MILITARY PSYCHOLOGY, 21:(Suppl. 1)S46–S54, 2009

ISSN: 0899-5605 print / 1532-7876 online
DOI: 10.1080/08995600802554631

Exploration and Exploitation in a Foraging Resource Acquisition Task: Implications From Sleep Deprivation

Brian D. Glass, W. Todd Maddox, Arthur B. Markman, and David M. Schnyer
University of Texas, Austin, Texas

Sleep deprivation effects in a fluid, real-time competitive environment are examined using a resource acquisition foraging task. The task is ideal for examining the exploration–exploitation tradeoff in decision making. The exploration–exploitation tradeoff is the balancing of previously successful strategies with the adoption of new strategies. The Generalized Exploration Model is used to develop tasks for which either exploitation or exploration is optimal. Preliminary results from an ongoing sleep deprivation study demonstrate the technique in practice and suggest that sleep deprivation leads to impairment in the exploitation task but not in the exploration task.

Sleep deprivation is a constant battle for many military personnel (Belenky et al., 1994). Therefore, it is important to understand the cognitive aspects of individuals who are required to engage in demanding tasks while sleep deprived. The common view of sleep deprivation is one of an overall slowing of cognitive ability, or an across-the-board deficit of reasoning and decision-making skills. Contrary to that view, studies reveal that some abilities may remain intact (Harrison & Horne, 2000). For example, Williamson, Feyer, Mattick, Friswell, and Finlay-Brown (2000) employed a variety of tasks and found relatively intact performance on tasks that involve visual search and logical reasoning. Still, they reported significant impairment for a range of perceptual, attentional, and memory tasks. However, a meta-analysis by Harrison and Horne (2000) supported intact performance on rule-based tasks and impaired performance on more complex integrating tasks.

Correspondence should be addressed to Brian D. Glass, Department of Psychology, University of Texas, 1 University Station, A80000, Austin, TX 78712. E-mail: glass@mail.utexas.edu.

Other studies indicate increased risk-seeking and decreased cognitive flexibility under sleep deprivation (Herscovitch, Stuss, & Broughton, 1980; McKenna, Dickinson, Orff, & Drummond, 2007). Clearly, the role of sleep deprivation on reasoning and decision-making is not well understood. Many of the tasks used in the study of sleep deprivation are monotonous and without any clear ecological validity.

In an effort to address the role of sleep deprivation on decision-making in a fluid, dynamic, and competitive situation, a model of resource acquisition is used to develop ecologically valid tasks. These tasks involve resource acquisition in a real-time two-dimensional environment in which the human participant competes against simulated opponent agents in a virtual resource acquisition foraging task. The environment was inspired by experiments and observations in behavioral ecology as well as observations and models of human group behavior in similar situations (Critchfield & Atteberry, 2003; Kennedy & Gray, 1993; Kraft & Baum, 2001; Madden, Peden, & Yamaguchi, 2002; Roberts & Goldstone, 2006; Sokolowski & Tonneau, 2004). The Generalized Exploration Model (GEM) was developed out of this work to pinpoint the optimal behavior of individual foragers as a function of the task parameters (Glass, Markman, & Maddox, in press). Task parameters range from the number of opponent agents, the placement of resource patches throughout the environment, and the relative replenishment rates of the available patches. By observing the optimal behavioral characteristics of simulated agents, tasks are constructed for which the optimal behavior is well understood. Of special interest is the behavioral aspects of the exploration–exploitation tradeoff in cognitive decision-making.

