DoorWay CTM: Reproduction and Explanation

Lan Lyu, Bowen Su, Chang Ma

1 Introduction

It's a specific case where people suddenly forget what they are going to do when they encounter a new environment. This is a everyday phenomenon, and we want to explore whether CTM can exhibit this phenomenon like human beings.

The phenomenon can be psychologically explained by Radvansky's Event Horizon Model[Radvansky, 2012]. According to this model, event boundaries are what causes sudden forgetting —— streams of information were segmented into event units, and the doorway or the boundaries between segments cause sudden forgetting, as people tend to update their event models when crossing event boundaries. Such update activate new model and deactivate previous model , and leads to the doorway effect.

Our proposal is to investigate the Doorway Effect via the Conscious Turing Machine. We hope to model an event horizon with the Conscious Turing Machine(CTM)[Blum & Blum, 2021], i.e. reveal the machine to two consecutive events and see if it experiences sudden forgetting, as well as finding a theoretically and psychologically plausible explanation as for why this phenomenon occurs.

2 Related Works

Radvansky conducted a series of studies[Radvansky et al., 2011, Radvansky et al., 2010] to explore the effect of walking through doorways. In one experiment, people were asked to move through virtual space, and transfer an object from one table to another. Researches design two settings of the experiment, where one involves location shift, while the other doesn't. In the shift condition, people moved from one room to another (with a doorway). In the no-shift condition, they traveled a similar distance but they did not change locations (without a doorway). The result was that it took people longer to respond for a response and candidates made more errors when there was an event shift than when there was not. In other words, walking through doorways caused forgetting.

This document is written for the course Conscious Turing Machine, Peking University.

The doorway effect not only occurs when there is a change of location. When an event shift occurs in reading, the reading time for sentences increase. Besides, people respond more slowly to memory probes about the detailed plots of a story if the text has shifted to a new event than if that event is still ongoing[Radvansky & Copeland, 2010, Zwaan, 1996]. In narrative comprehension, when people encounter event boundaries, they mentally update their models[Zwaan et al., 1995].

To develop a model to explain the phenomenon described above, we assume people (and CTM) segment the stream of information into event units. This assumption is drawn from the Event Segmentation Theory[Kurby & Zacks, 2008]. When people encounter and identify an event boundary, they segregate information into event models that are stored in memory. For example, if a person walks from one room to another, this is often a new event. Thus the information received before and after the doorway is segmented and separately input into model.

3 CTM Explanation for a Basic Scenario

Following Radvansky's work [Radvansky et al., 2011], we model our CTM in the Event Horizon scenario, in which there are two successive events A, B, each constitutes of a series of time series inputs $a_1, a_2 ..., a_t, b_{t+1}, b_{t+2}, ..., b_{2t}$. At time t, the CTM crosses the Event Horizon between A and B.

For example, there is a person holds a box and wants to put the box in the next room. The surroundings in room A is different from that in room B. When he walks through the room, he will not only be aware of the box he is holding, but also take in the view of his surroundings. The doorway between room A and B correspond to the Event Horizon, and t stands for the time he needs to pass through a room.

Below we first explain how CTM might achieve the Doorway effect in this scenario. Then we make minor alterations on CTM to better reproduce the phenomenon.

4 Preliminary Modeling of Forgetting

We first consider the following CTM model.

- Three chunks, one for remembering the task to put the box on the table (Chunk Action), one for visual input of surroundings(Chunk Visual), and one for audio input of surroundings(Chunk Audio).
- $Chunk_{p,t} =< address_p, t, gist, weight, intensity = |weight|, mood = weight >$, the weight here can be interpreted as the alertness level of a certain chunk. High weight shows that a chunk is on high alert.
- The competition function is $intensity + \frac{1}{2} mood$

