

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

3 Nicholas Grunloh¹ and Marc Mangel²

4 March 27, 2014

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

7 **Corresponding author**

8 Marc Mangel, Center for Stock Assessment Research and Department of Applied Math-
9 ematics and Statistics, University of California, Santa Cruz, CA 95064, USA.

10 Email: msmangel@ucsc.edu

11 **Abstract**

12 Understanding the evolutionary origins of play remains an unsolved problem. We show
13 how state dependent behavioral theory, as implement by Stochastic Dynamic Program-
14 ming (SDP) can illuminate the evolutionary biology of play. We first explain in generic
15 terms the components of a state dependent behavioral model. We then use this frame-
16 work to show how play may be an incidental aspect of behavior during a non-breeding
17 season. After that, we develop a state dependent life history model on the assumption
18 that social play is an adaptive behavior contributing toward the development of skills
19 that affect an individual's lifetime fitness. We discuss the interaction of theory and em-
20 pirical work, and how each can reinforce the other.

21

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

24

25 Date received: ; revised:; accepted:

26 Introduction

27 Burdghardt (2005) identifies play according to the criteria that

28 (i) Play is a behavior that is non-essential to the immediate survival of the playing
29 organism. (However, it could contribute to future fitness.)

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

31 (iii) Play differs from any serious version of a similar non-play behavior.

32 (i.e. play can be a non-serious version of other types of behaviors)

33 (iv) Play is heavily repeated (i.e. practiced often), yet loosely stereotyped.

34 (i.e. aspects of play behavior are learned or experimental in nature)

35 (v) Play only occurs in a stress free environment (a “relaxed field”).

36 (e.g. an environment with adequate food, that is free of predation or intense competi-
37 tion)

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

44 It is possible that play is purely incidental to ontogeny – that the neurological de-
45 velopment requires the networks associated with play be established before other neu-
46 ral networks can be built (see Burghardt 1988). In this case, play is likely not to
47 be adaptive, and may even be costly. However, it is important to be able to assess
48 those costs. On the other hand, nearly all organisms – even some of the simplest ones

49 (Jennings1976)– need to develop life history skills during ontogeny. This can be done
50 in many ways (Fagen 1981), either alone (e.g. Mather 1999) or in a social setting; in
51 this paper we explore the role of intraspecific social interaction via play as a means of
52 developing such skills, based on the assumption that social play facilitates the develop-
53 ment of life history skills with subsequent fitness consequences. In short, the benefits
54 of play can be summarized as the acquisition of skill to be used some time in the future
55 (e.g. Fagen and Fagen 2004, 2009). In a similar way, the costs associated with play
56 can be loosely grouped into manageable quantities. There are the costs associated with
57 not playing (e.g. loss of skill) and the costs that occur while playing (e.g. injury and
58 mortality). The observation that play occurs in the presence of its costs, suggests that
59 the benefits of play outweigh the costs. Thus, it is reasonable to assume play behavior
60 has adapted in order to allow individuals the benefits of play, in the face of those costs
61 (Burghardt, 2006).

62 If play is adaptive in this way, as opposed to a coincidental non-functional behav-
63 ior, then play decisions must follow some pattern of increasing an organism’s fitness
64 through skill (i.e. play occurs because it increases fitness). That is, even though indi-
65 viduals are driven to play because it is “fun” the functional interpretation as to why play
66 has become “fun” is that play at a given period of development increases an organism’s
67 fitness at some time in the future (Burghardt 2006; Caro 1988; Fagen 1981; Fagen and
68 Fagen 2004, 2009). We use Burghardt’s criteria for recognizing play behavior as the
69 rules of how and when play are allowed to occur, together with the assumption that play
70 occurs on the basis of increasing (or maximizing) fitness as a foundation for the model.

71 **Methods**

72 We begin with a description of the components of a state dependent behavioral model.
73 We then turn to the case in which play is incidental to other developmental processes,
74 considering that play may be costless or costly. After that analysis, we model a situa-
75 tion in which play is essential for the development of skills that contribute to subsequent
76 lifetime reproductive success. Neither of our models is intended to be ‘realistic’, in the
77 sense that we are not modeling a particular play situation by a particular animal. How-
78 ever, we expect that the models have much in common with many different organisms in
79 many different situations, and thus may inspire other colleagues to get into the specific
80 details.