The exploration–exploitation tradeoff is a decision-making paradigm well suited for empirical and modeling analysis with GEM as well as one that has neuropsychological implications for those experiencing sleep deprivation. *Exploration* is the willingness to try new strategies, and *exploitation* is the reliance on previous knowledge and strategies. Since one must abandon old strategies to adopt new ones, exploration and exploitation therefore trade off (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Worthy, Maddox, & Markman, 2007). The locus coeruleus-norepinephrine (LC-NE) system may drive this tradeoff process via connections with frontal brain areas responsible for monitoring and task utility and strategy shifting. During periods of extended sleeplessness and fatigue, the system may be prone to lead to more exploratory behavior (Aston-Jones & Cohen, 2005). To examine the exploration–exploitation tradeoff in sleep deprivation, two GEM environments are developed for which either exploration or exploitation is the optimal strategy. Next, we discuss the development of these tasks with GEM. Then, we outline the neurological predictions made by the LC-NE system under sleep deprivation. Lastly, we showcase results from an ongoing project that show that sleep deprivation leads to impairment in the exploitative environment despite intact performance in the exploratory environment.

THE GENERALIZED EXPLORATION MODEL

Here we provide an overview of GEM and how it relates to the current project. Glass et al. (in press) provide details of the model and its development from previous models. In the GEM environment, agents are represented as dots on a 120 × 120 grid; these agents try to acquire resources that are also represented as dots (shown in a different color; see Figure 1). Agents move in eight degrees of freedom and can move one grid space per turn. One run involves 1500 turns or time steps. Resources appear in the environment depending on environmental parameters such as the spatial patch arrangement, the relative replenishment rate of the patches, and the global replenishment rate.

A resource patch is a bivariate normal random distribution with a standard deviation of three grid spaces. If an environment has multiple resource patches, then a relative replenishment schedule is specified. For example, the standard two-patch environment features one patch with a center at the coordinate (40, 40) with a relative replenishment rate of 20% and another patch centered at the nearby coordinates (80, 80) with a relative replenishment rate of 80%. In this setup, when a resource dot is dropped into the environment, it has an 80% chance of arriving at one patch and a 20% chance of arriving at the other (Figure 2).

Agents cannot wrap around the grid walls and are allowed to occupy grid spaces with other agents. When an agent lands on top of a resource dot, that agent acquires the resource, and the agent's score changes to reflect the point value of the re-

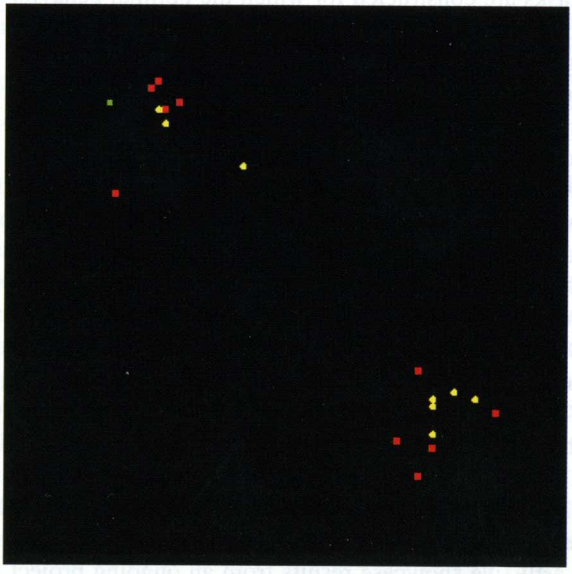

FIGURE 1 Screenshot of a two-patch exploitation environment after 750 time steps (resources are red, opponent agents are yellow, and the lone participant agent is green).

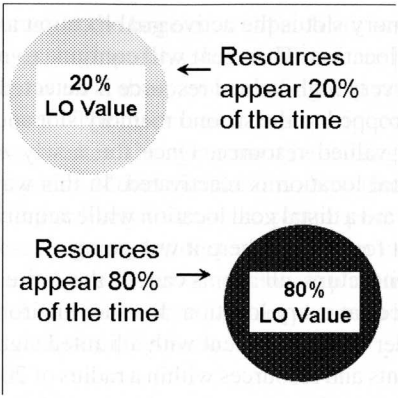

FIGURE 2 Exploitative task environment, with resources replenishing 80% at one patch and 20% at the other patch.

source. When multiple agents land on top of a resource dot, one agent is chosen at random to receive the resource dot.

