Between Frustration and Elation: Sense of Control Regulates the lntrinsic Motivation for Motor Learning

Beata J. Grzyb
Robotic Intelligence Lab, Jaume I University, Castellon, Spain, Email: grzyb@uji.es
Graduate School of Eng., Osaka University, Suita, Japan

Joschka Boedecker and **Minoru Asada**
Graduate School of Eng., Osaka University, Suita, Japan

Angel P. del Pobil
Robotic Intelligence Lab, Jaume I University, Castellon, Spain
Interaction Science Dept., Sungkyunkwan University, Seoul, Korea

Linda B. Smith
Cognitive Development Lab, Indiana University, Bloomington, USA

Abstract

Frustration has been generally viewed in a negative light and its potential role in learning neglected. We propose a new approach to intrinsically motivated learning where frustration is a key factor that allows to dynamically balance exploration and exploitation. Moreover, based on the result obtained from our experiment with older infants, we propose that a temporary decrease in learning from negative feedback can also be beneficial in fine-tuning a newly learned behavior. We suggest that this temporal indifference to the outcome of an action may be related to the sense of control, and results from the state of elation, that is the experience of overcoming a very difficult task after prolonged frustration. Our preliminary simulation results serve as a proof-of-concept for our approach.

Human infants are born with a tremendous amount of intrinsic motivation to explore the properties of their own bodies and the nearest surroundings. This motivation is aimed primarily towards actions whose results are immediately observable. The contingency between actions and their results can be rewarding and encourages to continue those valuable actions. Disruption of a learned contingency between behavior and reward leads to negative emotional reactions, even if rewards are still delivered but are not dependent on infants' actions. These negative emotional reactions, commonly known as frustration, have been generally viewed in a negative light. The hypothesis proposed by Wong (Wong 1979), that an optimal level of frustration leads to exploration and faster learning, has not found much attention from the scientific community.

On the other hand, many models of intrinsic motivation for artificial learning systems found their inspiration in Berlyne's famous monograph *Conflict, Arousal, and Curiosity* (Berlyne 1960). In there, exploration or information seeking has been greatly considered as a primary source of motivation. Thus, a significant number of models driven by novelty (Weng 2002; Barto, Singh, and Chentanez 2004; Marshall, Blank, and Meeden 2004), and curiousity (Schmidhuber 2010) has been proposed. The basic concept behind curiosity-driven models is that artificial agents are interested in learnable but yet unknown regularities and get bored by both predictable and inherently unpredictable things (eg. white noise). The mismach between expectations and reality is translated into curiosity rewards, which propels agents to actively create surprising events in the environmnet and thus learn novel patterns.

More recently a new computational approach to intrinsic motivation that is based on competence has been suggested (Baranes and Oudeyer 2010). In this framework, an agent sets up a "challenge", that is a self-determined goal associated with measures of difficulty and measures of actual performance. Herein, interesting learning challenges are those which promise the largest learning progress based on the agent's current level of competence.

In this paper, we also suggest a competence-based approach with the important difference that task performance does not affect which of several possible tasks should be selected, but rather how exploration and exploitation should be balanced while learning one particular task. The advantage of our method is that it changes this balance dynamically based on the level of competence of the agent. We propose to use a notion of *sense of control* that, in our understanding, is one's subjective sense of the capacity to successfully perform a desired action, fulfill the individual personal goals and desires, or instinctual drives and needs. A lack of such an ability causes the feeling of frustration, and decreases the overall sense of control. On the other hand, the experience of overcoming a very difficult challenge after prolonged frustration due to many trials and errors may result in an increase of one's sense of control. Therefore, frustration and sense of control are inversely related. In line with Wong's suggestion (Wong 1979), we assume that a medium (optimal) level of frustration leads to more explorative behavior, while low levels lead to exploitation.

The next section introduces basic concepts of our experiment with older infants along with a short discussion on the main finding from this work. In section III, we present our synthetic approach for designing an intrinsically motivated system. Section IV introduces basic concepts of our first approach to a model implementation and provides the details of our experiment with a simulated robot. We close the paper with conclusions and discussions of follow-up research.

