focal individual decides to play, or not.

177 We now assume that **ontogenic development ??redundant??** of the skill stops at time
178 T , at which time an individual with skill level i has future lifetime fitness $\phi(i)$ (Figure
179 2). For previous times we define a fitness function (Mangel & Clark (1988) as well as
180 Clark & Mangel (2000))

$$F(i, t) = \max \mathbb{E} \left[\phi(S(T)) \right]. \quad (12)$$

181 In this equation \mathbb{E} denotes the expectation over encounters with potential partners
182 and \max denotes the maximum over the behavioral decision to play or not. Thus,
183 individuals behave adaptively in that they choose whether or not to play based on maxi-
184 mizing their future fitness, not necessarily their immediate fitness. By considering focal
185 individuals with a range of skill levels at any given time within the model, we are able
186 to see how factors independent of energy reserves and predation affect an organism's
187 decision to play. **something about deep mathtalk about these details ??**

189 *Play Events*

190 We assume that all play partners are willing and available to enter play events with the
191 focal individual, contingent on the focal individual's decision whether or not to play
192 with them. When a play event occurs between the focal individual, of skill i , and a
193 play partner, of skill j , the focal individual receives an increment to its skill denoted
194 $\Delta S(i, j)$. In order to capture the idea that skill associated with play events is not
195 necessarily acquired instantaneously, the skill increment, $\Delta S(i, j)$, of a particular play
196 event is awarded to the focal individual a number of time periods, τ , after the play event
197 starts. Since individuals incur a per period decrement to their skill, α , every period of
198 the model, and it takes τ time periods to gain skill from a play event, it follows that the
199 total decrement to skill of a single play event is $\alpha\tau$ and the net change in skill when a
200 player of level i plays with a partner of level j is $\Delta S(i, j) - \alpha\tau$.

201 In cases where play events collide with the time horizon of the model, T , we re-
202 place $t + \tau$ by T and assume that the focal individual receives the net increment in skill
203 $\Delta S(i, j) - \alpha\tau$. Doing this keeps the relationship between skill increments and skill
204 decrements for truncated play events consistent with all other time periods. Although
205 it is worth noting that the effective per period decrement in these truncated play events
206 is higher due to compressing the total cost, $\alpha\tau$, over less than τ time steps.

207

208 *Skipping Play Events and Exiting the Playing Field*

209 The focal individual may skip a play event in a time period may be because of being
210 unable to find an appropriate play partner, or because the available play partners in
211 the environment do not allow $\Delta S(i, j)$ to be greater than $\alpha\tau$. In this case, the focal
212 individual only incurs the per period cost to skill, α , for a single time period.

213 Caro's (1988, 1995) results suggest that different types of play occur at differing
 214 periods of development and thus a model of play behavior must include the ability of
 215 playing organisms to stop considering social play as a behavioral option altogether.
 216 Thus, we include the behavioral option of exiting the play field entirely.

217

The Skill Increment

We assume that when individuals whose skills are closely matched play, the increment in skill is greater than if the individuals have skills that are widely different (Burghardt, 2006). Thus, $\Delta S(i, j)$ reaches a maximum, S_{max} , when $i = j$, and as i becomes more different from j , $\Delta S(i, j)$ decreases. For the computation we use the symmetric form

$$\Delta S(i, j) = \Delta S_{max} \exp \left\{ -\frac{(i - j)^2}{2\sigma^2} \right\}. \quad (13)$$

218 Here σ is a parameter that describes how similar the focal individual must be to the
 219 play partner in order to receive a meaningful skill increment from a play event (Figure
 220 3). $\Delta S(i, j)$ will always be maximized when the focal individual and the play partner
 221 have the same skill (i.e. $i = j$). Notice that the symmetry of Eqn 2 means that $\Delta S(i, j)$
 222 does not really depend on either i or j , but rather the absolute difference between i and
 223 j .

224 As a thought experiment to help understand how focal individuals are motivated by
 225 the acquisition of skill through $\Delta S(i, j)$, consider a focal individual that makes play
 226 decisions based only on the effects of those behaviors in the short-term.