Agents are controlled by the GEM value equation, based on the EPICURE model of human foraging (Roberts & Goldstone, 2006). In GEM, on each time step an agent selects a goal location to travel toward. The agent selects this goal location by evaluating all the known resources in the environment. Each resource is evaluated with a linear combination of weighted parameters. These parameters are the distance from agent, the density of other resources around the location, the density of other agents around the location, and the distance from the previous goal location (Equation 1).

$$Value(i,j) = \left(P_1 * \frac{1}{distance}\right) + (P_2 * resourcedensity) - (P_3 * agentdensity) + \left(P_4 + \frac{1}{goaldistance}\right) \quad (1)$$

After each potential resource location is given a value, one location is probabilistically chosen using the softmax decision rule (Equation 2). The probability of moving to the resource at coordinate (i, j) is shown in Equation 2.

$$P(\text{moving to i, j}) = \frac{e^{\text{Value(i,j)/K}}}{\sum_x \sum_y e^{\text{Value(x,y)/K}}} \quad (2)$$

The chosen resource location becomes the agent's goal location. The GEM agent has the ability to store the memory of two goal locations in its goal stack

memory. The top memory slot is the active goal location and the second memory slot is the stored goal location. The agent will continue to move toward the active goal location unless a very high-valued resource is detected, in which case the active goal location is dropped to the second memory slot and replaced by the location of the new high-valued resource. Once the newly active goal location is reached, the stored goal location is reactivated. In this way, the GEM agent can continue traveling toward a distal goal location while acquiring proximal resources along the way without forgetting where it was going.

In some environment setups, all agents can see the entire environment and all of the agents and resources at every location. In other environments, a limited sight radius is used. Consider an environment with a limited sight radius, where agents can only see other agents and resources within a radius of 20 grid spaces. In this environment, the agent must rely on memory when considering locations beyond the limited sight radius. GEM agents have this ability, by storing a reinforcement history map of the entire environment. An agent's memory map is updated when it acquires a resource. For potential goal locations outside the limited sight radius, only reinforcement history memory is used to determine value.

Exploration vs. Exploitation Task

The value K in Equation 2 is the exploration parameter that determines how exploratory or exploitative the decision process of the agent will be (Daw et al., 2006; Worthy et al., 2007). As the exploration parameter increases, the agent is more likely to select a resource location that is not necessarily the one with the highest value. As the exploration parameter approaches zero, the agent is more likely to select the resource location with the highest value on each time step. An agent with a high exploration parameter will switch between resource patches more often than an agent with a low exploration parameter. An agent with a low exploration parameter will tend to remain at one resource patch until the patch is exhausted of resources.

By varying the task environment parameters (e.g., spatial patch arrangement and relative replenishment schedule), tasks can be found for which agents with either high or low exploration parameters perform best (i.e., acquire the most resources). In the aforementioned two-patch environment with relative replenishment rates of 80 and 20%, agents with low exploration parameters perform best. The reason for this is that agents with high exploration parameters too often switch between the patches, missing out on resource acquisition while traveling between the patches. This cost of switching is avoided by agents with low exploration parameters that tend to remain at one patch throughout the run. Since resource patch exploitation is optimal in this task, it is deemed the *exploitation task* (Figure 2).

To develop a task for which a high exploration parameter is optimal, a limited sight radius is used in combination with an environment with a large central re-

source patch. The resources in the large central patch are low valued and are only worth one point. However, high-point-value resources are available at small peripheral patches that appear for a short amount of time at random locations along the edges of the environment. The resources at these peripheral locations are worth 100 points. To find these patches, the agent must be willing to abandon the large (yet low-valued) central patch. Model simulations show that agents with high exploration parameters are more likely to leave the central pool and find the high-valued peripheral resources. These exploratory agents stay a larger average distance from the center of the central pool. Agents with low exploration parameters tend to remain at the central pool. Since resource patch exploration is optimal in this task, it is deemed the *exploration task* (Figure 3).