Observation data

The primary motive for our experiment was to see how infants' knowledge about their own body capabilities changes with the acquisition of new motor skills. A reaching action was a good candidate for our test, as to sucessfully perform this action infants need to know not only the distance to the object, but also how far they can reach and lean forward without losing balance. Infants master this skill quite early in their development. As we were interested in how body perception changes with age, our experimental group consisted of 9-month-old infants (N=8) and 12-month-old infants (N=8). The basic setup of the experiment is shown in Fig. 1. The procedure of the experiment was like the following (for the details please refer to (Grzyb, del Pobil, and Smith a)). Infants were seated in a specially adapted car seat with the seatbelts fastened for security reasons. In order to keep infants engaged and attentive during the entire experimental session, a colorful stimuli display was placed in front of them. The colorful display also helped in separating the experimenter from the infants, making communication between infants and the experimenter impossible. A ball attached to a wooden dowel appeared through the opening of the frame at various distances (30, 37, 47, 60, 70 cm). The sequence of trials consisted of 9 distances and always begun and ended with two trials at close distances to keep infants motivated. The order of distances in the middle of the sequence was chosen pseudo-randomly. The sequence of distances was repeated up to three times. There was no explicit reward provided to the infants after the trial for any tested distance. This helped us to avoid situations where infants could learn to make reaching movements just to communicate their interest in obtaining a reward. The entire experimental session was recorded with two cameras. These recordings were subsequentally viewed and infants' behavior scored.

The results of the experiments showed that 12-month-old, but not 9-month-old infants constantly reached for the out-of-reach objects, which was quite surprising as typically we would expect older infants to know more than younger ones. As 12 months is the age around when the transition to walking occurs, we decided to extend our experiment and recruit more infants depending on their walking abilities (Grzyb, del Pobil, and Smith b). We segregated 12-month-old infants into three groups: non-walkers (N=8), walkers with help (N=8), and walkers without help (N=8). To see how reaching for far objects changes during the experimental session, we calculated the mean percentage of reaches for far distances for every sequence of trials. Fig. 2 shows the results for 60 cm distance. For comparison we also provide results for 9-month-old infants. It can easily be noticed that all 12-month-old infants reached for the out-of-reach object the first time, but only walkers (with or without help) continued reaching the second time and the third time. The probability of reaching, however, slightly decreased the third time, which may suggest that walkers indeed learn what is within their reachable space, but the learning rate in their case is much lower than in the case of non-walkers and 9-month-old infants.

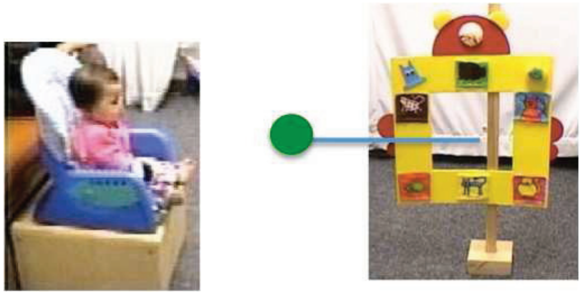


Figure 1: Experimental setup.

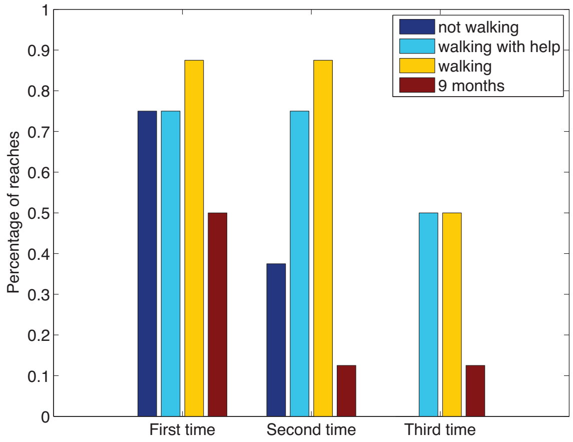


Figure 2: Mean percentage of reaches to far objects (60cm) for 12-month-old infants: not able to walk (navy blue), able to walk with help (light blue), or able to walk without help (yellow), and 9-month-old infants (red). Please notice that we use the term "time" here, and that these are not consecutive trials. There are several trials to various distances between the first and the second time presented for a given distance.