227 Such a myopic focal individual only considers whether a play event causes an in-
 228 crease or decrease in skill, regardless of any ill effects these decisions may cause in
 229 further time periods. For the myopic focal individual the decision to play, or not, is

really just a comparison between the skill decrement of the play event, $\alpha\tau$, and the skill increment, $\Delta S(i, j)$. If $\Delta S(i, j)$ is greater than $\alpha\tau$ then the myopic individual will always play, regardless of how small the difference, and if $\alpha\tau$ is the greater than $\Delta S(i, j)$, the myopic individual will never play. However, as long as $\tau > 1$ (which we assume it always is), there is an opportunity cost associated with playing with a poorly matched partner. Thus, optimally behaving individuals consider factors that introduce opportunity costs and lead to more selective behavior than in the myopic case.

237

Play Partners

We characterize the play environment through a probability distribution of potential partners. That is, we let $\lambda_j(t)$ be the probability that a focal individual encounters a potential play partner of skill j at t . For computations, we use an exponential distribution

$$\lambda_j(t) = \delta_n \exp\{-cj\} \quad (14)$$

where c is a scale parameter and δ_n is a normalization constant chosen so that $\sum_j \lambda_j(t) \leq 1$; $\lambda_0 = (1 - \sum_j \lambda_j(t))$ is the case in which the focal individual cannot find any play partner. The distribution of the potential social play partners in the environment, as an exponential, translates into an environment with initially many low skill individuals. As potential play partners develop their own skills, and leave the population, a decreasing number of high skill individuals are left in the population.

244

Fitness Function and SDP Equation

We assume that $\phi(i)$ is an increasing function of i , consistent with a higher level of skill at T providing greater future fitness. For computations, we choose a logistic function

that has some threshold skill level after which fitness rises rapidly with skill (Figure 2):

$$\phi(i) = \frac{(i - S_L)^\gamma}{(i - S_L)^\gamma + (S_o - S_L)^\gamma}. \quad (15)$$

245 Thus, $\phi(i)$ is normalized to a maximum value of 1, S_o is the skill at which half
 246 maximal fitness is achieved, and γ characterizes how quickly fitness increases with
 247 increased skill near the skill threshold.

In light of the definition of $F(i, t)$ we have the end condition $F(i, T) = \phi(i)$. At each time prior to the the time horizon, T , an individual with skill level i may exit the play field, thus obtaining future fitness $\phi(i)$ or may continue to seek partners. We let $V_{cont}(i, t)$ denote the future expected fitness of an individual with skill level i at time t who continues to seek play partners. Then

$$F(i, t) = \max[\phi(i), V_{cont}(i, t)] \quad (16)$$

248 the fitness value of continuing to seek play partners depends upon the skill level of the
 249 partner encountered at time t , so that

$$\begin{aligned} V_{cont}(i, t) = & \left(1 - \sum_j \lambda_j(t)\right) F(i - \alpha, t + 1) + \\ & \left(\sum_j \lambda_j(t)\right) \max[F(i + \Delta S(i, j) - \alpha\tau, t + \tau), F(i - \alpha, t + 1)] \end{aligned} \quad (17)$$

250 For example if the focal individual does not encounter a play partner (the first term
 251 on the right hand side) it is not awarded any skill, but still incurs the per period cost
 252 to skill, α . If a play partner of skill level j is encountered in period t , with probability
 253 $\lambda_j(t)$, the focal individual must decide between entering a play event or skipping the

254 play event with the encountered play partner of skill j . If the focal individual decides
 255 to play, the focal individual's skill is incremented by $\Delta S(i, j)$ and decremented by α
 256 for every period of the play event. Encountering a play partner and choosing not to play
 257 has the same fitness effect as having not encountered a partner in the first place. The
 258 solution of Eqns 17-19 leads to two matrices of decisions depending upon i, j , and t .
 259 The first, $D_e^*(i, t)$ characterizes whether an individual with skill level i at time t exits
 260 the play field or not. The second, $D_p^*(i, j, t)$, characterizes whether an individual, of
 261 skill i , chooses to play with a partner, of skill j , at time t of the model.

262

263 *Monte Carlo Implementation of Play Decisions Forward in Time*

264 To predict the behaviors of individuals, we use $D_e^*(i, t)$ and $D_p^*(i, j, t)$ to run a Monte
 265 Carlo simulation forward through time (Mangel and Clark (1988), Clark and Mangel
 266 (2000)). In particular, we simulate a **number of focal individuals, k** , making optimal
 267 play decisions as predicted by Eqns 16-17. We begin by considering K focal individ-
 268 uals in a much larger play field and assign skill levels that are randomly drawn from
 269 a uniform distribution on $[S_L, S_U]$. In each time period of the simulation, each of the
 270 focal individuals encounter a potential play partner drawn randomly from the probabil-
 271 ity distribution of encountering potential play partners of skill j . At each potential play
 272 encounter the focal individual either enters a play event, skips a play event, or exits the
 273 model according to the decision matrices, at the particular i, j, t conditions of the given
 274 play encounter. The simulation follows the following algorithm for each of the K focal
 275 individuals:

276 (1) $t = 0$

277 (2) Randomly assign the k^{th} focal individual an initial skill, $I_k(0)$, between S_L and S_U .