SLEEP DEPRIVATION AND THE EXPLOITATION–EXPLORATION TRADEOFF

There is evidence that the LC-NE system drives the exploration–exploitation tradeoff due to connections with frontal areas that monitor task related utility, and this system is modulated during sleep deprivation. An NE system with widespread cortical and subcortical innervation begins at the LC, which is in turn innervated by

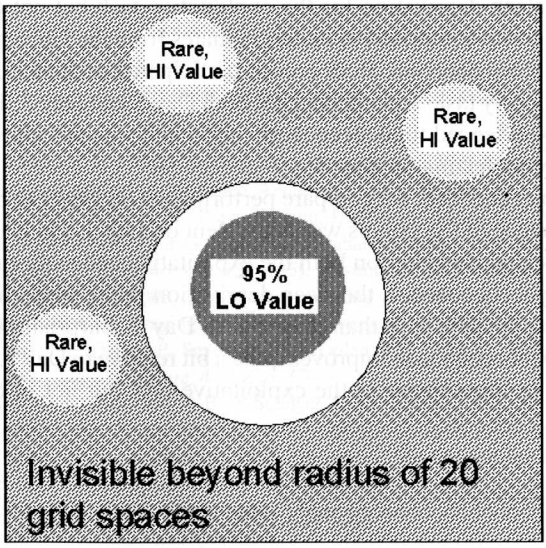

FIGURE 3 Exploratory task environment, with a limited sight radius, a large low-value central pool and small high-value pools that transiently appear at the periphery.

orbitofrontal (OFC) and anterior cingulated (ACC) cortices (Rajkowski, Kubiak, & Aston-Jones, 1993).

The ACC and OFC monitor the effectiveness of currently adopted strategies and cause the LC to modulate the NE system when task utility suffers. LC-NE activity ranges from low and high levels of tonic activity to moderate levels of phasic activity. Low and high levels of tonic LC activity lead to inattention or distractibility, respectively. Moderate phasic levels of LC activity result in task engagement and low distractibility. One speculation would be that top-down compensation for fatigue shifts LC into high levels of tonic activity, leading to exploration (Aston-Jones & Cohen, 2005).

This increased level of exploration in sleep deprivation is beneficial in tasks for which exploration is optimal, such as the exploration task described above. Conversely, sleep deprivation individuals would then suffer in tasks for which exploitation is optimal, as is the case in the exploitative task environment. Again, this is based on speculation about the LC-NE system under fatigue.

PRELIMINARY EMPIRICAL RESULTS FROM WEST POINT CADETS

In an ongoing investigation of sleep deprivation in undergraduate West Point cadets (Glass, Maddox, Markman, & Schnyer, in press), preliminary results support impairment in sleep deprivation for the exploitative task but not the exploratory task. A control group comprised of undergraduates from the University of Texas at Austin was not sleep deprived. Both groups participated in three exploitative tasks and three exploratory tasks, interleaved, on both Day 1 and Day 2. The sleep deprivation group experienced total sleep deprivation and was not allowed to sleep in the 24 hours before testing on Day 2.

For each group and task, we compare performance on Day 1 to performance on Day 2. Performance of the groups was equivalent on Day 1. On Day 2, the control group performs slightly better on both the exploitative and exploratory tasks than they did on Day 1. In contrast, the sleep deprivation group performs significantly worse on the exploitative task than they did on Day 1. Surprisingly, their performance on the exploratory task improves quite a bit relative to Day 1. Thus sleep deprivation leads to impairment in the exploitative task but not in the exploratory task.

CONCLUSIONS

A neuropsychological account of the exploration–exploitation tradeoff in sleep deprivation, based on the LC-NE system, hypothesizes that fatigue leads to increased

exploration. Using GEM, a model of resource acquisition in a spatially explicit real-time environment, tasks were constructed for which exploitation or exploration was optimal. Preliminary results from an ongoing study suggest that sleep deprivation individuals are not impaired in tasks for which exploration is the optimal strategy, despite difficulties in tasks for which exploitation is optimal.

