The dynamics of idealized attention in complex learning environments

Madeline Pelz[1], Steven T. Piantadosi[1], and Celeste Kidd[1,2]
1 Brain and Cognitive Sciences, 2 Center for Vision Sciences
University of Rochester, Rochester, NY
{mpelz, spiantadosi, ckidd}@bcs.rochester.edu,

***Abstract*—Effective allocation of attention is crucial for many cognitive functions, and attentional disorders (e.g., ADHD) negatively impact learning. Despite the importance of the attentional system, the origins of inattentional behavior remain hazy. Here we present a model of an ideal learner that maximizes information gain in an environment containing multiple objects, each containing a set amount of information to be learned. When constraints on the speed of information decay and ease of shifting attention between objects are added to the system, patterns of attentional switching behavior emerge. These predictions can account for results reported from multiple object tracking tasks. Further, they highlight multiple possible causes underlying the atypical behaviors associated with attentional disorders.**

I. Introduction

Individual differences in working-memory capacity can affect people in many aspects of cognitive function, including the allocation of attentional resources. Previous work has linked working memory to inattentional behavior. Children diagnosed with attention deficit hyperactivity disorder (ADHD) exhibit poor working-memory performance in both visuospatial and phonological loop tasks as compared to typically developing children [1].

We aim to create a model of attention that will allow for the exploration of the possible sources of these inattentional behaviors, as well as the possible interplay between these different cognitive systems in those with prototypical attentional control. Multiple object tracking (MOT) tasks are often used to behaviorally assess the connections between visual working memory and attentional allocation. In a typical MOT task, participants are shown a display that contains images of multiple objects with unique identifying features such as color, shape, and spatial orientation. These presentations are followed by memory tasks designed to reveal aspects of the organizational structure in which visual information is being held in short-term memory (for a recent review, see [2]).

In one influential MOT task, participants attended to an array of colored objects and were then tested using a change-detection paradigm. Performance was close to optimal until the number of objects in the array exceeded four, at which point accuracy began to decline, leading researchers to conclude that visual memory encodes each object as a whole [3]. Later studies challenged the idea of this one-to-one correspondence "slot"-based model of memory. Based on a study in which participants were asked to identify a target object in an array of objects that varied in complexity, Alvarez and Cavanagh (2004) concluded that memory capacity is flexible, and depends on the feature complexity of the target objects [4]. Despite this debate, there is a consensus in the literature that visual working memory has a limited capacity, with only a subset of objects able to be retained from a complicated display.

Attentional shifting determines how we move our focus between objects in the environment, as we gain information and hold it in our visual memory. This process has been studied extensively in infants, and it has been suggested that different visual search patterns and speeds yield different learning outcomes [5], [6]. Bronson (1991) used an eye-tracking paradigm to assess individual differences in how 12-week-old infants explored visual shape stimuli, and how this related to their novelty preferences. Infants who exhibited more frequent, shorter fixations encoded new stimuli faster (and, thus, habituated sooner). Infants who exhibited less frequent, longer fixations encoded new stimuli more slowly. Additionally, older infants were more likely to be among those who exhibited frequent, shorter fixations and habituated faster, suggesting an effect of maturation on the allocation of attention. The fact that more frequent, shorter fixations in infants was associated faster stimulus encoding could indicate that increased attentional shifting yields more efficient learning.

Working memory restricts how many items can be held in memory, while attentional regulation affects how we move between them, but the link between these two systems is not clearly defined. Here we develop a model of idealized attentional allocation and study how the efficient distribution of attentional resources might interface with memory to affect learning outcomes. Attentional switching is beneficial for learning in infants. However, when deregulated in ADHD and other attentional disorders, it can also lead to negative learning outcomes. Understanding which attentional behaviors are ideal for information gain would be helpful in many contexts across development, and in the study of attentional disorders.

