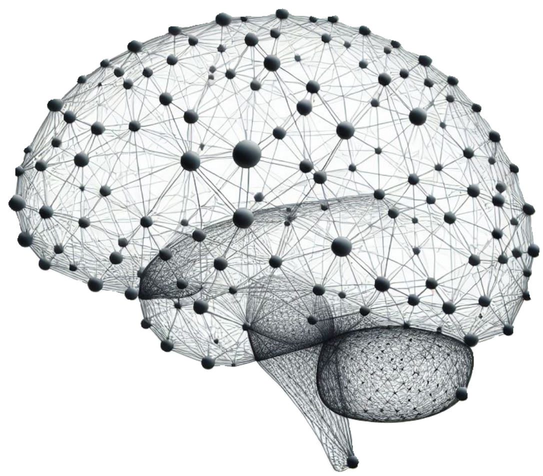
Prediction and pattern recognition

Predictive information decay and its impact on performance in learning tasks

Martijn Klop



Research Project Complex Systems Profile

Institute for Theoretical Physics
Utrecht University
The Netherlands
February 14, 2024

Abstract

Humans frequently have to learn sequential data through pattern recognition or prediction. Data can come from a wide variety of sources and in a wide variety of forms, such as melodies, numbers or texts. An underlying theory about how humans learn, that is, to make an underlying model of the data generation, is still lacking. In this study, we investigate the usefulness of an information theoretical concept, predictive information decay, in human information processing. A simple experiment is designed and piloted that distinguishes how well people perform on pattern recognition tasks when presented with sequential data with varying predictive information or varying predictive information decay. The underlying data is generated through random walks on k-regular graphs. After tentatively concluding that predictive information or predictive information decay rate do not provide clear correlates to performance on recognition tasks, we provide advice and inside for a more extensive follow-up study.

Acknowledgements and supervision

Dr. Clelia de Mulatier - Initiator and primary supervision on theory.
Dr. Samson Chota - Supervision from Utrecht University, grading and support with experimental part of the project.
Dr. Eugenio Piasini - Advise and attedance at biweekly meetings.

Contents

1 Introduction

Learning and memory are among the most studied subjects in the behavioural sciences. In daily life, people are subject to a wide array or learning and memorization tasks, be it recalling a password, studying a piece of music, learning a new skill or differentiating between the faces of actors in movies. Models of human memory are diverse and often distinguish between implicit and explicit memory or recognition or recall memory [1][2][3]. In this study, we are going to focus on the memorization of sequential data, specifically in terms of pattern recognition.

To perform recognition effectively, people need to be efficient at deciding what information they pick up on. For example, people can confidently recognize a familiar melody, without knowing how the melody continues when they have only heard part of it. An important part of effective memorization is thus to find out how to store some information in a short-hand way. One can imagine that recognition depends on an internal, implicit model being built next to which a stimulus can be compared for familiarity. If and how exactly humans do this and what maxes them effective pattern recognisers, remains elusive [4] [5].

To shed light on such questions, it is a logical choice to look at information theory for answers, as this domain of science is able to quantify and qualify otherwise vague notions such as 'patterns', 'relevance' and 'informativeness'. The research field of information theory aims to provide a scientific and quantitative framework for the analysis of information and communication. It quantifies the amount of information transmitted through some channel by analysing it in the context of a stochastic signal-generating process that follows an underlying statistical distribution, often combined with an external noise generating process. Formalisms in information theory generally depend on establishing a sender, receiver, information channel and a syntax. The latter states all possible messages that could in principle be sent from the sender to the receiver through an information channel. When this is known, and we know something about the prior expectations of the receiver, we can make quantitative statements about how much information the receiver gains from the data. [6].

In this study, data is going to take the form of a random walk on a network with fixed transition probabilities between its discrete states. Such random walks are *Markovian* as the information observable at the current time t constitutes the full past of the system, which fully characterizes the probability of transitioning to any possible next state[7]. This is one of the simplest possible setting in which pattern recognition can be studied, but allows for very specific manipulation of information theoretical features of the data. We will use a relatively novel information theoretical concept which is ideal for these type of settings, to investigate how people build models of the perceived data in recognition tasks.

The aim of this study is to investigate the usefulness of *Predictive Information Decay* in understanding human performance in information processing and pattern recognition. Predictive information is the amount of information that the past yields about the future in an ongoing data sequence, such as the random walks used here. In the past, studies by Lynn et al. (2020) have examined the relation between predictive information of random walks on networks and performance on prediction tasks. Here, we will extend this to *Predictive Information Decay* by examining differences in predictive information that the states of the random walks provide about states further in the future, instead of looking only one step ahead. The motivation for this is that it seems logical that when humans try to recognize patterns in data, they look for longer reoccurring trajectories, not only at frequent transitions from one state to another.