This pattern of results supports previous findings that more complex, reasoning skills are maintained in sleep deprivation (Williamson et al., 2000). In the current study, participants completed tasks that fundamentally differ from tasks in previous studies in their ecological validity and real-time nature. Further investigation should attempt to use modeling simulations to further delineate components of the current task and reveal new tasks that lead to a greater understanding of the complex nature of reasoning and decision making under sleep deprivation.

These findings do have implications for the military. Soldiers are trained with the aim of allowing them to perform reliably under adverse conditions including high stress and total sleep deprivation. Some tasks can be taught to the point where they are clearly driven by the habit learning system. For example, sharpshooters train until their performance is automatic. In contrast, dynamic conditions may not support the kinds of repeated trials that create habits. In these cases, Soldiers are given heuristics to enable them to perform under pressure. The present results suggest that in conditions in which Soldiers must follow rules, their performance will suffer under sleep deprivation. This behavior may cause them to wander through environments less carefully than may be warranted in dangerous situations.

REFERENCES

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450.

Belenky, G., Penetar, D. M., Thorne, D., Popp, K., Leu, J., Thomas, M., et al. (1994). The effects of sleep deprivation on performance during continuous combat operations. In B. M. Marriot (Ed.), *Food components to enhance performance* (pp. 127–135). Washington, DC: National Academy Press.

Critchfield, T. S., & Atteberry, T. (2003). Temporal discounting predicts individual competitive success in a human analogue of group foraging. *Behavioural Processes, 64*, 315–331.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*, 876–879.

Glass, B. D., Maddox, W. T., Markman, A. B., & Schnyer, D. M. (in press). *The effects of sleep deprivation on the exploration–exploitation tradeoff.* Manuscript submitted for publication.

Glass, B. D., Markman, A. B., & Maddox, W. T. (in press). The Generalized Exploration Model (GEM). A model of human foraging for empirical analysis.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied, 6*, 236–249.

Herscovitch, J., Stuss, D., & Broughton, R. (1980). Changes in cognitive processing following short-term cumulative partial sleep deprivation and recovery oversleeping. *Journal of Clinical Neuropsychology, 2*, 301–319.

Kennedy, M., & Gray, R. D. (1993). Can ecological theory predict the distribution of foraging animals? A critical analysis of experiments on the ideal free distribution. *Oikos, 68*, 158–166.

Kraft, J. R., & Baum, W. M. (2001). Group choice: The ideal free distribution of human social behavior. *Journal of the Experimental Analysis of Behavior, 76*, 21–42.

Madden, G. J., Peden, B. F., & Yamaguchi, T. (2002). Human group choice: Discrete-trial and free-operant tests of the ideal free distribution. *Journal of the Experimental Analysis of Behavior, 78*, 1–15.

McKenna, B. S., Dickinson, D. L., Orff, H. J., & Drummond, S. P. (2007). The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions. *Journal of Sleep Research, 16*, 245–252.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Social Neuroscience Abstracts, 19*, 974.

Roberts, M. E., & Goldstone, R. L. (2006). EPICURE: Spatial and knowledge limitations in group foraging. *Adaptive Behavior, 14*, 291–313.

Sokolowski, M. B. C., & Tonneau, F. (2004). Human-group behavior: The ideal free distribution in a three-patch situation. *Behavioural Processes, 65*, 269–272.

Williamson, A. M., Feyer A., Mattick, R. P., Friswell, R., & Finlay-Brown, S. (2000). Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis and Prevention, 33*, 313–326.

Worthy, D. A., Maddox, W. T., & Markman, A. B. (2007). Regulatory fit effects in a choice task. *Psychonomic Bulletin and Review, 14*, 1125–1132.