II. The Model

Our model is meant to formalize a default baseline for what type of attentional behavior should be expected in a system that tries to efficiently gather information from a complex environment. It models the attentional allocation of a participant (the "learner") viewing a display with multiple objects that he is asked to attend to, similar to a MOT paradigm. There are parameters set for (A) the learning curve that the learner follows as he attends to the stimulus, (B) memory decay rate, as a model of the learner's short term memory limitations, (C) the cost of switching attention between objects, and (D) the amount of information that the learner knows about the objects at the start of test. Attention allows accumulation of data about an object, but there is a diminishing return as the learner comes to know everything that can be learned about each object. Importantly, these objects are best thought of as dynamic stimuli that will change when unattended, meaning that information will be lost about them when attention is oriented elsewhere.

At each timestep, the model computes the amount of information that it expects to gain from each object, and chooses whether to continue to attend to the current object or to switch to another based on the expected information gains. Through this computation, the model demonstrates how these components (learning rate, memory decay, switching cost, and prior information) may interact to determine attentional behavior. As we show, these simple properties can lead to surprising, nontrivial dynamics and attentional patterns.

A. Learning Curve

In order to model the acquisition of information about an object that is attended, we use a Gompertz growth curve

$$y(t) = ae^{-be^{-ct}} \tag{1}$$

where *y(t)* is the amount of information known about each object at time *t*, *a* is the maximum amount of information that can be learned from each object (fixed for this demonstration at 100 bits of information), *b* characterizes the amount of initial information, and *c* is the learning rate, here fixed to 1.0 for simplicity.

This function allows us to model the slow learning rate when a novel object is first encountered, which is then followed by a sharp increase in information gain. Learning returns to a slower rate as the learner approaches knowing all of the information about each object (the point at which the asymptote is reached). The starting location on the curve represents the amount of information about the objects the learner has at the beginning of the simulation.

B. Information Decay

The model also takes into account the fact that information about objects in our environment decays for unattended objects. This would naturally occur for objects that *change* dynamically in the environment and whose changes can only be noticed or processed with sustained attention. It would also occur due to the fact that information about an unattended object must be sustained with noisy memory representations.

The information decay rate was computed using a power law decay function, as is observed in memory very broadly [8], [16]. The power law is given by

$$g(t) = (t+1)^{-\beta} \tag{2}$$

where t is the amount of time that has elapsed since the object was last attended and β is an information decay parameter that is representative of the loss of information about the present state of the world due to a dynamic environment. Here, when attention is moved away from an object that a learner has I bits of information about, the amount of information learners have about the object at time t later is decayed via (2), to be $g(t) \cdot I$.

There is also a debate in the literature about whether the passage of time itself can cause information or memory decay, or if it is only caused by interference of novel visual stimuli [10]. In the present model, decay is calculated at each time step based on the time that has elapsed since the learner was attending to the object in question. Despite using time to calculate the extent of the decay, the model does not rely critically on either account, as time or the interference caused by attending to other objects in the environment could be the source of this information loss.

C. Cost of Switching

Moving attention from one object in the scene to another requires the learner to disengage attention from the current object, locate the new target, and then refocus attention, which takes time as well as cognitive effort (for review, see [11]). This effect is captured by the addition of a *cost of switching* parameter to our model. When potential information gains of attending each object are calculated at each time step, this cost is the amount subtracted from the amount of information available from all objects *except* the one which the learner is currently attending.

D. Attentional Decision

The attentional decision is made using a standard *softmax* choice rule [9]:

$$P(\text{choosing to attend object i}) = \frac{e^{\gamma \cdot I_i}}{\sum_j e^{\gamma \cdot I_j}} \tag{3}$$

where the information I_i of object i is computed as the expected information gain of attending that object via (1) after taking infomation decay into account, including the penalty for switching. The softmax parameter γ interpolates between random choice ($\gamma = 0$) and perfect maximization ($\gamma = \infty$). For the figures presented in this paper, γ was fixed (a priori) to 30, a value which corresponds intuitively to a 95% chance of choosing an option (an object) that will provide 0.10 bits of information more than its alternative.