- 278 (3) Randomly draw a potential play partner skill level, J , from Eqn 14.
- 279 (4) Look up the appropriate play decision, $D_p^*(I_k(t), J, t)$.
- 280 (5.1) If the play decision is “play”; $I_k(t+\tau) = I_k(t) + \Delta S(I_k(t), J) - \alpha\tau$ and $t \rightarrow t+\tau$.
- 281 (5.2) If the play decision is “skip”; $I_k(t+1) = I_k(t) - \alpha$ and $t \rightarrow t+1$.
- 282 (5.3) If the play decision is “exit”; $I_k(t+1) = I_k(t)$ and $t \rightarrow T$.
- 283 (6.1) If $t < T$ go to step (3).
- 284 (6.2) If $t \geq T$ then $I_k(T) = I_k(t)$.

285 **Results**

286 We separately discuss the different cases of play being incidental in development and
 287 play essential for the development of life history skills.

288 **Play Is Incidental**

289 In Figure 4, we show the boundary x_{th} for three values of Y , which measures the
 290 richness of the environment when play is costless (i.e. $a_p = a_0 = 1$). It accords with
 291 intuition that in richer environments, even if individuals have the same probability of
 292 finding food, they need to forage less frequently, so can remain in the burrow – thus
 293 avoiding the risk of predation. In the case of an environment that is less rich, even
 294 when an individual finds food the amount of food found does not increase the state
 295 as much, hence the boundary is lower. This boundary translates into the fraction of
 296 individuals remaining in the burrow as a function of time, which we show in Figure 5.

297 Three phenomena appear in this figure. First, early in the time series, say $t = 1 - 15$,
 298 no individual is in the burrow, simply because states are so low. For moderate times,
 299 say $t = 15 - 50$, play is more frequent in the richer environments. Finally, for $t > 50$,
 300 we see the effect of the end condition, interacting with the richness of the environment
 301 and the risk of mortality when foraging.

302 In Figure 6a, we show a heat map for the fitness cost of playing when $a_p = 2$, and
 303 the threshold for playing is $x_p = 75$. Note that play is predicted to have very small
 304 fitness consequences except for large values of state, and that it is predicted to have
 305 some consequences, of the order of 4 % for values of state between about 50 or 60 and
 306 x_p , depending upon time. However, in Figure 6b we show the average and standard
 307 deviation of the state in this case and we see that on average state is about 40, so that
 308 even though play is costly, the cost is rarely paid.

309 **Play Develops Skills**

310 *From the Backward Equation*

311 Focal individuals choose to play with a range of similarly skilled individuals about the
 312 diagonal of $D_p^*(i, j, t)$ where $i = j$ (Figure 7). If the cost of play, $\alpha\tau$, is larger than,
 313 $\Delta S(i, j)$, a focal individual is predicted not to play with the partner whose skill level is
 314 j . Thus, $\alpha\tau$ is a major driver in determining the extent to which i must be similar to j
 315 in order for the focal individual to enter a play event.

We observe patterns in the total range of playable j 's based on the focal individuals skill and the time period of the model in which a play event occurs. That is, at some t and i , there exists a maximum j that is beneficial for i to play with; we denote this maximum playable j , \hat{J}_i . Similarly there is some minimum j that is beneficial for i to

play with, denoted by \check{J}_i . We show the total range of potential play partners for every combination of i and t

$$R(i, t) = \hat{J}_i - \check{J}_i. \quad (18)$$

in Figure 8.

Forward Monte Carlo Simulation

In Figure 9a we show the final skill distribution of $k = 100000$ individuals making optimal decisions for 40 periods. This distribution is a clear trimodal distribution. From the left to right, firstly we see the mode at about skill 15 representing the most common skill for individuals who have not yet exited the model. Secondly, the mode around skill 30 is representative of the accumulation of individuals that begin the simulation below the exit threshold, but play to build enough skill to exit the playing stage. Thirdly, the right-most mode around a skill of 35, represents individuals that enter the model with skills high enough to automatically exit play behavior. In Figure 9b we show the same final skill distribution of $k = 100000$ individuals making optimal play decisions when the initial skill allocation is truncated to skills just prior to the automatic exit skill. Comparison of Figures 9a and 9b demonstrate the basis of the Figure 9a's third mode around skill 35.