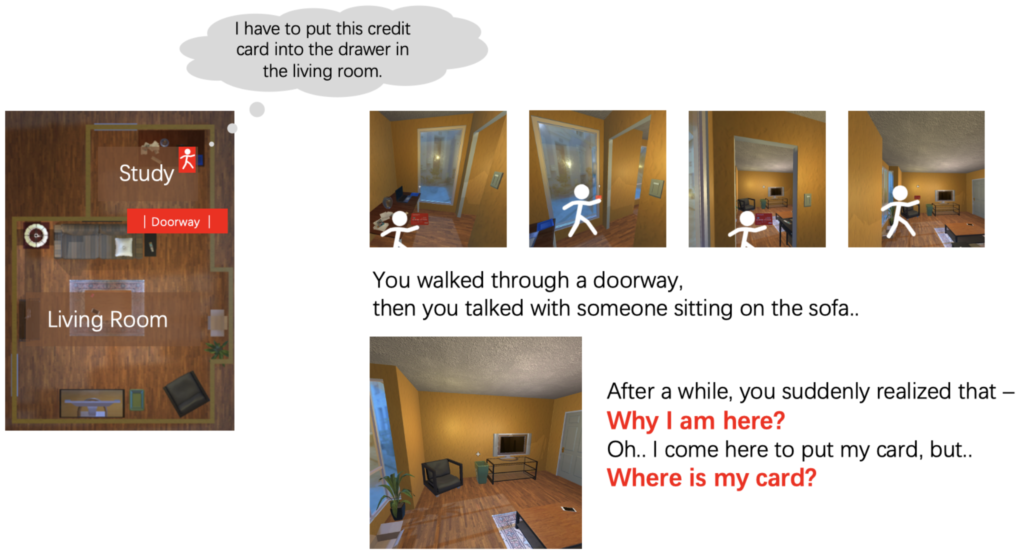


FIGURE 1: The scenario - one walked through a doorway and forget why he/she came here

- The processors give the result for input in the same room, i.e, $p(a_0) = p(a_1) = \ldots = p(a_t), p(b_{t+1}) = p(b_{t+2}) = \ldots = p(b_{2t})$

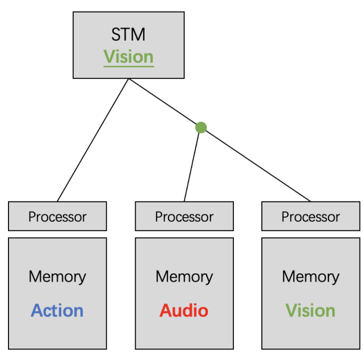


FIGURE 2: The CTM Tree Structure

Now we consider the following scene: a person holds a small box through room A, which has a calm surrounding. Therefore, he doesn't need to put much attention on visual and audio surroundings. The chunks have weight :

$$Chunk_{action} =< weight = 10 >$$

$$Chunk_{visual} =< weight = -15 >$$

$$Chunk_{audio} =< weight = -4 >$$

Therefore, the action chunk wins the competition, and the note that "I need to place the box in room B" stays in Short Term Memory(STM).

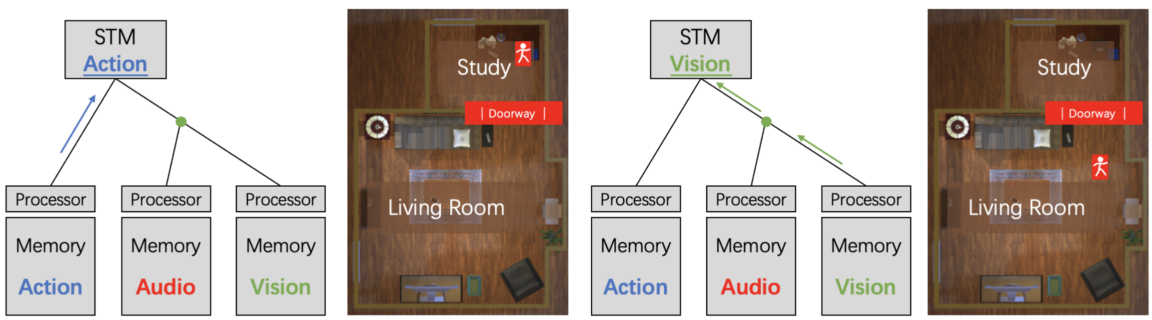
However, after entering the room B, he sees some friends having a party, and the surrounding is extremely noisy, maybe a friend of him will chat with him. Therefore, the visual and audio chunks are on high alert, and have positive weights. The action chunk doesn't change because the box that person is holding hasn't changed. The weights become:

$$Chunk_{action} =< weight = 10 >$$

$$Chunk_{visual} =< weight = 15 >$$

$$Chunk_{audio} =< weight = 4 >$$

The visual and audio chunks win the competition and the action to be done is removed from STM. Therefore, the person experiences sudden forgetting. He may never thinks about the box until he comes back to room A, where the weights are restored and the action goes back to STM. This accounts for why we sometimes remember what we sudden forgot after returning back to the previous surrounding.



a)in study "action" wins the competition b)in living room "vision" wins the competition

FIGURE 3: The CTM tree when in study and living room.

5 Another Explanation from Processor View

However, the explanation above can't explain the case - the man entered room B, and talked with some friends, he may smoothly put down the box somewhere nearby without noticing it. Afterwards when the box comes back into his mind, he can hardly remember where the box has been placed. So we improved our CTM model to fit this situation.

The preliminary CTM model mentioned in section 3 reveals that the main cause of doorway effect is the change in our ways to process information. When we encounter a new environment like room B, we will develop a different way to sense and understand the surroundings, thus our goal generated in room A will be obscured.

We assume that when walking into a new room, CTM will produce a new processor to learn about the new scene. So if a CTM moves across between room A and room B, two different processors gain control, specifically, when in room A processor A wins the

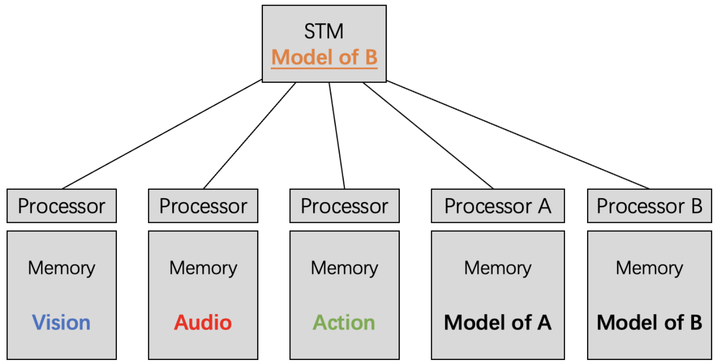


FIGURE 4: Improved CTM Tree Structure in room A

competition, when in room B processor B will win the competition and replace A as the winner.

After B wins the competition, it will broadcast to all chunks, thus influencing the action part, making the person to put down the box as he is still in high alert in visual and audio cortex. Because the chunk of action didn't win the competition, the action of placing the box will not be noticed.

6 Encoder-Decoder Model

In field observation the Doorway effect is a small probability event, for most time a person won't entirely forget what to do, but often his memory would not be precise due to the Doorway problem. To make our model closer to the reality, we expect the CTM to display different degrees of forgetting, being capable of remembering things to do after moves across varied environments, partly forgetting what should be done , or entirely forgetting actions.

Inspired by CTM and autoencoders[Tschannen et al., 2018], we designed an event model with varying parameter so as to better describe the process.

First, we want the model to encode the information of the environment, so that it can encode in one place, and decode in another, with the abstract learnt knowledge stored in it. We define f_θ, a mapping function that transforms the first-hand raw information into cipher (θ could only be changed by previous information input). f_θ^{-1} is the inverse function of f, denoting the decode process. For example, when the man walked into room B to look around, he takes in the information of room B in the form of $f_{\theta_B}(B)$. When he tries to recall something, that involves the inverse process ,during which the stored cipher is decoded into specific event $f_{\theta_B}^{-1}(B)$. When he succeeds in remembering what to do - the idea encoded in the room A, this process can be expressed as $f_{\theta_B}^{-1}(f_{\theta_A}(A)) = A$. And if the equation doesn't hold, it indicates a Doorway effect. This means you can't think of things remembered in room A.