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

82 *The Environment*

83 Organisms respond to the environment, so it must be described. In general, we must
84 consider acquisition of food in the environment and the risk of acquiring that food
85 (some older readers may recall the ‘crisis of the common currency’ in behavioral ecol-
86 ogy in the late 1970s and early to mid-1980s). Clearly, such risks are minimized for
87 juveniles whose parents provide food and shelter. In the case of social play, we also
88 need to characterize the availability of potential partners.

89

90 *States and Their Dynamics*

91 Organisms are interesting to us because they have physiological states (e.g. body mass,
92 length, fat reserves) that change in time, in response to the environment and behavior.

93 When we consider play that has fitness consequences, skill developed through play be-
94 comes another state variable. We will give specific examples below.

95

96 *The Fitness Increment and Lifetime Fitness*

97 Biology is well-suited for economic thinking because there is a natural pay-off from
98 behavior: the representation of genes in future generations. Often, we use a proxy such
99 as accumulated lifetime reproductive success or the number of grand offspring (Man-
100 gel and Clark 1988, Clark and Mangel 2000, Mangel 2006). In the case of a juvenile
101 organism, the situation is simpler because by definition a juvenile is not accumulating
102 reproductive success. However, at the end of the juvenile period, we may assess future
103 fitness (Figure 1). The question is this: how do we associate play behavior during the
104 juvenile period with future fitness? The equations of Stochastic Dynamic Programming
105 (SDP) allow us to formalize this question mathematically and derive many predictions
106 – both quantitative and qualitative – that both can be tested empirically and provide
107 insight into the biological world.

108

109 *Thinking, Analysis, and Numerical Implementation*

110 As will be seen, deriving the equation of SDP forces one to think deeply about the
111 biology (which is, after all, the goal of such work). Often, some kinds of preliminary
112 analysis can be conducted on the model. However, and especially in the 21st century,
113 numerical solution of the SDP equation can provide exceptional insight – both qualita-
114 tive patterns and detailed numerical predictions. Indeed, as noted in Mangel and Clark
115 (1988) and Clark and Mangel (2000), very often the intuition from a numerical model
116 can be so powerful that one no longer needs the model to understand the phenomenon.

117 And that, of course, is what we are aiming for – understanding the natural world.

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

We begin by considering the case in which play is an incidental outcome of development, so that skills is irrelevant to the fitness outcomes of play. We consider an individual 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 $1 - \lambda$

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

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

121 When foraging, the individual is at risk of mortality. We assume that the probability
 122 of surviving a single period of foraging given that $X(t) = x$ is $\exp(-m_0 - \frac{m_1}{x})$ where
 123 m_0 and m_1 are the mass-independent and mass-dependent components of mortality
 124 risk, so that larger individuals have lower rates of predation.

At the end of the non-breeding interval, the focal individual has future expected reproductive success $\Phi(x)$ determined by its state $X(T)$ at that time. 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

$$F(x, t) = \max E[\Phi(X(T)) | X(t) = x] \quad (6)$$

125 where ‘max’ indicates that we are to take the maximum over behavioral choices (here
 126 remain in the burrow, resting or playing, or forage) and ‘ E ’ denotes the mathematical
 127 expectation over the stochastic events of finding food or not and surviving predation. In
 128 words, $F(x, t)$ is the maximum average value of future expected reproductive success
 129 at the end of the non-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

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

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.

138

139 *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

147 and Mangel 2000).

148 That is, imagine that we simulate the behavior of K individuals and let $X_k(t)$ denote
149 the state of the k^{th} individual at time t . To begin, we specify the initial state, $X_k(1)$ for
150 each individual. We then go forward in time. At any time t , if an individual is still alive
151 we use the decision matrix $D^*(X_k(t), t)$ to determine whether the individual remains
152 in the burrow or not. If the decision is to remain in the burrow, then $X_k(t+1) =$
153 $X_k(t) - a_0$. If the decision is to forage, then we draw a random variable uniformly
154 distributed between 0 and 1, U , and compare it with λ . If $U \leq \lambda$ then this individual
155 finds food and $X_k(t+1) = X_k(t) - a_f + Y$; otherwise this individual does not find food
156 and $X_k(t+1) = X_k(t) - a_f$. If this value falls below the critical level then the individual
157 is dead. We then draw another random variable, still denoted by U , to determine if the
158 individual survives mortality by comparing the value of U and $\exp(-m_0 - \frac{m_1}{X_k(t)})$. In
159 the next section, we summarize this kind of forward simulation in a formal algorithm.