III. Results

A. Number of Objects

Figure 1 shows the general dynamics of the model, giving the amount of information (y-axis) about each of five objects the learner holds over time (x-axis). Although this was not explicitly built into the model, one interesting prediction is that an ideal learner under constraints of rapid information decay will not have the ability to switch between all objects in the

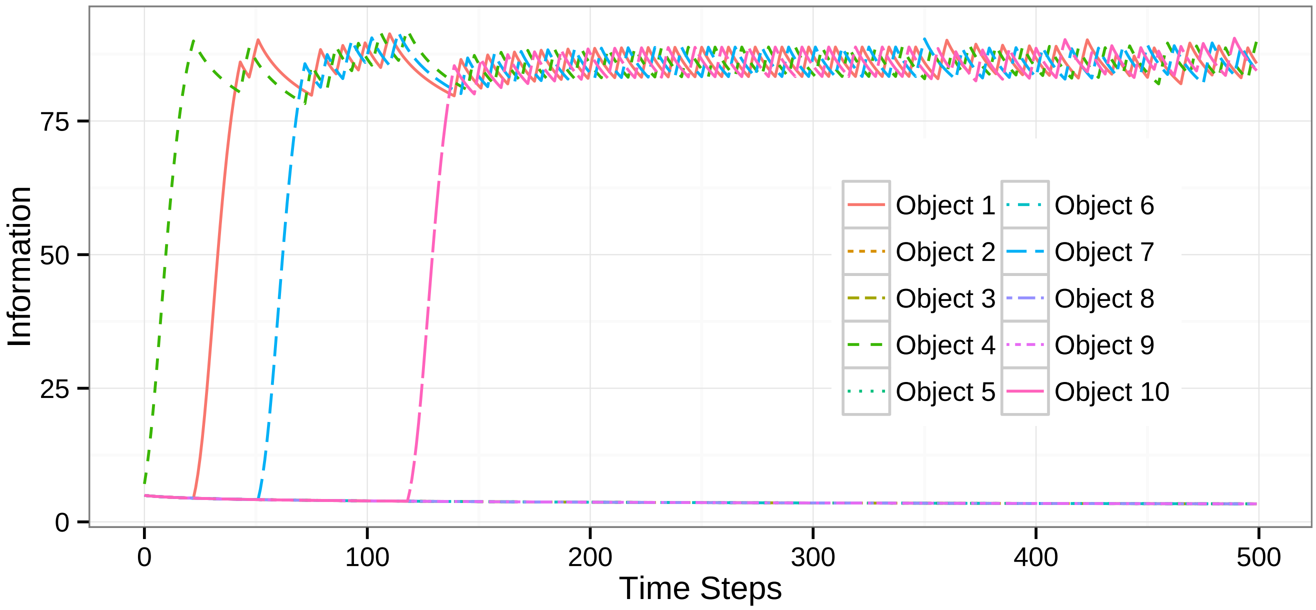

Fig. 1. A representative example of switching behavior across 500 time steps with five objects in the environment. With $\beta = 0.1$ and $switchingcost = 0.5$, an ideal learner is only able to pick up four of the objects, and learn a maximum of 85% of the available information. Once the objects are fixated upon, the learner maintains the information held in memory by switching between all of the previously attended objects. After the maximum number of objects has been picked up, switching behavior becomes cyclical and the amount of information remembered for each object stays relatively constant.

environment, demonstrating an effect like selective attention. If the decay rate is low, all of the five objects can be picked up and learned within 500 time steps. However, with a higher decay rate, only a subset of objects can be maintained at a high enough level, and some objects are left untouched (Figure 2).

There is a maximum rate at which the model can acquire information from each object (given by the maximum derivative of (1)). However, for unattended objects, information is constantly lost. This means that the most effective strategy is naturally to pick as many objects as can be supported (relative to the information decay and learning rates) and attend to them, allowing the system to gather *no* information about the remaining unattended objects. In this way, the model is much like a juggler: only a fixed number of objects can be supported because the hand can only move so fast relative to the rate at which objects fall.