In our opinion, such a slow rate of learning what is reachable or not, makes infants excercise more their walking behavior, as a primary motive for walking is to reach for something. It is possible that if infants learned faster what is within their reachable space there would be less incentive for mastering further walking behavior.

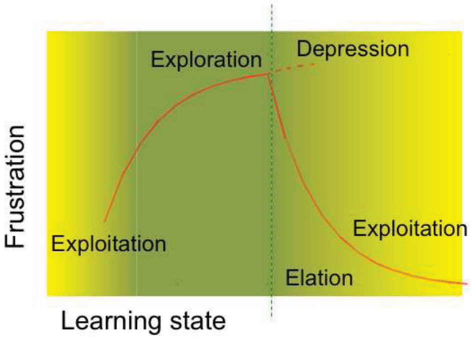


Figure 3: The dynamics of balancing between exploration and exploitation based on the level of frustration

Our approach

We favor a synthetic approach that brings together explanation and design theory and is expected to fill the gap between existing disciplines instead of staying in one closed discipline, to further our understanding of human cognitive development (Asada et al. 2009).

The basic premise of our approach is that a need for control is innate, and exercising control is extremely rewarding and beneficial for an individual's wellbeing (Leotti, Iyengar, and Ochsner 2010). People's ability to gain and maintain a sense of control is essential for their evolutionary survival (Shapiro, Schwartz, and Astin 1996). The concept of sense of control in our work was introduced in the introduction. Fig. 3 shows the dynamics of balancing between exploration and exploitation that is tightly related to frustration and therefore inversely to the sense of control. The level of *frustration* increases when a selected action is no longer rewarding. An optimal level of frustration favours more explorative behavior. Prolonged frustration may result in two different states. When a new action that is rewarding has been found it leads to a state of *elation*, that is characterized by a sudden decrease of frustration. On the other hand, when a new action has not been encountered, prolonged frustration will lead to a state of *learned helplessness*.

Frustration and exploration

The timing of infants' transition to upright locomotion was associated with temperament (Scher 1996). More specifically, earlier walkers become more easily frustrated and stressed when physically constrained. They also reveal more persistence in reaching a blocked goal as compared to later walkers during the transition to walking (Biringen et al. 2008). We suggest that being easily frustrated could be caused by the perception of limits of self-efficacy. As suggested by Zelazo (Zelazo 1983) 12-month-old infants are more skilled in making associations, and that may stimulate their interest in distant objects. The failures in obtaining these new challenging goals may significantly decrease infants' sense of control, increasing at the same time their level of frustration. In our opinion, growing emotional distress associated with a decreasing level of control in pre-walking infants can trigger the process of exploration. Fustration-motivated exploration, as proposed by Wong, may play the function of widening the scope of an agent's response repertoire (Wong 1979). Although our observational data do not allow us to perform an exact analysis of variability of infants' reaching trajectory, we observed that pre-walkers slightly more than other groups of infants use their left, right, or both hands. We speculate that infants before the transition to walking may vary their reaching behavior more.

In classical reinforcement learning, one possibility for the agent to choose an action is a *softmax* action selection rule (Sutton and Barto 1998):

$$P_t(a) = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^{n} e^{Q_t(b)/\tau}}; \qquad (1)$$

where $P_t(a)$ is a probability of selecting an action a, $Q_t(a)$ is a value function for an action a, and τ is a positive parameter called the temperature that controls the stochasticity of a decision. A high value of the temperature will cause the actions to be almost equiprobable and a low value will cause a greater difference in selection probability for actions that differ in their value estimates. The parameter τ is usually fixed. For an adaptive agent, however, this parameter should be flexible in order to dynamically regulate the balance between exploration and exploitation. We suggest that a level of frustration, that reflects the agent's sense of control, could be used in the *softmax* choice rule instead of the parameter τ. It has been shown that frustration leads to higher levels of noradrenaline in the right amygdala (Young and Williams 2010). Thus, our suggestion seems to be consistent with Doya's proposal (Doya 2002) that noradrenaline may control the temperature τ.

