Curiosity as Pattern Matching: Simulating the Effects of Intrinsic Rewards on the Levels of Processing

Curiosity as Pattern Matching: Simulating the Effects of Intrinsic Rewards on the Levels of Processing

Kazuma Nagashima (nagashima.kazuma.16@shizuoka.ac.jp),
Junya Morita (j-morita@inf.shizuoka.ac.jp) ,
Yugo Takeuchi (takeuchi@inf.shizuoka.ac.jp)
Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University,
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka-ken, 432-8011 Japan

Abstract

Many studies have been conducted concerning curiosity, a type of intrinsic motivation in humans and artificial agents. However, the specifics of the correspondence between curiosity in humans and artificial agents have not yet been fully explained. This study examines this correspondence on the Adaptive Control of Thought–Rational (ACT-R) cognitive architecture by exploring situations in which curiosity effectively promotes learning. We prepared three models of path planning, representing different levels of thinking, and had them learn in multiple-breadth maze environments while manipulating the curiosity strength. The results showed that curiosity in learning an environment negatively affected the model with a shallow level of thinking. However, it positively affected the model with a deliberative level of thinking. We consider that the results show some commonalities with human learning.

Keywords: cognitive modeling; intrinsic motivation; curiosity; ACT-R

Introduction

Although curiosity is assumed to be an effective source of motivation that encourages humans to engage in long-term learning, it does not always work effectively. To explore the conditions in which intrinsic motivation works well, we examine the influence of the levels of the thinking process that many cognitive scientists have discussed (e.g., Brooks, 1986; Kahneman, 2011). This study does not go into the details and differences of such theories but assumes broad distinctions between the shallow automatic level in which a person does not think carefully (fast process) and the deep deliberative level in which a person takes time to think carefully (slow process).

This study aims to clarify the role of intrinsic motivation in such levels of thinking. To accomplish this, we prepared models that instantiate the information processing of each thinking level. The prepared models were constructed based on a cognitive architecture, which is a structure enabling cognitive functions in various tasks by various individuals (Anderson, 2007). By assuming a common structure, differences in the thinking levels are represented as combinations of primitive processes provided by the architecture.

Of the several cognitive architectures developed to date, this study uses Adaptive Control of Thought–Rational (ACT-R; Anderson, 2007) because this architecture has the most publications showing the details of the models for various tasks (see Kotseruba & Tsotsos (2018) for a quantitative review). By referring to these models, we can implement several thinking levels with the validation made by the previous studies. Furthermore, ACT-R has two types of knowledge (declarative and procedural), which seem useful to represent different levels of thinking.

As a representation of curiosity in ACT-R, this study uses a mechanism proposed in our previous study (Nagashima, Morita, & Takeuchi, 2020). Although there are other options for motivation theory in ACT-R (e.g., Juvina, Larue, & Hough, 2018), our previous proposal has the advantage of implementing curiosity as rewards accompanied by pattern matching. We consider this characteristic effective to examine the complex relations between curiosity and levels of thinking. However, our previous study failed to demonstrate that the mechanism relates to human learning. Therefore, the current study newly implements models of different processing levels and tries to find common features with human learning by examining the relation between those levels and the mechanism of intrinsic motivation.

In order to clearly present the goal of this study, the following section shows previous studies concerning intrinsic motivation and ACT-R. Following this, we briefly introduce a curiosity mechanism proposed by Nagashima et al. (2020). We then discuss this mechanism's implementation and run simulations of a specific task. Finally, we summarize the implications of the study and indicate future directions.

Related Works

As noted above, curiosity is regarded as a type of intrinsic motivation. Therefore, this section introduces studies concerning intrinsic motivation to explore situations in which curiosity works effectively. Following this, a brief introduction of ACT-R is presented, focusing on the relationship with levels of thinking.

Intrinsic Motivation in Humans and Artificial Intelligence

To date, a large body of studies has been created concerning learning as facilitated by intrinsic human motivation. For example, Malone (1981) categorized intrinsic motivation into three types: "challenge," which comes from goals of appropriate difficulty; "fantasy," which leads to the imagination of unrealistic experiences; and "curiosity," which is stimulated by something surprising, interesting, or fun. These types are not independent but interrelated. Therefore, reviewing the categories other than curiosity can also help to place the study in a broader context.