Kahneman, Treisman, and Gibbs (1992) suggested that information about objects in the environment is stored in "object files" that are consistently updated as information about the object increases, as well as when the scene containing the object moves and changes [12]. This imagery fits well with our model, with some slight modifications. In this model, only objects that can be maintained in working memory are picked up from the environment, and this can be thought of as the point at which a file is created. But in addition to adding information to the object file corresponding to the object to which the learner is attending, some information about other objects in the scene is being lost from their respective files. This information will need to be relearned and refiled with the

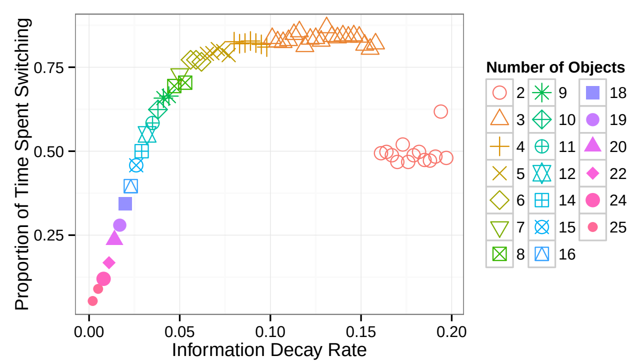

Fig. 2. As information decay becomes more rapid, switching behavior increases and fewer objects can be attended to and learned. The visual environment being modeled contains 50 objects, but the subset of objects selected remains the same for any number of objects >25.

other information as the learner switches back to each object in turn.

B. Attentional Switching

The key part of the "juggling" property of the model is that objects must be continually "held up" in order to prevent loss of information about them. In order to achieve this, the model must switch between objects in a roughly cyclic pattern.

These dynamics are caused in large part by the Gompertz function (1). Growth along this curve is slowest at the beginning and end, leading to the greatest learning potential occurring at the center of the curve. Due to this, objects that have already been previously attended to offer more information gain per time step than one that has yet to be

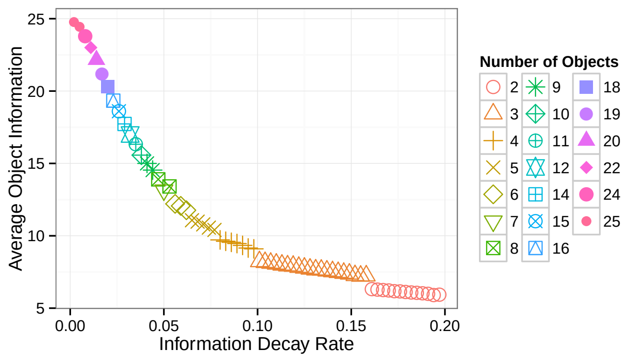

Fig. 3. As information decay increases, both the number of objects attended to as well as the average amount of information learned per object decreases. The visual environment being modeled contains 50 objects, but the subset of objects selected remains the same for any number of objects >25.

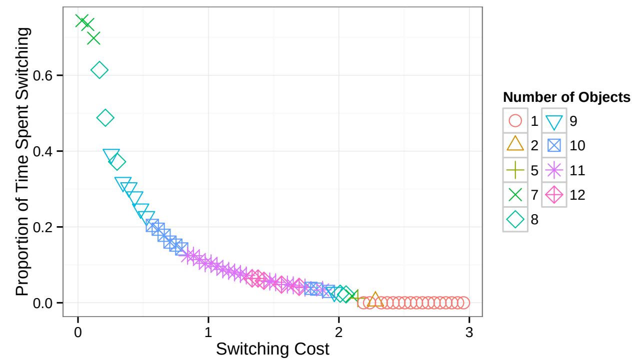

Fig. 4. As the cost of switching increases, time spent switching during the simulation and the number of objects that were attended to decrease. The visual environment being modeled contains 50 objects, but the subset of objects selected remains the same for any number of objects >12.

picked up and remains in the early section of the function. This also leads to a decrease in learning potential once the learner has a lot of information about an object. When an object that is being attended to reaches the point at which its potential information gain at the next time step is less than what can be gained from picking up a new object, the learner switches and fixates the new object.

The number of objects that the learner can maintain is limited depending on the rate of information decay (β) and the cost of switching attention between objects. When there is pressure on the learner due to high information decay, objects that have already been picked up take precedence over those that are untouched. The learner then begins to switch through the objects being held in memory in an attempt to maintain the highest possible information about each. As information about the non-fixated objects decays, those objects return one by one to a location on the learning curve that offers a greater information gain than the object currently being fixated upon, leading to a repeated switching pattern.