Elation and fine-tuning

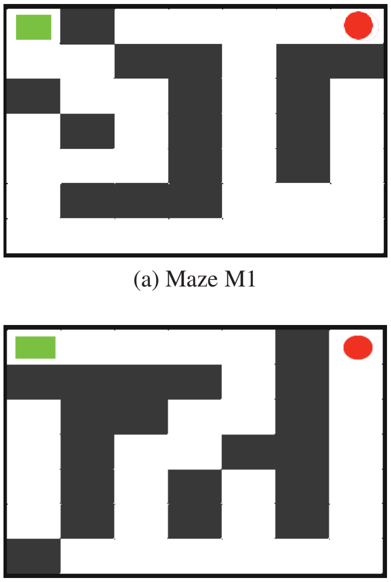
The newly walking infants are described as "euphoric" in relation to the first steps away from their mother (Biringen et al. 1995). The experience of overcoming a prolonged state of frustration that was caused by an inability to reach for a desired distant object results in an extremely high level of sense of control. We call such a state *elation*, and relate it to a sudden decrease of frustration. As our experimental data suggest such a state may contribute to a decreased ability of learning from a negative feedback. In our opinion, the temporary omission of errors plays an important role in fine-tuning a newly acquired behavior.

As the result of our experiment suggested, low learning rate may be helpful in fine-tuning the newly learned behavior. In temporal difference reinforcement learning a value function V_t is updated after each choice has been made, according to the following formula:

$$V_{t+1}(c_t) = V_t(c_t) + \alpha_v * \delta_t; \qquad (2)$$

where V_t is a value function, c_t a set of options, α_v is a free learning rate parameter and δ_t is the difference between the received and expected reward amounts. This formula has been adapted from (Beeler et al. 2010), for a detailed description of temporal difference learning algorithm please refer to (Sutton 1988).

We propose here that the free learning parameter α_v can also be flexible. The state of elation, that is triggered after



(a) Maze M1

(b) Maze M2

Figure 4: The mazes used in our testing scenario. The green square indicates the start and the red circle the destination.

the agent makes a discovery following a period of prolonged frustration, should temporarily decrease the learning rate.

Learned helplessness

Consistend and repetitive lack of reward will lead to an extremely high level of frustration, which if not overcome should eventually result in a feeling of *helplessness* and *depression*. Although such a state is not desirable for the agent's benefit, it might serve as an good indicator for selecting a less challenging goal. In order to re-gain its sense of control, the agent should lower its expectations and attempt to practice less demanding tasks.

Simulation and results

As discussed in the previous section, a sense of control may play an important role as a possible mechanism for regulating the intrinsic motivation for learning. Two different simulation scenarios were used to test whether the frustration dynamics could lead an agent to more adaptive behavior.

For the purpose of our simulations, frustration was represented as a simple leaky integrator. We chose the leaky integrator model because it captures the dynamics of a rapid rise in frustration level and also the possible rapid decrease over time if no input is provided:

$$df/dt = -L * f + A_o \tag{3}$$

where f is the current level of frustration, A_o is the outcome of the action and L is the fixed rate of the 'leak' ($L = 1$ in our simulations).