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

166

167 *The Fitness Effect of Costly Play*

168 When $a_p > a_0$ it is easy to see that play can never be optimal if it does not provide a
169 benefit to skill. We may ask, however, how much fitness is lost if the animal plays. John
170 McNamara and Alasdair Houston, in their seminal paper on state variable modeling

171 of behavior, called this the ‘canonical cost’ of non-optimal behavior (McNamara and
 172 Houston 1986). We illustrate the idea here with a simple case. Imagine another value
 173 of the state $x_p > x_{th}$ such that if $x > x_p$ then the animal, which is already in the burrow,
 174 plays rather than rests.

We now let $F_p(x, t)$ denote the value of $E[\Phi(X(T)|x)]$ 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

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

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

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

175 **Play Develops Fitness Enhancing Skills**

176 In order to simplify the dynamics of social play , we consider a focal individual sepa-
 177 rately from all of the other potential play partners in the environment and ignore physi-
 178 ological state. This would occur, for example, if juvenile individuals are fed by parents
 179 and do not have to forage themselves. Individuals can have skill levels ranging from a

180 minimum skill, S_L , to a maximum skill, S_U . We let $S(t)$ denote the level of skill at time
 181 t , with particular value noted by i . Similarly, potential play partners have particular skill
 182 levels denoted by j . In one timer period, the skill of the focal individual decrements by
 183 α regardless of activity, to capture the idea that skill requires maintenance through re-
 184 peated practice. At each time period, the focal individual may encounter a play partner
 185 or not, and if a partner is encountered, then decides to play or not.

186 We now assume that ontogenic development of the skill stops at time T , at which
 187 time an individual with skill level i has future lifetime fitness $\phi(i)$ (Figure 2). As before,
 188 we define a fitness function

$$F(i, t) = \max E\{\phi(S(T))\}. \quad (12)$$

189 In this equation E denotes the expectation over encounters with potential partners
 190 and \max denotes the maximum over the behavioral decision to play or not. Thus,
 191 individuals behave adaptively in that they choose whether or not to play based on maxi-
 192 mizing their future fitness, not necessarily their immediate fitness. By considering focal
 193 individuals with a range of skill levels at any given time within the model, we are able
 194 to see how factors independent of energy reserves and predation affect an organism's
 195 decision to play. happening here then in the deep math talk about these details) ??

196

197 *Play Events*

198 We assume that all play partners are willing and available to enter play events with
 199 the focal individual, contingent on the focal individual's decision to play with them.
 200 When a play event occurs between the focal individual, of skill i , and a play partner,
 201 of skill j , the focal individual receives an increment to its skill $\Delta S(i, j)$. In order
 202 to capture the idea that skill associated with play events is not necessarily acquired

203 instantaneously, the skill increment, $\Delta S(i, j)$, of a particular play event is awarded to
 204 the focal individual a number of time periods, $\tau > 1$, after the play event starts. Since
 205 individuals incur a per period decrement to their skill every period of the model, and it
 206 takes τ time periods to gain skill from a play event, it follows that the total decrement
 207 to skill of a single play event is $\alpha\tau$ and the net change in skill when a player of level i
 208 plays with a partner of level j is $\Delta S(i, j) - \alpha\tau$.

209 In cases where play events collide with the time horizon of the model, T , we re-
 210 place $t + \tau$ by T and assume that the focal individual receives the net increment in skill
 211 $\Delta S(i, j) - \alpha\tau$. Doing this keeps the relationship between skill increments and skill
 212 decrements for truncated play events consistent with all other time periods.

213

214 *Skipping Play Events and Exiting the Playing Field*

215 The focal individual may skip a play event in a time period because of being unable to
 216 find an appropriate play partner, or because the available play partners in the environ-
 217 ment do not allow $\Delta S(i, j) > \alpha\tau$. In this case, the focal individual only incurs the per
 218 period cost to skill, α , for a single time period.