This pattern of predictions offers an interesting new perspective on results from past studies on visual attention. It has been demonstrated that when some features of an object were relevant to a task and some were not, only the relevant features were held in memory rather than the object being stored as a whole [13]. Perhaps objects are not coded in their entirety not because we do not have room for that number of features in each memory slot, but instead because we are switching between objects (sampling) in order to keep them in our short-term memory. If there are enough objects in the environment that there is not enough time to cycle among them, it would not be possible to gain all of the available information about their features.

C. *Information Decay*

When all other parameters are held constant, increases in information-decay rate lead to an increase in the switching rate across objects (Figure 2). Fast alternations between objects are necessary, as unattended objects quickly slip from memory. In addition, as the information decay increases, fewer objects can be supported, and the model automatically focuses on the maximum number of objects it can support. Changes in information-decay rate also affect the amount of information that can be gathered from each object in the environment. As the rate of information decay increases, the average knowledge reached between all of the objects drops by almost a third (Figure 3).

Vul, Alvarez, Tenenbaum, and Black (2009) demonstrated that in a MOT task, the number of objects that can be held in memory decreases as the motion of the objects on-screen increases in speed [14]. The model's sensitivity suggests a natural explanation: when attending an object, the speed at which the other objects move around the screen will affect the rate at which information about those objects (positional or velocity information) decreases, thus affecting the number of objects that can be "juggled."

D. *Cost of Switching*

This model also allows for the exploration of differences relating to the ease with which people can disengage and re-engage attention to different objects. If one person can more readily switch between objects in the environment then that might result in an increase in switching behavior and inattention. Based on the organization of this model, it is intuitive that when the cost of switching increases, attentional switching will decrease accordingly (Figure 4). But a subtler issue concerns how these dynamics influence learning. At a set information-decay value, increasing the cost of switching leads to a relatively stable average amount of information known about each object (Figure 5). At a certain point following this high information capacity, the cost of switching becomes so high that switching can no longer be used to effectively maintain all of the objects in the environment, leading to a catastrophic crash in the number of objects that can be supported.

Luck and Vogel (1997) found that participants could successfully learn to identify four different objects each with two features (for a total of eight features), but were not able to do so with eight objects each with only one identifying feature [3]. Based on this, the authors hypothesized that there is a

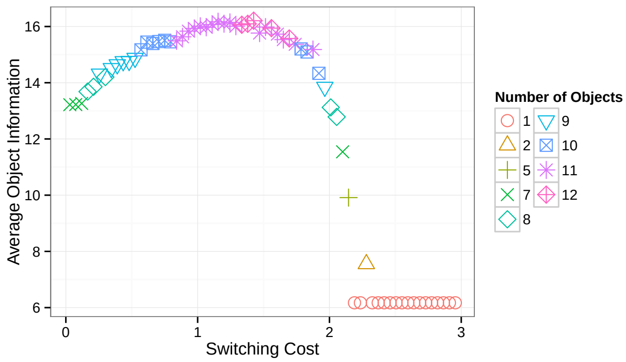

Fig. 5. As the cost of switching increases, the average knowledge about each object remains relatively stable, until the point at which fewer objects can be held in memory and there is a decrease in the amount of information learned. The visual environment being modeled contains 50 objects, but the subset of objects selected remains the same for any number of objects >12.

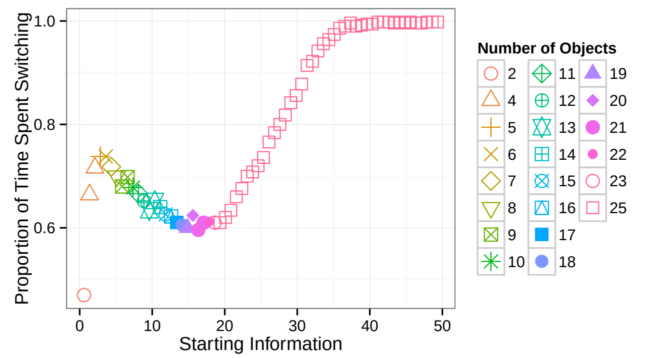

Fig. 6. Increasing the amount of prior information that the learner has about the objects leads to an increase in switching behavior, as well as an increase in the number of objects selected. Total possible information per object = 100 bits. At high values of information, all available objects are selected. Here, 25 are modeled, but he outcome is qualitatively equivalent for any number of objects.