Another way of visualizing the results of Figure 9 is in the scatter plot seen in Figure 10, in which we show the relationship between the initial and final levels of skill of the simulated individuals. The dotted red one-to-one line in Figure 10 shows the final skill level required to maintain the initial skill level. By considering individuals that enter the model with a uniform distribution over the range of possible skills, we ensure that the results will cover all of the possible play strategies in the environment.

336 In such a case, of course, some individuals start with high enough skill to exit the model
337 immediately. These are the individuals with initially high skill, on the one-to-one line
338 in the region labeled “Exit”. Individuals with initial skills below the initial exit skill all
339 play to some degree, but the lower the initial skill the more selective the play decisions
340 become. With lower levels of skill, individuals are predicted to be more selective when
341 choosing a play partner. Playing organisms that have high enough final skills to find
342 themselves above the one-to-one line, in the region labeled “Lucky” are individuals
343 that were able to successfully find the play partners that they need to improve their skill
344 from their initial state. Playing organisms that end up below the one-to-one line, in the
345 region labeled “Unlucky” are individuals seeking play partners, but were not able to
346 find the play partners that they need to improve their skill. For low skill individuals it
347 is relatively easy to find appropriate play partners, and thus they most often end up in
348 the “Lucky” region.

349 Discussion

350 The major message of our first model is that even if play is costly and has no direct
351 consequences on the development of life history skills, the fitness costs of play may be
352 relatively small, so that we predict individuals will play even if it is not adaptive. Our
353 second model assumes a relaxed field (sense Burghardt, 2006), and thus allows us to
354 understand skill-dependent play decisions independent of energy reserves or predation
355 risk while playing. However, it is relatively easily to consider play behavior with re-
356 spect to these factors, but as a starting point it is instructive to understand the basics
357 of play behavior within this simple model first. As more intricate models are made on
358 play behavior, added considerations may make it hard to see some of the basic forces

359 driving play behavior as seen in this model. In this model, for simplicity, we have al-
360 lowed play events between a single focal individual and a single play partner at one
361 time, but there is no reason that this has to be the case. For example litters of kittens
362 often play in groups. This may present interesting results considering that the results of
363 this model suggest that playing individuals tend to develop skill as a group. In addition
364 to adding multiple play partners, adding mortality would give insight into a potentially
365 strong cost of play t .

366 As described above, there is a pocket of time and skill where the general patterns
367 do not hold true (Figure 9). We propose that this can be explained by the finite time
368 horizon of the model, and its relation to play events as defined by the model. Recall
369 that for time periods near T , play events cause $t + \tau$ to be greater than T . Due to the
370 construction of the model the skill increments and decrements for play events in these
371 periods are consistent with all other time periods of the model, however the fitness
372 values associated with these skill levels must be truncated at $F(i, T) = \phi(i)$ because
373 by definition fitnesses for time periods beyond T are defined by $\phi(i)$. This has the
374 effect of decreasing $R(i, t)$ for time periods just prior to the final time periods of the
375 model. Skills high enough to exit the model have lower than expected values for $R(i, t)$
376 several time periods before these individuals exit the model. Individuals several time
377 steps before the end of the model are predicted to be very selective in their choice of
378 play partners because the fitness associated with any skill level in these time periods of
379 the model has been truncated to $F(i, T) = \phi(i)$. Since play is not actually bounded in
380 this way, it is useful to run the model with large values of T and consider the general
381 trends prior to this exception pocket.

382 When considering the general trends o prior to the exception pocket, we find that

low skill individuals are predicted to be relatively selective in their play decisions, seeking with other low skill individuals. As individuals gain high skill, they become more willing to play with individuals of very dissimilar skill levels. In nature, high skill individuals may often self-handicap; this could be included in a version of the model in which self-handicapping is another behavioral decision.