The result of this study is a pilot of an experiment that could result in a full scale study, with insight and advice on how to proceed. Ultimately finding links between information theoretical concepts and human performance on pattern recognition tasks could help with developing models of human recognition memory that can make falsifiable predictions [8].

2 Network Topology and Predictive Information Decay

To give the proper context to the experimental design, we will introduce some necessary concepts related to network structures and information theory. For people familiar with graphs as mathematical objects, the first subsection can is optional.

2.1 k-Regular, complete graphs

Graphs are mathematical objects that form graphical representations and interpretations of matrices. Graphs consist of discrete units called *nodes* that are connected to one another through *edges*. A graph with n nodes can be represented by an n-by-n matrix T, the elements of which (T_{ab}) represent if node a is connected with node b. There are many varieties of networks, but we will limit my focus to *undirected* graphs without self-loops. That is, all nodes that are connected are connected bi-directionally and a node can not be connected with itself. In mathematical terms, T is symmetric and has no diagonal terms. Furthermore, we are going to look at graphs in which all edges are of equal weight. The values of our matrix T will hence take a binary form. They are either 0 or $1/k$, with k the amount of edges of a certain node, which we will call the *degree*.

The study of networks concerns itself mainly with questions about *architecture* and other global properties of graphs. As the values on matrices are identified with connections between nodes, we can ask questions such as 'what are the average distances between any two nodes?' or 'what distribution do these distances follow?' or 'how are the degrees of the nodes distributed?'. We will once again limit ourselves to a particular type of graphs, the k-regular graphs. In these graphs, all nodes have the same amount (k) of edges. This leaves us to vary only two things: the amount of edges k and the *structure* of the graph.. There are multiple metric through which to formalize structure, and in the next subsection we will look at one that is particularly relevant to information processing.

2.2 Predictive Information

In this study, k-regular graphs will be used to perform random walks on. The procedure is as follows:

1. Pick a starting node x_0.
2. Find the list of all its neighbours, corresponding to row x_0 of the matrix T
3. Randomly pick one of the nonzero elements $T_{x_0,y}$ of the matrix, weighted by its values (which are all equal here)
4. Repeat this process from node y and mark it as x_1

From now on, we will sometimes refer to the nodes as 'states' and as the matrix T as the 'transition matrix'. This random walk is an example of a Markovian process, as the probabilities for the next possible states x_{t+1} only depend on the currently observable state x_t, that is $P(x_{t+1} = j|x_0, ..., x_t) = P(x_{t+1} = j|x_t)$.

In information theory, information is usually quantified as

$$S_P = \sum_X P(x) \log P(x) \tag{2.1}$$

Where S_P stands for the *Shannon information* contained in the probability distribution $P(X)$ for all possible x in some set X.

Predictive Information is roughly defined as the amount of information provided about the future by observing the past. In mathematical terms it is the expectation value for the difference in information between seeing some future x_{future} and some past x_{past} together and your prior of observing these future and past independently [9], or

$$I_{\text{pred}} = I(X_F; X_P) = \left\langle \log_2 \frac{P(x_{\text{future}}, x_{\text{past}})}{P(x_{\text{future}})P(x_{\text{past}})} \right\rangle_{x_{\text{future}}, x_{\text{past}}}. \tag{2.2}$$

This can be rewritten in terms of Shannon information as

$$\begin{aligned} I_{\text{pred}} &= -\langle \log_2[P(x_f)] \rangle_{x_f, x_p} - \langle \log_2[P(x_p)] \rangle_{x_f, x_p} + \langle \log_2[P(x_f, x_p)] \rangle_{x_f, x_p} && (2.3) \\ &= S[X_F] + S[X_P] - S[X_F, X_P] && (2.4) \end{aligned}$$

Or, in terms of conditionality, using $P(A|B) = P(A,B)/P(B)$, we write

$$I_{\text{pred}} = S[X_f] - S[X_f|X_p] \tag{2.5}$$

Now we can make use of the Markov Property to show that.

$$\begin{aligned} P(x_f) &= P(x_1)P(x_2|x_1)...P(x_T|x_{T-1}), && (2.6) \\ P(x_p) &= P(x_{-\tau})P(x_{-\tau+1}|x_{-\tau})...P(x_0|x_{-1}), && (2.7) \\ & && (2.8) \end{aligned}$$

Such that

$$\begin{aligned} P(x_f, x_p) &= P(x_{-\tau})P(x_{-\tau+1}|x_{-\tau})...P(x_1|x_0)...P(x_T|x_{T-1}) \\ &= P(X_F)P(x_1|x_0)P(X_P). && (2.9) \end{aligned}$$