memory constraint on the number of objects held in memory, regardless of feature complexity. Similarly, Fougnie, Asplund, and Marois (2010) demonstrated that remembering details about a single object that has many features has a lower cost on memory than multiple objects that only have one feature dimension [15].

Another possible interpretation of these results is that the cost of switching is measured at the object level. With attention, it will be easy to gain information about the features in a single object, but the additional penalty on switching will prevent attending to *many* objects in order to learn about the same number of features.

E. Starting Information

There is also a possibility that boredom plays a role in inattentional behavior. If a student in a classroom already knows everything about an object or activity that they are being asked to attend to, a tendency to search the environment for other things to learn from seems intuitive. It has been previously demonstrated that infants prefer to attend to information that is both novel and able to be learned within a pre-existing framework of knowledge [17], [18], [19], [20]. In this way, time is not wasted on attending to fully learned information, and more complex cognitive representations can be built.

The model suggests that as the the amount of information that the learner has at the starting point increases, switching behavior also increases (Figure 6). It is possible that some inattentional behavior originates from the learner having previous experience with the objects in the environment and seeking other objects or locations that might offer a higher information gain.

IV. Discussion and Future Directions

A. Behavioral Tasks

Due to its relevance to disorders of attention and novel predictions for the interaction of the visual working memory and attentional systems, it will be important to test the detailed predictions of this model. Similar parameters to those utilized in the creation of the model could be tested using MOT or eye-tracking paradigms in which the relevant parameters are varied.

The cost of switching attention can also be manipulated by changing the visual information load of each object. Past studies have manipulated feature complexity as a way to assess the size of the visual memory store [4]. Similar techniques could be used to alter the cost of disengaging and re-engaging attention by manipulating the complexity of the visual stimuli and evaluating the impact of attentional cost on switching behavior.

B. Model Development

The current model does not break down the information held by each object into distinguishable visual features. This makes formal analysis easier, but human perception may be more easily thought of in features (e.g., color, shape) rather than bits of information. In the future, it would be interesting to create computational representations of features that can be learned after a certain amount of time focusing on the object of interest. In this way, modeling results might be able to be more directly compared to behavioral studies, rather than to information-theoretic ideas that may lie behind psychological systems.

Secondly, the model also does not take into account the context of the objects in the environment and the spatial relationships between them that occur during simultaneous presentation [7]. This could be further explored in future studies, although we would predict that objects that were closer in space might have a lower cost for switching, while those further away might be "picked up" less often, and perhaps remembered with less accuracy due to a greater cost to move attention towards or away from them.

Future versions of the model might also allow for the manipulation of the amount of information held by each object individually rather than one value shared among all of the objects in the environment. In this way, the allocation of attentional resources based on previous knowledge of objects can be further assessed in what might be a more realistic representation of a real-life visual scene.

C. Understanding Disordered Attention

The predictions from this model suggest that high rates of information decay lead to increased switching behavior, as well as a decrease in average object information and number of objects able to be held in memory. This fits well with the observed inattentional behavior and learning difficulties of children diagnosed with attentional disorders such as ADHD, who also demonstrate lower working-memory spans [1].

Modern educational environments require children of increasingly younger ages to sustain focus on classroom activities for extended amounts of time, which makes inattentional behavior even more pronounced. This model demonstrates that ideal learners will show increased switching behavior in response to changes to multiple different cognitive systems.

The use of pharmacological interventions to force sustained attention in the absence of an understanding of the direct cause could be harmful to long-term learning outcomes if the intervention forces the attentional system into a non-optimal pattern of switching. Predictions from this model might provide alternative hypotheses for the origin of disorders of attention and could encourage the development of novel and more effective behavioral treatment options.