Frustration and exploration

We used a non-stationary environment to test whether frustration can actually lead to more adaptable behavior. The experimental setup was similar to the one used in (Zhang and Pan 2006). The simulation started in a form of 7-by-7 Maze M1 (Fig. 4a), and after 80 learning epochs, the maze changed to M2 (Fig.4b). We compared the performance of a general reinforcement learning agent with an agent equipped with the frustration module. The learning algorithms were based on the standard Q-learning algorithm. The Q values were updated as follows:
$Q(s_t, a_t) = 0 + \gamma * V(s_{t+1})$, for successful state transition, $Q(s_t, a_t) = -0.5 + \gamma * V(s_{t+1})$, if the agent bumped into the wall, where $V(s_{t+1}) = \max_a Q(s_{t+1}, a)$, and $\gamma = 0.99$. When the agent reached the destination point, a reward of 10 was received. A Boltzmann action selection was used. The Boltzman temperature was initialized to 9, and discounted by 0.9 at each time step.

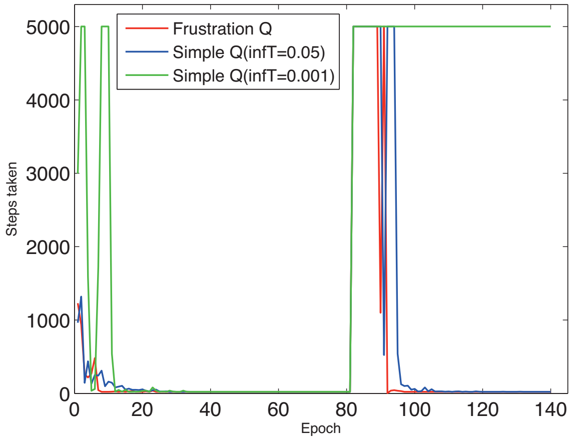


Figure 5: A comparison of the performance for different Q learning agents. The change of maze was at epoch 80.

In case of the frustration agent, the outcome of a state transition (A_o) was fed into its leaky integrator as follows. $A_o = 10$, when the goal has been reached, $A_o = -0.5$ when an agent bumps into the wall, $A_o = 0.1$ in any other condition. Herein, the level of frustration also built up for successful transitions. In this way we could optimize the length of the trajectory from the starting position to the goal position. The current level of frustration was used directly as a Boltzman temperature in the Bolzmann action selection process. The duration of the simulation was 140 epochs. In each epoch, the number of steps taken for the agent to reach the destination was recorded. If the agent failed to reach the destination in 5000 steps, the epoch terminated and the number 5001 was recorded.The results of the simulation are presented in Fig. 5. These two Q agents differed in their lower range of the exploration temperature ($infT$). As it can easily be seen the agent with $infT = 0.001$ could not adapt to the environmental changes. Due to its relative small degree of exploration, after change of the maze the agent could not reach the goal position. Only keeping the $infT$ at the optimal level 0.05 allowed the agent to relearn the correct path to the destination. The agent with frustration also was able to adapt to the new changes in the environmnent, and did it

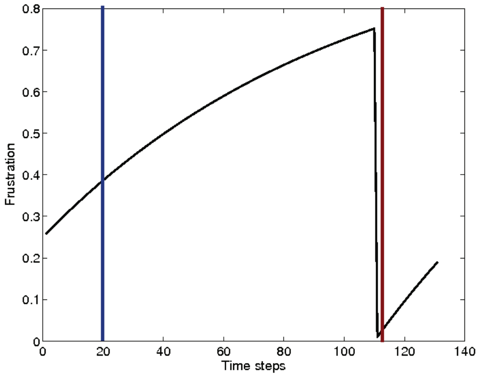


Figure 6: The dynamics of global frustration

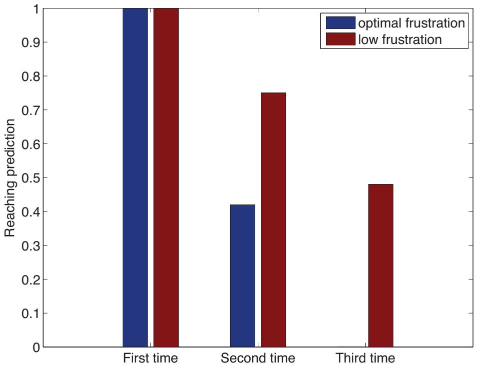


Figure 7: Reaching prediction for the out-of-reach object.

slightly faster than the Q agent.