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

The Increment in Skill

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 computations we use the symmetric form

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

219 Here σ is a parameter that describes how similar the focal individual must be to the play
 220 partner in order to receive a meaningful skill increment from a play event (Figure 3).
 221 $\Delta S(i, j)$ will always be maximized when the focal individual and the play partner have
 222 the same skill (i.e. $i = j$). Notice that the symmetry of Eq.(13) means that $\Delta S(i, j)$
 223 does not really depend on either i or j , but rather the absolute difference between i and
 224 j .

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

Such a myopic focal individual only considers whether a play event causes an increase or decrease in skill, regardless of any effects these decisions may cause in further time periods. For the myopic focal individual the decision to play or not is thus 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 playing with a poorly matched partner in terms of lost time. Thus, optimally behaving individuals consider factors that introduce opportunity costs and lead to more selective behavior than in the myopic case.

Play Partners

We characterize the play environment through a probability distribution of potential partners. We let

$$\lambda_j(t) = Pr(\text{a focal individual encounters a potential play partner of skill } j \text{ at time } t). \quad (14)$$

For computations, we use an exponential distribution

$$\lambda_j(t) = \delta_n(t) \exp\{-cj\} \quad (15)$$

where c is a scale parameter and $\delta_n(t)$ 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.

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):

$$\phi(i) = \frac{(i - S_L)^\gamma}{(i - S_L)^\gamma + (S_o - S_L)^\gamma}. \quad (16)$$

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

In light of the definition of $F(i, t)$ we have the end condition

$$F(i, T) = \phi(i). \quad (17)$$

At each time $t < T$, 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

$$F(i, t) = \max[\phi(i), V_{cont}(i, t)] \quad (18)$$

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

$$\begin{aligned} V_{cont}(i, t) = & \left(1 - \sum_j \lambda_j(t)\right) F(i - \alpha, t + 1) + \\ & \sum_j \lambda_j(t) \max[F(i + \Delta S(i, j) - \alpha\tau, t + \tau), F(i - \alpha, t + 1)] \end{aligned} \quad (19)$$

233 For example if the focal individual does not encounter a play partner (the first term
234 on the right hand side) it is not awarded any skill, but still incurs the per period cost to
235 skill, α . If a play partner of skill level j is encountered in period t , with probability
236 $\lambda_j(t)$, the focal individual must decide between entering or skipping a play event with
237 the encountered play partner of skill j . If the focal individual decides to play with that
238 partner it is awarded the skill increment for playing with that play partner, $\Delta S(i, j)$,
239 and a skill decrement, α , for every period of the play event. Encountering a partner
240 and choosing not to play has the same fitness effect as having not encountered a partner
241 The solution of Eqns 17-19 leads to a two matrices of decisions depending upon i, j ,
242 and t . The first, $D_e^*(i, t)$ characterizes whether an individual with skill level i at time t
243 exits the play field or not. The second, $D_p^*(i, j, t)$, characterizes whether that individual

chooses to play with a partner whose skill level is j or not, if such a partner is encountered.

Monte Carlo Implementation of Play Decisions Forward in Time

To predict the behaviors of individuals, we use $D_e^*(i, t)$ and $D^*(i, j, t)$ to run a Monte Carlo simulation forward through time (Mangel and Clark 1988, Clark and Mangel 2000). In particular, we simulate a number of focal individuals, k , making optimal play decisions as predicted by Eqs. (18,19). We begin by considering K focal individuals in a much larger play field and assign skills levels that are randomly drawn from a uniform distribution on $[S_L, S_U]$. In each time period of the simulation, each of the focal individuals encounters a potential play partner drawn randomly from the probability distribution of encountering potential play partners of skill j . At each potential play encounter the focal individual either enters a play event, skips a play event, or exits the model according to the decision matrices, at the particular i, j, t conditions of the given play encounter. The simulation follows the following algorithm for each of the K focal individuals:

(1) $t = 0$

(2) Draw a random uniform focal individual skill level of individual k , $I_k()$, on the interval $[S_L, S_U]$.

(3) Draw a random potential play partner skill level, J , from Eq.(15).