Malone's classification of motivation as challenge has been related to the discussion of the optimal level of intrinsic motivation (Csikszentmihalyi, 1990; Yerkes & Dodson, 1908). In humans, there are appropriate levels of task difficulty at which intrinsic motivation is stimulated. Based on this idea, Baranes, Oudeyer, and Gottlieb (2014) found through experiments that intrinsic motivation that is neither too high nor too low for a task is effective. Furthermore, the appropriate level of difficulty for an individual depends on the individual's preferred level of thinking. Based on this discussion, we assumed the dependency of the appropriateness of the challenge on an individual's level of thinking. In other words, intrinsic motivation can be enhanced by providing tasks that are suitable for the level of thinking that the individual prefers.

We consider that the above discussion of challenge cannot be separated from a discussion of curiosity. Rather, we treat curiosity as a mechanism of intrinsic motivation evoked by the appropriate difficulty of a task. Various studies of artificial agents have addressed the mechanisms of curiosity. The key principle of modeling curiosity can be obtained from the theory of prediction error (Friston, 2010). The emotions of surprise, interest, and enjoyment that trigger curiosity are caused by discrepancies between perceptions of the external world and predictions derived from experience. Based on this theory, autonomous agents have been constructed to learn an environment based on curiosity (Aubret, Matignon, & Hassas, 2019; Schmidhuber, 2010; Singh, Barto, & Chentanez, 2005). In contrast to conventional reinforcement learning, in which one receives a reward directly from the external environment (Sutton & Barto, 1998), the rewards generated from intrinsic motivation fluctuate depending on the state of the internal environment.

In recent years, this topic has progressed remarkably with a framework for deep reinforcement learning through an end-to-end approach (Burda et al., 2018; Mnih et al., 2015; Pathak, Agrawal, Efros, & Darrell, 2017). In particular, Burda et al. (2018) have shown that agents with curiosity can learn a wide range of environments and improve their game scores without explicit extrinsic rewards.

Levels of Thinking in ACT-R

The studies presented in the previous section implemented curiosity-based agents using a reinforcement-learning framework. However, with the framework alone, it is difficult to explore situations in which intrinsic motivation functions effectively. Thus, a framework that seamlessly connects the learning algorithms and the process of inference in a task is needed.

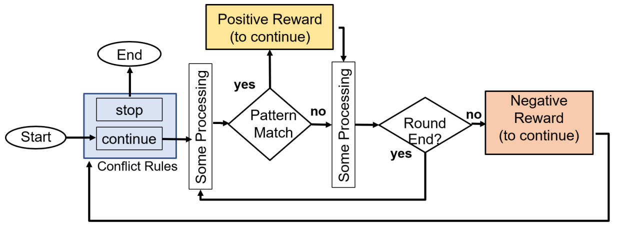

Figure 1: Flowchart of the task continuation framework presented in (Nagashima et al., 2020). A positive reward is generated by "pattern matching" accompanied by memory retrieval.

As noted previously, we use ACT-R as such a framework to connect multiple levels of thinking and curiosity-based learning. ACT-R has modules corresponding to brain regions. For example, the declarative module (prefrontal cortex) retains experience and knowledge, and the goal module (anterior cingulate cortex) manages states in tasks. The production rules stored in the production module (basal ganglia) are selected based on the status of such modules, and they send commands to the modules as actions (e.g., they search for knowledge that meets the conditions and update the current state of the task). These rules include variables that realize flexible correspondence (pattern matching) with module states. According to the ACT-R theory, pattern matching is realized in the cerebral cortex, specifically the prefrontal cortex. If knowledge retrieval from the declarative module becomes unnecessary in the task, the role of the basal ganglia becomes dominant in the process of proceduralizing the task. Therefore, we consider pattern matching accompanied by memory retrieval the criterion that distinguishes a deliberative level of thinking from a shallow level of thinking in ACT-R. In other words, the shallow level of thinking involves little pattern matching, while the deliberative level of thinking involves extensive pattern matching.

Mechanism of Curiosity in ACT-R

This section presents the mechanism of intrinsic motivation proposed by Nagashima et al. (2020). According to their basic idea, curiosity, especially that involving a higher cognitive function, is connected with pattern matching. Several authors have stated that enjoyment, a source of curiosity (Friston, 2010), is related to discovering novel patterns in the environment (Caillois, 1958; Csikszentmihalyi, 1990; Huizinga, 1939; Koster, 2004; Schmidhuber, 2010). Following such discussion, Nagashima et al. (2020) focused on a mechanism of pattern matching by computers as a concept that corresponds to pattern discovery by humans.