Elation and fine-tuning

The results from our infant study suggested that the state of elation may be helpful in fine-tuning newly learned behavior. In this simulation we roughly test how sense of control could influence the decision making process. The goal for the simulated humanoid robot, similarly to our infant experiments, was to decide if the object is reachable or not depending on its previous experience. The whole experiment consisted of several sessions. Each session always begun with high probability of reaching for any object. During an experimental session the robot updated its prediction of object reachability based on the reaching outcome and its current level of sense of control. At this point, the level of frustration was artificially reset in order to simulate the transition to walking.

The robot was equipped with a ”know-how” module that was responsible for performing the reaching action (an analytical inverse kinematics solver). After each session, the reaching prediction error was updated according to the following formula:

$$P_e(t+1) = P_e(t) + (P_e(t) - A_o) * E_f; \qquad (4)$$

where P_e stands for the prediction error, A_o actual result of the action, and E_f is a function of frustration defined as follows:

$$E_f = 1 - exp(-\frac{f}{2}); \qquad (5)$$

f is the current level of frustration. This function was selected mainly because it gives low values for low level of frustration, and it rapidly approaches 1 for increased values of frustration. The low values given by this function will result in less error to be taken into account during updating the prediction P_e of the future errors, but will leave the amount of error almost unchanged for optimal and higher values of frustration.

The experiment started with the robot being in an optimal level of sense of control that resembles a state of 12-month-old infants far from the transition period to walking. We assumed here an increased interest of the robot in the distant object, and for that reason each experimental session always begun with high probability of reaching. During the whole experiment the global level of frustration was increasing as a result of non-rewarding reaching behavior. The dynamics of frustration is shown in Fig. 6 where two different stages of sense of control are marked, the first one corresponds to the optimal level of frustration (blue) and the second one to the extremely low level of frustration (red), that comes after overcoming the prolonged frustration. As it easily can be noticed in Fig. 7, the robot being in the optimal level of sense of control, after two unsuccessful trials learned that the object was not reachable and gave up on trying to reach for it. On the other hand, the robot in a state of overconfidence (elation) needed more repetitions to correctly predict the outcome of its action. The behavior of the robot in this stage resembles the behavior of the infant right after the transition to walking, when she persistently continues to reach for unreachable objects. Thus, the state of elation, that is characterized by a low level of frustration, leads to a temporal decrease in learning rate. That as we suggested previously may help infants practice more their newly acquired walking behavior.

Discussion and future work

The shortage of experimental studies concerning intrinsically motivated learning has been pointed out by Kaplan and Oudeyer (Kaplan and Oudeyer 2007). The results obtained from our experiments with infants could be the first step in filling in this gap. Although our infant study suggested a possible mechanism behind infants' learning to walk, we believe that it can be extended to a more general form of intrinsically motivated open-ended learning. The main issue that needs to be addressed first is how walking and reachability are tied tied together in a larger learning framework. The answer to this question could shed light on a more general form of existing relationship between different behaviors and on the process of their acquisition. Another important issue to address is whether a sense of control could serve as a meta-parameter of intrinsic motivation for reinforcement learning.

The preliminary simulation results seem to confirm the viability of our approach. The next step in our research is to perform a series of experiments with a real humanoid robot. The reaching in our simulation was performed by an analytical inverse kinematic controler that excluded the possibility of online learning. As it seems that an optimal level of frustration leads infants to more explorative behavior, we should introduce variations in robot reaching attempts depending on its level of frustration. The robot will have several built-in behaviors, like for example a stepping reflex. As an upright posture is very unstable a constant reaching from this new posture will result on many occasions in a loss of balance. A step made by infants in order to recuperate the balance can trigger the process of learning to walk. We hypothesize that the discovery of a solution that brings the desired goal closer, rapidly decreases the level of frustration, and boosts up the level of control. The state of high level of self-efficacy causes the errors to be omitted until the newly learned skill has been mastered. In this way we believe that our model for the mechanisms that balance exploration and exploitation could trigger learning to walk in a robot.