(4) Look up the appropriate play decision, $D_p^*(I_k(t), J, t)$.

(5.1) If the play decision is play; $I_k(t + \tau) = I_k(t) + \Delta S(I_k(t), J) - \alpha\tau$ and $t \rightarrow t + \tau$.

266 (5.2) If the play decision is skip; $I_k(t + 1) = I_k(t) - \alpha$ and $t \rightarrow t + 1$.

267 (5.3) If the play decision is exit; $I_k(t + 1) = I_k(t)$ and $t \rightarrow T$.

268 (6.1) If $t < T$ go to step (3).

269 (6.2) If $t \geq T$ then $I_k(T) = I_k(t)$.

270 **Results**

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

273 **Play Is Incidental**

274 In Figure 4, we show the boundary x_{th} for three values of Y , which measures the
275 richness of the environment when play is costless (i.e. $a_p = a_0 = 1$). It accords with
276 intuition that in richer environments, even if individuals have the same probability of
277 finding food they need to forage less frequently, so can remain in the burrow – thus
278 avoiding the risk of predation. In the case of an the environment that is less rich, even
279 when an individual finds food the amount of food found does not increase the state
280 as much, hence the boundary is lower. This boundary translates into the fraction of
281 individuals remaining in the burrow as a function of time, which we show in Figure 5.
282 Three phenomena appear in this figure. First, early in the time series, say $t = 1 - 15$,
283 no individual is in the burrow, simply because states are so low. For moderate times,
284 say $t = 15 - 50$, play is more frequent in the richer environments. Finally, for $t > 50$,

we see the effect of the end condition, interacting with the richness of the environment and the risk of mortality when foraging.

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

Play Develops Skills

From the Backward Equation

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

$$R(i, t) = \hat{J}_i - \check{J}_i. \quad (20)$$

in Figure 8.

Forward Monte Carlo Simulation

In Figure 9 we show the final skill distribution of $k = 250$ individuals making optimal decisions for 40 periods. This distribution appears to be bimodal, and if k is increased, the final skill distribution becomes a clear bimodal distribution. The mode centered around skill 30 is representative of the accumulation of all exiting individuals throughout the modeled periods. The mode centered around skill 15 is the most common skill for individuals who have not yet exited the model.

Another way of visualizing the results of Figure 9 is in the scatter plot 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. In such a case, of course, some individuals start with high enough skill to exit the model immediately. These are the individuals with initially high skill, on the one-to-one line in the region labeled “Exit”. Individuals with initial skills below the initial exit skill all play to some degree, but the lower the initial skill the more selective the play decisions become. With lower levels of skill, individuals are predicted to be more selective when choosing a play partner,. Playing organisms that have high enough final skills to find themselves above the one-to-one line, in the region labeled “Lucky” are individuals that were able to successfully find the play partners that they need to improve their skill from their initial state. Playing organisms that end up below the one-to-one line, in the

325 region labeled “Unlucky” are individuals seeking play partners, but were not able to
326 find the play partners that they need to improve their skill. For low skill individuals it
327 is relatively easy to find appropriate play partners, and thus they most often end up in
328 the “Lucky” region.

329 **Discussion**

330 The major message of our first model is that even if play is costly and has no direct
331 consequences on the development of life history skills, the fitness costs of play may be
332 relatively small, so that we predict individuals will play even if it is not adaptive. Our
333 second model assumes a “relaxed field”(sense Burghardt, 2006), and thus allows us to
334 understand skill-dependent play decisions independent of energy reserves or predation
335 risk while playing. However, it is relatively easily to consider play behavior with re-
336 spect to these factors, but as a starting point it is instructive to understand the basics
337 of play behavior within this simple model first. As more intricate models are made on
338 play behavior, added considerations may make it hard to see some of the basic forces
339 driving play behavior as seen in this model. In this model, for simplicity, we have al-
340 lowed play events between a single focal individual and a single play partner at one
341 time, but there is no reason that this has to be the case. For example litters of kittens
342 often play in groups. This may present interesting results considering that the results of
343 this model suggest that playing individuals tend to develop skill as a group. In addition
344 to adding multiple play partners , adding mortality would give insight into a potentially
345 strong cost of play t.