As shown in Figure 8, the Monte Carlo, individuals with initially low skill (perhaps the most common natural occurrence) play to increase their skill, and on average they increase their skill level and exit play behavior in the same proportions as other playing individuals. However, one may expect that individuals entering the model with high pre-exit skill levels should have a developmental advantage, and exit the model more quickly and in higher proportions. In general this is not the case, unless playing individuals enter the model virtually at the the exit threshold. Generally, individuals with initially high pre-exit skill levels quickly fit into very similar skill distributions as individuals with initially low skill. This is due to the scarcity of favorable play partners in the pre-exit upper skill range. On average individuals in a confined social environment will develop their skill as a group. Regardless of an initially playing individuals initial skill, the skill development of all individuals in the group converges toward the average skill development of the group.

Individuals with initially very high skill are immediately able to exit. In these cases play behavior is never displayed. This is clearly a hypothetical, and largely unattainable situation for many social species, but these initially exiting individuals could have a meaningful interpretation when one considers behaviors that are not learned via play, or even the evolution of innate behaviors or reflexes.

406 **Acknowledgements**

407 We thank the Mangel Lab Group for helpful comments during the development of these
408 ideas and **XX, YY, and ZZ** for comments on the manuscript. This work was conducted
409 as a part of the Evolution of Play Working Group at the National Institute for Mathemat-
410 ical and Biological Synthesis, sponsored by the National Science Foundation, the U.S.
411 Department of Homeland Security, and the U.S. Department of Agriculture through
412 NSF Awards EF-0832858 and DBI-1300426, with additional support from The Univer-
413 sity of Tennessee, Knoxville.

414 **References**

- 415 [1] Burghardt, G. M. (2006). **The genesis of animal play: Testing the limits**. MIT
416 Press, Cambridge, MA.
- 417 [2] Caro, T. M. (1980). Effects of the mother, object play, and adult experience on
418 predation in cats. *Behavioral and Neural Biology*, 29, 29-51.
- 419 [3] Caro, T. M. (1988). Adaptive significance of play: Are we getting closer?. *TREE*,
420 3(2), 50-54.
- 421 [4] Caro, T. M. (1988). Adaptive significance of play: Are we getting closer?. *TREE*,
422 3(2), 50-54.
- 423 [5] Caro, T. M. (1995). Short-term costs and correlates of play in cheetas. *Animal*
424 *Behaviour*, 49, 333-345.
- 425 [6] Clark, C. and Mangel, M. (2000). **Dynamic state variable models in ecology:**
426 **Methods and Applications**. Oxford University Press, New York.
- 427 [7] Mangel, M. (2006). **The Theoretical Biologist's Toolbox. Quantitative Methods**
428 **for Ecology and Evolutionary Biology**. Cambridge University Press, Cambridge,
429 UK.
- 430 [8] Mangel, M. and Clark, C. (1988). **Dynamic modeling in behavioral ecology**.
431 Princeton University Press, Princeton, NJ.

- 432 [9] McNamara, J.M. and Houston, A.I. 1986. The common currency for behavioral
433 decisions. *The American Naturalist* 127: 358-378.

Table 1. Variables, Parameters, Their Interpretation, and Values for the Model in Which
⁴³⁵ **Play is Incidental.**

<i>Symbol</i>	<i>Interpretation</i>	<i>Value</i>
t	Time in the non-breeding period	$t=1-70$
T	End time	70
$X(t)$	State variable at time t	Eqn XX
X_{max}	Maximum value of state	100
x	Particular value of $X(t)$	$x=1-X_{max}$
a_f	Cost of foraging	2
a_0	Cost of remaining at home inactive	1
a_p ⁴³⁶	Cost of remaining at home playing	1 (costless play) 2 (costly play)
Y	Value of food if found	13
λ	Probability of finding food	0.15
x_c	Critical level for starvation	5
m_0	Mass-independent rate of mortality	
m_1	Mass-dependent rate of mortality	
$\Phi(x)$	Future expected reproductive success, given $X(T) = x$	Eqn XX

Table 1 continued

<i>Symbol</i>	<i>Interpretation</i>	<i>Value</i>
t	Time in the non-breeding period	t=1-70
$D_e^*(x, t)$	Optimal behavioral decision given that $X(t) = x$	Output of Eqn 7
K^{438}	Number of simulated individuals	100
$X_k(t)$	State of the k_{th} simulated individual	Simulation
U	Uniformly distributed random variable	$0 \leq U \leq 1$
x_p	Threshold for costly play	$x_p = 75$
F_p	Fitness function with costly play	Eqn XX

Table 2. Variables, Parameters, Their Interpretation, and Values for the Model in Which
⁴⁴⁰ **Play Develops Skills.**