Here we formally define the problem.

$Definition 1. f_\theta : E \longrightarrow R^n$, E is the space for all events,R^d is the representation space for all events. f_θ stands for the encoding process, and inversely f_θ^{-1} stands for the decoding process.

$Definition 2$. The doorway process occurs when $f_{\theta_B}^{-1} f_{\theta_A}(A) \neq A$, A,B are names for events, and θ_A denotes for the parameters used decoding the event.

The reason a parameterized encoder f_θ is needed is because we don't often exhibits "hard forgetting", ie.,entirely forgetting all the details , it is more often that you know you forget some details, but you still recalls part of the picture. We design an experiment based on colored objects. It showed in this experiment that f_θ^{-1} can remember some parts while make mistake in others.

For example, in room A we expose the model to 10 object - "blueball, bluecube, bluecube, bluebook, darkbook, darkball, pinkbook, azurlane, pinkball, pinkbook" - and f_A works well and encodes and stores all these things. In room B, after looking around and seeing 5 objects in B, we question the model about what were the last five objects it had seen in room A, the result was half-right, in line with our vague memory in Doorway effect.

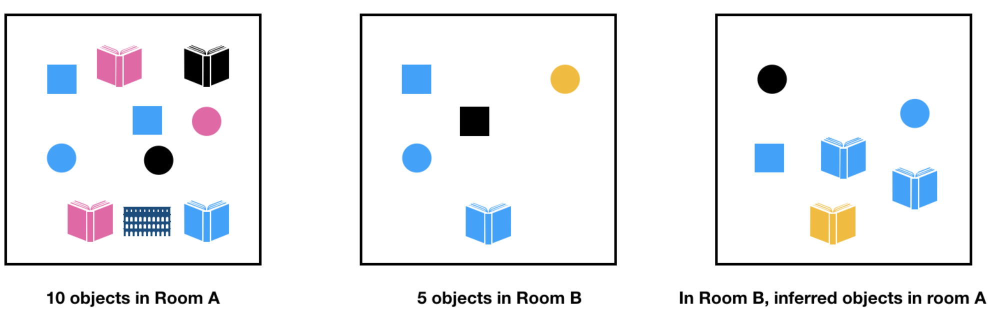


FIGURE 5: Illustration of Experiment: The person can recall blue and dark book in room A because their shape and color are both encoded in the perception of present room B. However, there are no pink objects in room B, so pink is not currently encoded and cannot be recalled.

We model this phenomenon with a C++ program. Implement details will be developed later.

6.1 An example of f_θ based on linear basis

Definition3.Encoder on linear basis : $f_\theta : R^n \to R^n, f_\theta(v_t) = \sum_{i=1}^{t-1} \theta_{t,i} * v_i + \theta_{t,t} * v_t, v_i \in R^n$

In this example, we assume that input information are vectors with a fixed length. We regard the vector in space R^n. Since R^n is a linear space, we can maintain the maximal linearly independent set of inputted vectors. [Knapp, 2007]This set is the basis of current span space, thus we can encode the inputted vectors in terms of linear combinations of the basis (recording the coefficients).

We take this encoding method as f_θ, and the addition of basis refers to a change of θ. When a new vector(information) is inputted, we add it into the basis set if it is linearly independent with the vectors in the set. Then we write all input as linear combinations of the basis and record the coefficients. The model record the basis set, along with coefficients.

The model initially has an empty basis when it is created.