346 As described above, there is a pocket of time and skill where the general patterns
347 do not hold true (Figure 9). We propose that this can be explained by the finite time

348 horizon of the model, and its relation to play events as defined by the model. Recall
 349 that for time periods near T , play events cause $t + \tau$ to be greater than T . Due to the
 350 construction of the model the skill increments and decrements for play events in these
 351 periods are consistent with all other time periods of the model, however the fitness
 352 values associated with these skill levels must be truncated at $F(i, T) = \phi(i)$ because
 353 by definition fitnesses for time periods beyond T are defined by $\phi(i)$. This has the
 354 effect of decreasing $R(i, t)$ for time periods just prior to the final time periods of the
 355 model. Skills high enough to exit the model have lower than expected values for $R(i, t)$
 356 several time periods before these individuals exit the model. Individuals several time
 357 steps before the end of the model are predicted to be very selective in their choice of
 358 play partners because the fitness associated with any skill level in these time periods of
 359 the model has been truncated to $F(i, T) = \phi(i)$. Since play is not actually bounded in
 360 this way, it is useful to run the model with large values of T and consider the general
 361 trends prior to this exception pocket.

362 When considering the general trends o prior to the exception pocket, we find that
 363 low skill individuals are predicted to be relatively selective in their play decisions, seek-
 364 ing with other low skill individuals . As individuals gain high skill, they become more
 365 willing to play with individuals of very dissimilar skill levels. In nature, high skill in-
 366 dividuals may often self-handicap; this could be included in a version of the model in
 367 which self-handicapping is another behavioral decision.

368 As shown in Figure 8, the Monte Carlo, individuals with initially low skill (perhaps
 369 the most common natural occurrence) play to increase their skill, and on average they
 370 increase their skill level and exit play behavior in the same proportions as other playing
 371 individuals. However, one may expect that individuals entering the model with high

372 pre-exit skill levels should have a developmental advantage, and exit the model more
373 quickly and in higher proportions. In general this is not the case, unless playing indi-
374 viduals enter the model virtually at the the exit threshold. Generally, individuals with
375 initially high pre-exit skill levels quickly fit into very similar skill distributions as indi-
376 viduals with initially low skill. This is due to the scarcity of favorable play partners in
377 the pre-exit upper skill range. On average individuals in a confined social environment
378 will develop their skill as a group. Regardless of an initially playing individual's initial
379 skill, the skill development of all individuals in the group converges toward the average
380 skill development of the group.

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

386 **Acknowledgements**

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

394 **Literature Cited**

395 Burghardt, G.M. 1988. Precocity, play, and the Ectotherm-Endotherm transition. pg
396 107-148 in **Handbook of Behavioral Neurobiology, Vol 9**, E.M. Blass (ed.), Plenum
397 Publishing Corporation.

398 Burghardt, G. M. (2006). **The genesis of animal play: Testing the limits**. MIT
399 Press, Cambridge, MA

400 Caro, T. M. (1980). Effects of the mother, object play, and adult experience on
401 predation in cats. *Behavioral and Neural Biology*, 29, 29-51.

402 Caro, T. M. (1988). Adaptive significance of play: Are we getting closer?. *TREE*,
403 3(2), 50-54.

404 Caro, T. M. (1995). Short-term costs and correlates of play in cheetas. *Animal*
405 *Behaviour*, 49, 333-345.

406 Clark, C. and Mangel, M. (2000). **Dynamic state variable models in ecology:**
407 **Methods and Applications**. Oxford University Press, New York

408 Fagen, R. 1981. **Animal Play Behavior** Oxford University Press, New York

409 Fagen, R. and J. Fagen. 2004. Juvenile survival and benefits of play behavior in
410 brown bears, *Ursus arctos*. *Evolutionary Ecology Research* 6:89-102

411 Fagen, R. and J. Fagen. 2009. Play behavior and multi-year juvenile survival in
412 free-ranging brown bears, *Ursus arctos*. *Evolutionary Ecology Research* 11:1053-
413 1067

414 Jennings, H.S. 1976. **Behavior of the Lower Organisms**. Indiana University Press,
415 Bloomington, IN

416 Mangel, M. (2006). **The Theoretical Biologist's Toolbox. Quantitative Methods**
417 **for Ecology and Evolutionary Biology**. Cambridge University Press, Cambridge, UK