Moreover, based on the correspondence between human curiosity and pattern matching, Nagashima et al. (2020) proposed a framework for task continuation in a general environment (Figure 1). This framework assumes a task that consists of several rounds. At the start of each round, the model decides whether to continue or stop the task by selecting production rules corresponding to each option. After it decides to continue the task, the model proceeds with the round. When the model encounters a condition that ends the round, a new round begins.

The selection of production rules is controlled by utility learning in ACT-R (Wai-Tat & Anderson, 2006). In the above process, the initial utility value of the *continue rule* is considered higher than that of the *stop rule*. The process of becoming bored from this initial state can be modeled by assigning a trigger of a negative reward to the rule that recognizes the end of each round.

To prevent boredom and consider the conditions that result in positive rewards and continued learning, curiosity is required. In this mechanism, rules that trigger positive rewards are defined as rules that fire as a result of the successful retrieval of declarative knowledge in the task. The search for declarative knowledge requires pattern matching between the conditional clauses of the rule (the current situation) and the memory in declarative knowledge. However, this rule gradually becomes used for repeated executions; that is, "production compilation" in ACT-R integrates the two rules and generates a compressed hierarchical rule. After integration occurs, it becomes routine and cannot be related to a reward. Then, the utility value of the *continue rule* decreases, and the *stop rule* fires.

Therefore, the framework represents the decrease in curiosity that comes from the discrepancy between the model's predictions (routine compiled knowledge) and the results of the action. In short, long-term task continuation is achieved by keeping the model engaged in pattern matching between the conditional clauses of production rules and declarative knowledge. Thus, the mechanism is consistent with the key principle of curiosity (Friston, 2010), while it utilizes the distinction between declarative and procedural knowledge in ACT-R.

Simulation

To examine the conditions in which curiosity functions effectively, we conducted a simulation study using the mechanism presented above. In this section, we first clarify the purpose of the simulation. Following this, the actual manipulations of the simulation are defined, and the results are presented.

Aims and Indicators

The purpose of the simulation was to address the following two successive questions:

1. What kinds of factors stimulate curiosity?
2. How does stimulated curiosity affect task learning?

To address the first question, this simulation manipulated the learning factors from both the internal and external viewpoints. The external factor can be considered the breadth of the learning environment (difficulty of the task), while the internal factor corresponds to the cognitive strategies (levels of thinking) implemented in which the model can be used. The influence of these internal/external factors on curiosity is measured as (a) the number of continuations of a task (number of firings of the *continue rule* in Figure 1).

The second question is explored because task continuation does not always contribute to task learning. To assess the effects of intrinsic motivation on task learning, we examined (b) the goal achievement rate, (c) the behavior pattern of the environment search, and (d) the number of newly generated rules. The index (b) is the outcome of task learning, and the index (d) indicates the internal changes in the model caused by task continuation. Regarding the connection between outcomes and internal changes, the present study computed the behavioral index (c) as the information entropy of the environment search:

$$Hr = \frac{-\sum_{i \in \mathrm{n}} p(x_i) \log p(x_i)}{\log n} \tag{1}$$

where x_i and n indicate each location and the number of locations in the map, respectively. This index increases if the model explores the environment extensively but decreases if the model insists on the same behavioral pattern during the task.

Simulation Conditions

We used maze searching as a task and manipulated the external factor by changing the size of the map; the internal factor was searching strategies as the model's level of thinking. Figure 2 shows the overview of the manipulations.

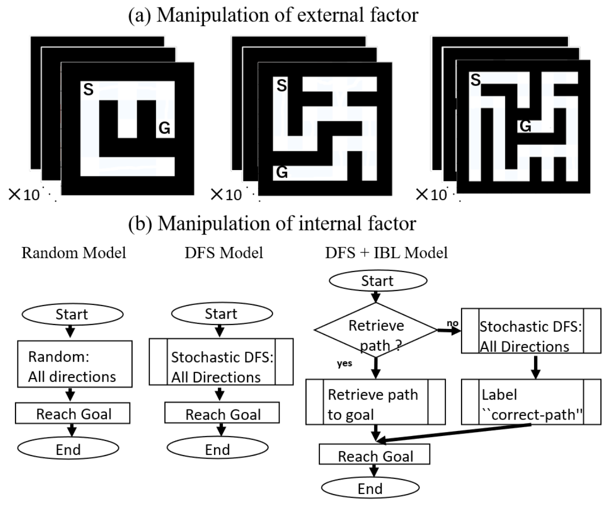

Figure 2: Manipulations of external and internal factors.