Conclusion

This paper presented a new approach to an important aspect of intrinsically motivated learning. Sense of control was suggested to be a key factor that allows to dynamically balance exploration and exploitation while learning new skills. The level of frustration also determines how much the negative outcome of an action is taken into account. Omission of the errors while learning was suggested to be helpful in fine-tuning a newly learned behavior. The plausibility of this mechanism was tested using a simulated humanoid robot, and our preliminary results qualitatively replicated the result obtained from our experimental data.

Acknowledgment

This research was supported by: Ministerio de Ciencia y Innovacion (FPU grant AP2007-02565); Grants-in-Aid for Scientific Research (Research Project Number: 22220002); WCU program, Korea National Research Foundation (R31-2008-000-10062-0).

References

Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; and Yoshida, C. 2009. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development* 1(1):12–34.

Baranes, A., and Oudeyer, P.-Y. 2010. Maturationally-constrained competence-based intrinsically motivated learning. In *Proceedings of the Ninth IEEE International Conference on Development and Learning*.

Barto, A.; Singh, S.; and Chentanez, N. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *Proc. 3rd Int. Conf. Development Learn.*

Beeler, J. A.; Daw, N.; Frazier, C. R.; and Zhuang, X. 2010. Tonic dopamine modulates exploitation of reward learning. *Frontiers in behavioral neuroscience* 4:1–14.

Berlyne, D. E. 1960. *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.

Biringen, Z.; Emde, R. N. .; Campos, J. J.; and Appelbaum, M. I. 1995. Affective reorganization in the infant, the mother, and the dyad: the role of upright locomotion and its timing. *Child Development* 66(2):499–514.

Biringen, Z.; Emde, R. N. .; Campos, J. J.; and Appelbaum, M. 2008. Development of autonomy: role of walking onset and its timing. *Perceptual and Motor Skills* 106:395–414.

Doya, K. 2002. Metalearning and neuromodulation. *Neural Networks* 15(4-6):495–506.

Grzyb, B. J.; del Pobil, A. P.; and Smith, L. B. Reaching for the unreachable: (mis)perception of body effectivity in older infants. Manuscript in preparation.

Grzyb, B. J.; del Pobil, A. P.; and Smith, L. B. Reaching for the unreachable: the cause or the consequence of learning to walk. Manuscript in preparation.

Kaplan, F., and Oudeyer, P.-Y. 2007. In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience* 1:225–236.

Leotti, L.; Iyengar, S.; and Ochsner, K. 2010. Born to choose: the origins and value of the need for control. *Trends Cogn Sci.* 14(10):457–463.

Marshall, J.; Blank, D.; and Meeden, L. 2004. An emergent framework for self-motivation in developmental robotics. In *Proc. 3rd Int. Conf. Development Learn.*

Scher, A. 1996. The onset of upright locomotion and night wakings. *Perceptual and Motor Skills* 83:11–22.

Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *Autonomous Mental Development* 2(3):230–247.

Shapiro, D. J.; Schwartz, C.; and Astin, J. 1996. Controlling ourselves, controlling our world. psychology's role in understanding positive and negative consequences of seeking and gaining control. *Am Psychol.* 51(12):1213–30.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Sutton, R. 1988. Learning to predict by the methods of temporal differences. *Machine Learning* 3(1):9–44.

Weng, J. 2002. A theory for mentally developing robots. In *Proc. 2nd Int. Conf. Development Learn.*

Wong, P. T. 1979. Frustration, exploration, and learning. *Canadian Psychological Review* 20(3):133–144.

Young, E. J., and Williams, C. L. 2010. Valence dependent asymmetric release of norepinephrine in the basolateral amygdala. *Behavioral Neuroscience* 124(5):633–644.

Zelazo, P. 1983. The development of walking: new findings and old assumptions. *J Mot Behav.* 15(2):99–137.

Zhang, K., and Pan, W. 2006. The two facets of the exploration-exploitation dilemma. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*.