418 Mangel, M. and Clark, C. (1988). **Dynamic modeling in behavioral ecology.**
419 Princeton University Press, Princeton, NJ

420 Mather. J.A. 1999. Exploration, play, and habituation in octopuses (*Octopus dofleini*).
421 Journal of Comparative Psychology 113:333-338

422 McNamara, J.M. and Houston, A.I. 1986. The common currency for behavioral
423 decisions. The American Naturalist 127: 358-378

424

425 **Table 1. Variables, Parameters, Their Interpretation, and Values for the Model**

426 **in Which 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	Equation 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
427 a_p	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

428 **Table 1 continued**

	<i>Symbol</i>	<i>Interpretation</i>	<i>Value</i>
	t	Time in the non-breeding period	t=1-70
	$d^*(x, t)$	Optimal behavioral decision given that $X(t) = x$	Output of Eqn 7
429	K	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

430 **Table 2. Variables, Parameters, Their Interpretation, and Values for the Model**

431 **in Which 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	XX
S_U	Maximum level of skill	XX
α	Per-period loss of skill	
T	Time at which development of skill stops	XX
432 $\phi(i)$	Future lifetime fitness for an individual with $S(T) = i$	Eqn 16, Figure 2
$F(i, t)$	Fitness function for an individual with $S(t) = i$	Eqn XX
$\Delta S(i, j)$	Gain in skill when an individual whose whose skill is i plays with an individual whose skill is j	Eqn 13, Figure 3 Eqn XX
τ	Length of a play bout	XX
$\lambda_j(t)$	Probability focal individual encounters a play partner with skill level j	Eqns 14, 15

433 **Captions for Figures**

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

439

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

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

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

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

451 Figure 6 a) A heat map for the fitness cost of playing when $a_p = 2$, and the threshold
452 for playing is $x_p = 75$. b) The average and standard deviation of the state in this case
453 and we see that on average state is about 40,

454 Figure 7 The focal individual fitness plotted against skill level. Each line is a single
455 time period of the model. Three time periods of the model are plotted. Notice when
456 many time periods remain in the model, fitness is relatively high for all skill levels, due

457 to the prospect of gaining skill in the future. As the number of periods remaining in the
 458 model decreases, the fitness of low skill individuals decreases due to reduced prospect
 459 for the future. Additionally, the dotted vertical lines mark the skill at which $F(i, t)$
 460 converges with $\phi(i)$. These dotted lines mark the skill at which the focal individual
 461 stops considering play behavior at the given time period of the model. Notice that with
 462 many time periods of the model remaining only very high skill individuals exit the
 463 model, and as the number of time periods remaining in the model decreases this exit
 464 skill decreases.

465 Figure 8 A grey scale representation of the focal individual play range as a function
 466 of both time and focal individual skill level. Dark cells are representative of focal
 467 individuals willing to play with play partners of many different skill levels, while light
 468 cells are representative of focal individuals with relatively small play ranges. In general
 469 as skill increases focal individual play range increases. Additionally as t approaches T ,
 470 in general, play range increases to the myopic condition, at $T - 1$. However, a pocket of
 471 lower than expected play ranges does violate these general trends. This pocket occurs
 472 at relatively high values for t and extends across all of the playing skill levels. This
 473 pocket is produced by truncating play events as t approaches T .

474 Figure 9 Final skill distribution of $k = 250$ Monte Carlo simulated individuals.
 475 Each individuals starts the simulation with a uniform random skill level on the interval
 476 $[S_L, S_U]$. Each individual makes optimal decisions, based on $D^*(i, j, t)$, for 40 time
 477 periods. Notice the bimodal distribution of the final skills.

478 Figure 10 Final skill distribution of $k = 250$ Monte Carlo simulated individuals
 479 plotted against the initial skill distribution. The red dotted line indicates the one-to-one
 480 relationship between initial and final skill. Individuals on the one-to-one line, in the

481 region labeled “Exit”, enter the simulation with high enough skills to immediately exit
482 play behavior. Notice for each initial skill below the initial exit skill, the final skill
483 distributions are very similar, both to each other, and to the final skill distribution seen
484 in Figure 7.