Manipulation of the External Factor Figure 2a depicts several maze maps, which were automatically created in a grid world using a maze generation algorithm that randomly changed wall positions with the constraint of avoiding a loop structure. As a parameter for the algorithm, we varied the map sizes between 5×5, 7×7 and 9×9. Ten maps were prepared for each size. These maps were given a start and a goal by choosing two positions having the largest number of intersections to traverse. The map was included in the model as location information that uses declarative knowledge to construct a topological map (Reitter & Lebiere, 2010). The model's current location is maintained as a slot content of the goal buffer. In each round, the model was first located at the start position and moved to the next location by retrieving the path from the declarative module. When the current location matched the goal, a new round began. The model repeated this for each map until it became bored (fired the *stop rule* in Figure 1).

Manipulation of the Internal Factor To manipulate the internal factor, we prepared three models corresponding to the different levels of thinking in the above environment. Figure 2b illustrates the abstract flow of the models, presenting the shallowest model (the left model), the most deliberative model (the right model), and the intermediate model (the middle model). Because these models were developed based on Nagashima et al. (2020), please refer to the original literature for the details of the base model. Brief descriptions of each model are as follows.

1. *Random model (random)*
   At each point in the movement, the model stochastically fires a rule representing the next direction (east, west, south, or north). Based on the direction and current position, the model searches for a path from declarative knowledge. If the model succeeds in finding a path, it moves to the location according to the direction searched (changing the state of the goal buffer). If the model fails to find a path, it repeats the same procedure. As the rounds proceed, the model compiles such retrieved declarative knowledge into procedural rules.
2. *Stochastic DFS model (DFS)*
   This model uses a stochastic depth-first search (DFS), as presented in Reitter and Lebiere (2010). This strategy determines a path by backtracking with the stack structure implemented by ACT-R's declarative and imaginal modules. As with the random model, this model first stochastically determines the direction of movement. After successfully retrieving a path linking the current location to the directed location, the model creates a new chunk linking the two locations as "already searched" and stores it in the declarative module. The model repeats this process until it reaches a goal or fails to retrieve a path. When the model fails to retrieve a path (reaches a dead end), it reverts to the previous location using a memorized chunk (already searched). Like the random model, the stochastic DFS model learns new rules by compiling declarative knowledge on paths, but it can repeat more rounds because it has internal resources that allow it to reach a goal effectively.
3. *Stochastic DFS and IBL model (DFS+IBL)*
   This model is the same as that presented in Nagashima et al. (2020). The model performs a combination of the probabilistic DFS (Reitter & Lebiere, 2010) and instance-based learning (IBL: Gonzalez, Lerch, & Lebiere, 2003).[1] At the beginning of the task, the model uses the stochastic DFS to explore the maze. Each time the model reaches a goal, it labels all the chunks used in the current round as the "correct path." In the next rounds, if the model can retrieve the knowledge, it uses it. If it cannot retrieve it, the model uses a probabilistic DFS to reach the goal from the current position. Among the three models, this model has the most deliberative and costly strategy. It always tries to memorize chunks and retrieve correct paths from its memory. As the round proceeds, however, the model accumulates the "correct path" and eventually compiles it into procedural knowledge, which leads to the most effective goal achievement behavior.

Parameters The simulation used the default ACT-R 7.14. The initial utility value of the *continue rule* was set to 10, and the initial utility value of the *stop rule* was set to 5.[2] We also assigned negative reward triggers ($r = 0$), which were lower than the initial utility value of the *stop rule*, to rules that recognized the end of the round (reaching a goal or recognizing that the time limit of each round had passed) and assigned positive reward triggers to rules that included pattern matching as curiosity. In this study, we selected the *path finding rule* accompanied by the pattern matching to the declarative memory as a positive reward trigger and varied the value from 1 to 20 as the simulation conditions (strength of curiosity). For each condition of the positive reward value, the model ran the task 1000 times at a maximum of 80 rounds each time. In addition, we set the time limit for each round to 100 s in ACT-R simulation time. When the time limit was reached, the model was forced to move to the next round.