V. Conclusion

We present a model of an ideal learner in an environment containing a fixed number of objects. Each object contains a set amount of information to be learned, some of which can be assumed to have been acquired previously, based on a set starting level of knowledge. At each time step, the potential information gain from each object is calculated based on a combination of potential for information gain and the cost of shifting attention between objects. The learner then chooses to fixate the object that offers the highest information gain, using a soft maximization. Memory capacity for previously attended objects is limited, so once an object is learned, maintenance is required to sustain the information about each object.

We demonstrate that a model of an ideal learner defined by learning rate, memory decay, cost of switching, and amount of starting information, will show unexpected patterns of attentional switching in order to maximize information gain. Based only upon these parameters, we also find emergent memory limits as well as limitations on the amount of encodable information that can be obtained from the environment.

This formal model allows us to generate testable hypotheses about the relationship between the visual memory and attentional systems. This is an important step towards gaining an understanding of the systems that govern the underlying dynamics of attention and learning. Further knowledge of these systems can aid in the development of more effective interventions for disorders that impair their function.

References

[1] Kasper, L.J., Alderson, R.M., & Hudec, K.L. Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): a meta-analytic review. *Clinical Psychology Review*, vol. 32, no. 7, pp. 605-617, 2012.

[2] Brady, T.F., Konkle, T., & Alvarez, G. A. A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, vol. 11 ,no. 5, pp. 4, 2011.

[3] Luck, S.J., & Vogel, E.K. The capacity of visual working memory for features and conjunctions. *Nature*, vol. 390, no. 6657, pp. 279-281, 1997.

[4] Alvarez, G.A., & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, vol. 15, no. 2, pp. 106-111, 2004.

[5] Bronson, G. W. Infant differences in rate of visual encoding. *Child Development*,vol. 62, no. 1, pp. 44-54, 1991.

[6] Jankowski, J. J., & Rose, S. A. The distribution of visual attention in infants. *Journal of Experimental Child Psychology*, vol. 65, no. 2, pp. 127-140, 1997.

[7] Krose, B. J., Julesz, B. The control and speed of shifts of attention. *Vision research*, vol. 29, no. 11, pp. 1607-1619, 1989.

[8] Anderson, J.R., & Schooler, L.J. Reflections of the environment in memory. *Psychological Science*, vol. 2, no. 6, pp. 396-408, 1991.

[9] Bishop, C.M. *Pattern recognition and machine learning*, vol. 4, no. 4, p. 12, New York: Springer, 2006.

[10] Cowan, N., & AuBuchon, A.M. Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review*, vol. 15, no. 1, pp. 230-235, 2008.

[11] Monsell, S. Task switching. *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 134-140, 2003.

[12] Kahneman, D., Treisman, A.,& Gibbs, B.J. The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, vol. 24, no. 2, pp. 175-219, 1992.

[13] Droll, J.A., Hayhoe, M.M., Triesch, J., & Sullivan, B.T. Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 6, pp. 1416, 2005.

[14] Vul, E., Alvarez, G., Tenenbaum, J.B., & Black, M. J. Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, pp. 1955-1963, 2009.

[15] Fougnie, D., Asplund, C.L., & Marois, R. What are the units of storage in visual working memory? *Journal of Vision*, vol. 10, no. 12, pp. 27, 2010.

[16] Wixted, J.T., & Carpenter, S.K. The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, vol. 18, no. 2, pp. 133-134, 2007.

[17] Kidd, C., Piantadosi, S.T., & Aslin, R.N. The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE*, vol. 7, no. 5, pp. e36399, 2012.

[18] Kidd, C., Piantadosi, S.T., & Aslin, R.N. The Goldilocks effect in infant auditory cognition. *Child Development*, vol. 85, no. 5, pp. 1795-804, 2014.

[19] Kinney, D.K., & Kagan, J. Infant attention to auditory discrepancy. *Child Development*, pp. 155-164, 1976.

[20] Piantadosi, S.T., Kidd, C., & Aslin, R.N. Rich Analysis and Rational Models: Inferring individual behavior from infant looking data. *Developmental Science*, vol. 17, no. 3, pp. 321337, 2014.