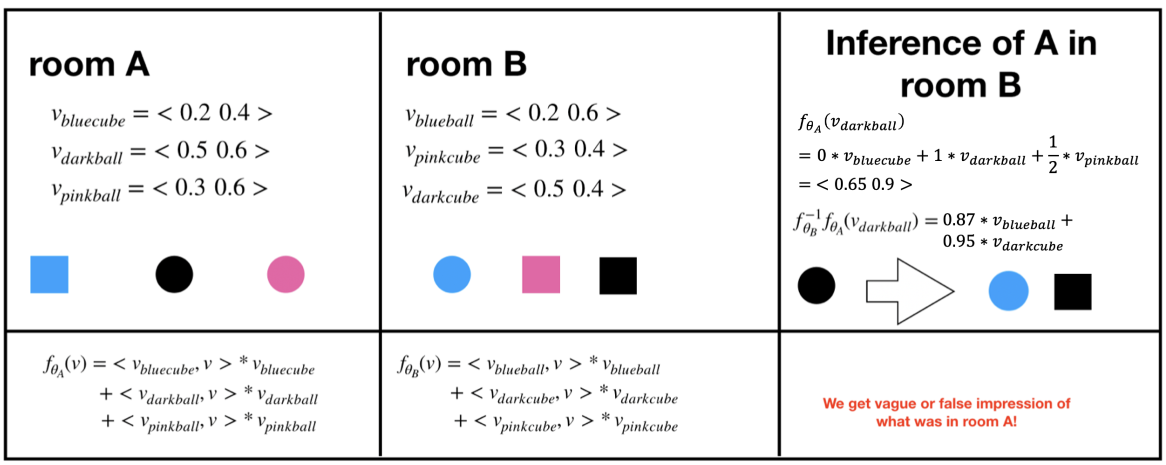


FIGURE 6: Illustration of vague or false memory - The <v1,v2> denotes the similarity between objects.

In the above experiment, we provide 10 pieces of information for model A, 5 pieces for B, then try to decode the last five recorded cipher in A with both θ_A and θ_B, it is expected that the decoding with θ_A is always correct (nothing forgot without a doorway) ,while decoding with θ_B sometimes there are errors (Doorway Effect).

Note that model B can sometimes correctly recall part of the information (like when recalling darkball, it is remembered as darkcube). This explains the vague memory we exhibit when remembering some objects.

7 Conclusion and Future Work

In this paper,we discussed the Doorway Effect in the context of CTM. Our contributions come in three folds.

- We formalize the definition for the Doorway Effect.
- We model the Doorway Effect with Conscious Turing Machine, and discussed the impact of event horizon on forgetting and recalling memory.
- Inspired by CTM, we built a encoder-decoder model that exhibits vague or false memory.

However, the encoder-decoder model meet hash linear basis requirements in order to correctly decode. Also, we only model the event horizon between two events. The situation would be more complicated when there are multiple consecutive or overlapping events. We will focus on refining the model and settings of our discussion.

References

[Blum & Blum, 2021] Blum, M. & Blum, L. (2021). A Theoretical Computer Science Perspective on Consciousness.

[Knapp, 2007] Knapp, A. W. (2007). *Advanced algebra*. Springer Science & Business Media.

[Kurby & Zacks, 2008] Kurby, C. A. & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences*, *12*(2), 72–79.

[Radvansky, 2012] Radvansky, G. A. (2012). Across the Event Horizon. *Current Directions in Psychological Science*, *21*(4), 269–272, https://doi.org/10.1177/0963721412451274 https://doi.org/10.1177/0963721412451274.

[Radvansky & Copeland, 2010] Radvansky, G. A. & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 210.

[Radvansky et al., 2011] Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *Quarterly journal of experimental psychology*, *64*(8), 1632–1645.

[Radvansky et al., 2010] Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic bulletin & review*, *17*(6), 900–904.

[Tschannen et al., 2018] Tschannen, M., Bachem, O., & Lucic, M. (2018). Recent advances in autoencoder-based representation learning. *arXiv preprint arXiv:1812.05069*.

[Zwaan, 1996] Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, memory, and cognition*, *22*(5), 1196.

[Zwaan et al., 1995] Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, *21*(2), 386.