Results

Figure 3 displays the results of the simulation as a function of the reward values for path finding. Each point in the graphs indicates the average scores of the four indices ($n = 10000$). The effects of the external factor (the map sizes) are shown in the difference of the three lines in each graph, and the influence of the internal factor can be seen by comparing the three graphs vertically aligned in the figure. The horizontal alignment of the graphs corresponds to the four indices presented at the beginning of this section, and the rightmost figures are correlation matrices of the indices and the two dependent variables (rewards and map size). In the following section, we examine the details of the results according to the two questions presented as the aims of this simulation.

Factors Stimulating Curiosity The left three graphs in Figure 3-a indicate the number of continued rounds. The strong effects of the internal factor on this index are clearly seen. In the upper two models (DFS+IBL and DFS), greater intrinsic rewards increased the number of task continuations (DFS+IBL: $r = .94$; DFS: $r = .97$). In contrast, the random model indicated a weaker correlation between the rewards and the number of rounds ($r = .67$), exhibiting an inverted U shape. Greater intrinsic rewards promoted task continuation until approximately 14 and then decreased task continuation. This inverted U shape suggests the existence of an

[1] Although the original IBL used a blending mechanism, the current model does not use the mechanism; it only utilizes the learning in declarative memory.

[2] Following Anderson et al. (2004), noise parameters were set as follows: ans (activation noise level) = 0.4 and egs (production noise level) = 0.5.

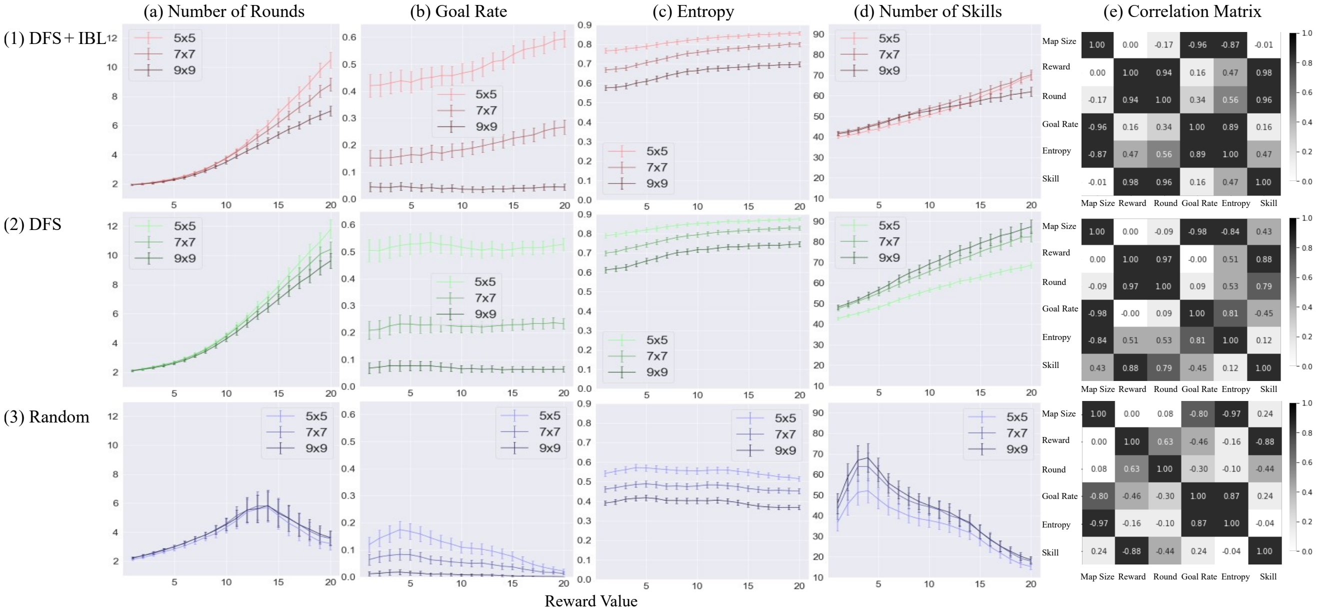

Figure 3: Results of simulation. The panels distinguish models vertically and indicators horizontally. The error bars in each graph represent the standard deviation scaled by 10.[3] The vertical axes of the graphs show the titles of the graphs. a: number of rounds continued, b: goal rates, c: entropy, and d: number of new rules generated by the production compilation. The rightmost graphs show e: correlation matrices between each variable.