<i>Symbol</i>	<i>Interpretation</i>	<i>Value</i>
$S(t)$	Skill of focal individual at time t	Eqn XX
i	Particular value of skill level	Varies
j	Particular value of the skill of a possible play partner	Varies
S_L	Minimum level of skill	0
S_U	Maximum level of skill	50
α	Per-period loss of skill	0.9
T	Time at which development of skill stops	40
⁴⁴¹ $\phi(i)$	Future lifetime fitness for an individual with $S(T) = i$	Eqn 15, Figure 2
$F(i, t)$	Fitness function for an individual with $S(t) = i$	Eqn 16, Figure 2
$\Delta S(i, j)$	Gain in skill when an individual whose whose skill is i plays with an individual whose skill is j	Eqn 13, Figure 3
τ	Length of a play bout	3
$\lambda_j(t)$	Probability focal individual encounters a play partner with skill level j	Eqns 14, 15

Captions for Figures

Figure 1. By definition a juvenile is not accumulating reproductive success. However, at the end of the juvenile period, which lasts from $t = 1$ to $t = T$ in our models, fitness is assessed depending upon the state variables. The equations of SDP allow us to formalize the link between state variables, future fitness, and behavior during the juvenile period.

Figure 2. Three possible choices for terminal fitness of an individual with skill level i , $\phi(i)$. Notice the greater the steepness parameter γ the more quickly and dramatically the organism matures once it reaches adolescence.

Figure 3. The increment in skill $\Delta S(i, j)$ for player with skill level i playing against a player with skill level j . The horizontal line at $\alpha\tau$ is the decrement in skill during the play interval, so that net gain in skill is $\Delta S(i, j) - \alpha\tau$.

Figure 4. The boundary x_{th} for three values of Y , which measures the richness of the environment. For values of $x > x_{th}$ we predict that the individual will remain in its burrow.

Figure 5. The fraction of individuals remaining in the burrow as a function of time, with the associated standard deviation.

Figure 6. a) A heat map for the fitness cost of playing when $a_p = 2$, and the threshold for playing is $x_p = 75$. b) The average and standard deviation of the state in this

466 case and we see that on average state is about 40.

467

468 Figure 7. The focal individual fitness plotted against skill level. Each line is a single
469 time period of the model. Three time periods of the model are plotted. Notice when
470 many time periods remain in the model, fitness is relatively high for all skill levels, due
471 to the prospect of gaining skill in the future. As the number of periods remaining in the
472 model decreases, the fitness of low skill individuals decreases due to reduced prospect
473 for the future. Additionally, the dotted vertical lines mark the skill at which $F(i, t)$ con-
474 verges with $\phi(i)$. These dotted lines mark the skill at which the focal individual stops
475 considering play behavior at the given time period of the model. Notice that with many
476 time periods of the model remaining only very high skill individuals exit the model, and
477 as the number of time periods remaining in the model decreases this exit skill decreases.

478

479 Figure 8. A grey scale representation of the focal individual play range as a func-
480 tion of both time and focal individual skill level. Dark cells are representative of focal
481 individuals willing to play with play partners of many different skill levels, while light
482 cells are representative of focal individuals with relatively small play ranges. In general
483 as skill increases focal individual play range increases. Additionally as t approaches T ,
484 in general, play range increases to the myopic condition, at $T - 1$. However, a pocket of
485 lower than expected play ranges does violate these general trends. This pocket occurs
486 at relatively high values for t and extends across all of the playing skill levels. This
487 pocket is produced by truncating play events as t approaches T .

488

489 Figure 9. Final skill distribution of $k = 100000$ Monte Carlo simulated individu-
 490 als. Each individual makes optimal decisions, based on $D_p^*(i, j, t)$, for 40 time periods.
 491 a) Individuals start the simulation with a random uniformly chosen skill level on the
 492 interval $[S_L, S_U]$. Notice the a trimodal final skill distribution. b) Individuals start the
 493 simulation with a random uniformly chosen skill level on the interval $[S_L, 34]$. Notice
 494 the resulting bimodal distribution of the final skills, due to the lack of initially exiting
 495 individuals.

496

497 Figure 10. Final skill distribution of $k = 10000$ Monte Carlo simulated individuals
 498 plotted against the initial skill distribution. The red dotted line indicates the one-to-one
 499 relationship between initial and final skill. Individuals on the one-to-one line, in the
 500 region labeled “Exit”, enter the simulation with high enough skills to immediately exit
 501 play behavior. Notice for each initial skill below the initial exit skill, the final skill
 502 distributions are very similar.