optimal level, which is frequently pointed out in theories of intrinsic motivation. For example, according to Yerkes and Dodson (1908)'s classical theory, as a learner's arousal level increases, performance increases to a certain point and decreases beyond that point.

The effects of manipulating the external factor also varied between the three models (DFS+IBL: $r = -.17$, DFS: $r = -.09$, random: $r = .08$). In the upper two models, we observed the negative effects of the size of the external factor, especially in high-reward conditions. Greater intrinsic rewards were more effective in smaller maps, where the model could more easily reach a goal. In other words, it suggested that applying DFS to a wide range of environments was not effective in obtaining internal rewards. The DFS strategy needed to backtrack to find a path, but the time cost of backtracking increased with the size of the map, which can be interpreted as decreasing the chances of finding a path within a limited time period.

Effects of Task Continuation on Learning The remaining of the indices in Figure 3 indicate how the stimulated intrinsic motivation affected task learning. As can be seen in the upper two models, larger rewards increased the entropy (IBL+DFS: $r = .47$, DFS: $r = .51$) and the number of learned rules (IBL+DFS: $r = .98$, DFS: $r = .88$), indicating that the deliberative strategy made use of intrinsic rewards to expand the search of the environments. Furthermore, the learning outcome results reveal the effects of IBL. The model with IBL (the upper model) showed the positive effects of intrinsic rewards ($r = .16$), especially in the smaller maps (5×5: $r = .97$, 7×7: $r = .97$, 9×9: $r = -.17$). However, the intrinsic rewards in the model without IBL had no effect on goal achievement ($r = -.00$).

Note that the IBL itself did not always work effectively in terms of goal rates. In the smaller-reward conditions, the IBL model had lower goal rates than the DFS-only model. However, when greater intrinsic rewards were given, the performance of the model with IBL exceeded that of the model without IBL. As described previously, IBL is a costly and slow strategy that always tries to retrieve a correct path. Therefore, it takes time to make use of such experiences to improve performance. However, it can be assumed that it is difficult to learn to reach a goal without labels of the correct paths. As the flat pattern of the goal rates in the DFS model without IBL (Figure 3-2-b) indicates, the lack of explicit correct labels led to disoriented wandering behaviors in the environment.

Unlike the other two models, the random model with a shallow strategy had a different overall trend. Like the results of the number of rounds, the number of skills (correlation with the reward: $r = -.88$) and goal rates (correlation with the reward: $r = -.46$) exhibited inverted U-shaped trends. Furthermore, in Figure 3-3-a, the peaks of the inverted U shapes in these two indices are smaller than that in the number of rounds, reflecting negative correlations of the two indices with the intrinsic rewards. More critically,

[3] The standard deviation rather than the standard error, which varies with n, is used to indicate the variability in the data.

in Figure 3-3-c, intrinsic reward has a negative effect on entropy ($r = -.16$). These results suggest that higher intrinsic rewards triggered by path finding strengthen irrational low-level behaviors (repetitive visits of the same locations without expanding the search) rather than leading to the creation of additional rules to achieve the goal. The compiled rule in the random model can be found in the appendix.

Conclusion

The purpose of this study was to examine the conditions in which intrinsic motivation affects learning. To achieve this purpose, we modified the previous model for intrinsic motivation in ACT-R (Nagashima et al., 2020) to represent different levels of thinking. Unlike the conventional methods for reinforcement learning (Aubret et al., 2019; Schmidhuber, 2010; Singh et al., 2005), the ACT-R architecture makes it possible to represent a detailed strategy for different thinking levels with realistic time constraints. We consider those features of ACT-R (different knowledge representations and assumptions of simulating reaction time) useful for representing the distinctions of levels of thinking and examining complex interactions with curiosity.

In the simulation, we manipulated the external and internal factors. As a result, the deliberative models showed the positive effects of intrinsic motivation on task continuation, learning skills, and searching behaviors. Regarding the outcomes of learning, however, only the slowest and most costly model benefited from intrinsic motivation. The model that did not evaluate the correctness of retrieval exhibited disoriented wandering through the environment. Moreover, the model that did not memorize the environment was negatively influenced by intrinsic motivation.

Summarizing these findings, we were able to characterize the effects of curiosity on behaviors in different levels of thinking. There are claims that intrinsic motivation works well with deliberative thinking, which requires “autonomy,” “mastery,” and “purpose,” and that extrinsic motivation works well with shallow thinking, which is usually used in routine work (Pink, 2011). Our model’s behaviors follow this idea, thereby corresponding to the human learning process.

In the future, we will analyze the causal relationship between each variable in detail to disentangle the complexities of the results presented in Figure 3. In addition, we will arrange the task setting to include the process of obtaining initial motivation. In the present study, we assumed that humans start with high motivation for a task. However, in reality, a person’s motivation for a task is likely to vary depending on the difficulty and contents of the task (Malone, 1981), as presented in the DFS and IBL model in the 9×9 condition. We considered that those results indicating the relative ineffectiveness of curiosity in difficult tasks were caused by the time limit. Therefore, we also need to examine the effect of time limits on the relation between the level of thinking and the strength of curiosity. By conducting studies addressing such limitations, we can explore more detailed conditions of task continuation, especially those before a model reaches optimal levels.

Appendix

Listing 1 presents rules relating the movement of locations in the random model (Check-Path and Check-Goal) and a rule that was generated through the production compilation of those rules (Check-Path-And-Check-Goal).

Listing 1: Productions rules in the random model. Strings in brackets indicates variables.

```
Check−Path
If
  The current task status is ‘confirming’
  The current location is <location1>
  The retrieved path has <location2>
Then
  Change the current task status to ‘check−goal’
  Change the current location to <location2>

Check−Goal
If
  The current task state is ‘check−goal’
  The current location is <location>
  The goal is not <location>
Then
  Change the current task status to ‘check−goal’

Check−Path−And−Check−Goal
If
  The current task status is ‘confirming’
  The current location is the <location1>
  The retrieved path has the <location2>
  The goal is not <location2>
Then
  Change the current location to <location2>
  Change the current task status to ‘check−goal’
```

Check-Path moves the current location in the goal buffer to the location described in the retrieved path. Check-Goal confirms that the moved location is not the goal in order to continue searching for the goal location. The compiled rule integrates those rules, having the condition that checks the retrieved destination is not the goal location and the action that leads to non-goal locations. This production is further integrated with rules retrieving the path with specific destinations and becomes a rule conflicting with the rule leading to the goal location. It can be considered that the inverted U shape presented in the random model of Figure 3 occurs as a result of generating such goal-avoiding rules.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe*. Oxford Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036.

Aubret, A., Matignon, L., & Hassas, S. (2019). A survey on intrinsic motivation in reinforcement learning. *Corr*.

Baranes, A. F., Oudeyer, P.-Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in neuroscience*, *8*, 317.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation*, *2*(1), 14–23.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *CoRR*.

Caillois, R. (1958). Les jeux et les hommes.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Huizinga, J. (1939). *Homo ludens versuch einer bestimmung des spielelementest der kultur*.

Juvina, I., Larue, O., & Hough, A. (2018). Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making. *Cognitive Systems Research*, *48*, 4 - 24.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Koster, R. (2004). *Theory of fun for game design*. O'Reilly Media, Inc.

Kotseruba, I., & Tsotsos, J. K. (2018, Jul 28). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … others (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Nagashima, K., Morita, J., & Takeuchi, Y. (2020). Modeling intrinsic motivation in act-r: Focusing on the relation between pattern matching and intellectual curiosity. In *ICCM 2020: 18th International Conference on Cognitive Modeling*.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 16–17).

Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Penguin.

Reitter, D., & Lebiere, C. (2010). A cognitive model of spatial path-planning. *Computational and Mathematical Organization Theory*, *16*(3), 220–245.

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

Singh, S., Barto, A. G., & Chentanez, N. (2005). Intrinsically motivated reinforcement learning. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17* (pp. 1281–1288). MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Wai-Tat, F., & Anderson, J. R. (2006, 6). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology. General*, *135*, 184-206.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.