This allows us to rewrite the predictive information that the past gives about the future as

$$\begin{aligned} I_{\text{pred}}(\tau, T') &= \left\langle \log_2 \left[\frac{P(x_f, x_p)}{P(x_f), P(x_p)} \right] \right\rangle_{X_F, X_P} && (2.10) \\ &= \left\langle \log_2 \left[\frac{P(x_f)P(x_1|x_0)P(x_p)}{P(x_p)P(x_1)P(x_p)} \right] \right\rangle_{X_F, X_P} && (2.11) \\ &= \left\langle \log_2 \left(\frac{P(x_1|x_0)}{P(x_1)} \right) \right\rangle_{x_1, x_0} && (2.12) \\ &= \sum_{x_1, x_0} P(x_1, x_0) \log_2 P(x_1|x_0) - \sum_{x_1, x_0} P(x_1, x_0) \log_2(P(x_1)) && (2.13) \\ &= \sum_{x_0} P(x_0) \left(\sum_{x_1} P(x_1|x_0) \log_2 P(x_1|x_o) \right) - \sum p(x_1) \log_2(p(x_1)) && (2.14) \\ &= S[P(x_1)] - \langle S[P(x_1|x_0)] \rangle_{x_0}. && (2.15) \end{aligned}$$

And hence, we only have to look at the difference in prior information on the next state x_1 and the information on x_1 upon observation of x_0, averaged over all possibilities. For a random walk on a k-regular network of degree m and with n nodes, it easily follows that $I_{pred} = \log_2 \left[\frac{n}{m} \right]$

2.3 Predictive Information Decay

In order to connect the measure of predictive information to pattern recognition, we need to expand it towards predictions further than one step ahead in the future, as patterns necessarily involve multiple transitions. This is also a more relevant measure for random walks on k-regular networks, because networks with the same amount of edges but a different architecture generate different data, even though they have the same predictive information, which is fully determined by the amount of nodes and the degree. If a network has more clusters, we expect it to stay in the same regions of the network for longer, whereas if the network is more lattice-like, the random walk will quickly cover all areas of the graph. In more modular graphs, there are often multiple ways to get from a node n to a nearby node m using t steps, such that this transition occurs more frequently. This is not captured in predictive information, which only considers the next state in a Markov process. Hence, we need to change the concept of *predictive information* to include how much information the observation of the current state x_0 gives about any of the future states x_t.

From the Markov property, we can see that Predictive Information generalises to

$$I_{\text{pred}}(k) = S[P(x_k)] - \langle S[P(x_k|x_0)] \rangle_{x_0} . \tag{2.16}$$

with

$$S[P(x_k)] = S\left[\sum_{x_1,\ldots,x_{k-1}} P(x_0)P(x_1|x_0)\ldots P(x_k|x_{k-1})\right] \tag{2.17}$$

This gives us a new way to characterize k-regular networks. Information about the future propagates through the multiplication of the probability distribution over all states with the transition matrix T, as

$$\mathbf{P}(x_k) = \mathbf{P}(x_0)T^k. \tag{2.18}$$

In the long term, one can show this means that predictive information decays exponentially as

$$\lim_{k \to \inf} I_{pred}(k) = C \exp\{-\alpha k\}; \tag{2.19}$$

$$\alpha = -2\log(\lambda^*). \tag{2.20}$$

In which *lambda** is the second largest positive eigenvalue of the transition matrix T. We will refer to α as the *predictive information decay rate* and $\tau = 1/\alpha$ as the decay time.

3 Methodology

In this experiment, we show participants data generated by random walks on k-regular graphs, without giving them information on the underlying graphs or even on any other aspects of the data generating mechanism. What they observe, is a circle made up of 14 dots. One after another, these dots will light up in a seemingly irregular fashion which corresponds to a random walk on a 14 node network. The participants are instructed to look for any regularities in the pattern and are queried by a forced choice task. After 25 data points, the color of the pattern changes (from red to blue). The participants are instructed to look for a change in the pattern that might or might not occur during the color change by looking at 15 more blue data points. When both the red and the blue data is shown to them completely, participants are asked to judge two data streams as similar or different.

The test consists of two parts. In part I, graphs used for data generation vary in terms of the amount of degrees per node, and therefore the *Predictive Information* ($\log(\frac{n}{m})$). In part II, the graphs are all 4-regular, but have qualitatively very different network architectures that range from a very low to the very highest *Predictive Information Decay Rate*.

Part I consists of 19 trials of a total of 40 steps with a 50% probability of transitioning to a fully random walk and a 50% percent probability of continuing unchanged after 25 steps. A fully random walk implies a random walk on a fully connected undirected network without self-loops. For a 14 node network, this is the same as a 13-regular network. In this part of the experiment, participants are asked to discriminate between typical k-regular networks of varying k and this 13-regular graph. They see 4 random walks each on graphs with degrees between 3 and 6 and 3 of degree 7. See Figure 1 for examples of these graphs. The networks are shown in a random order and which option is correct ('similar' or 'different') is randomized. After seeing 25 red transitions and 15 blue ones, participants are asked to make a decision by answering if the red and the blue patterns were similar or different. See Figure 2 for a schematic representation of what participants are shown.

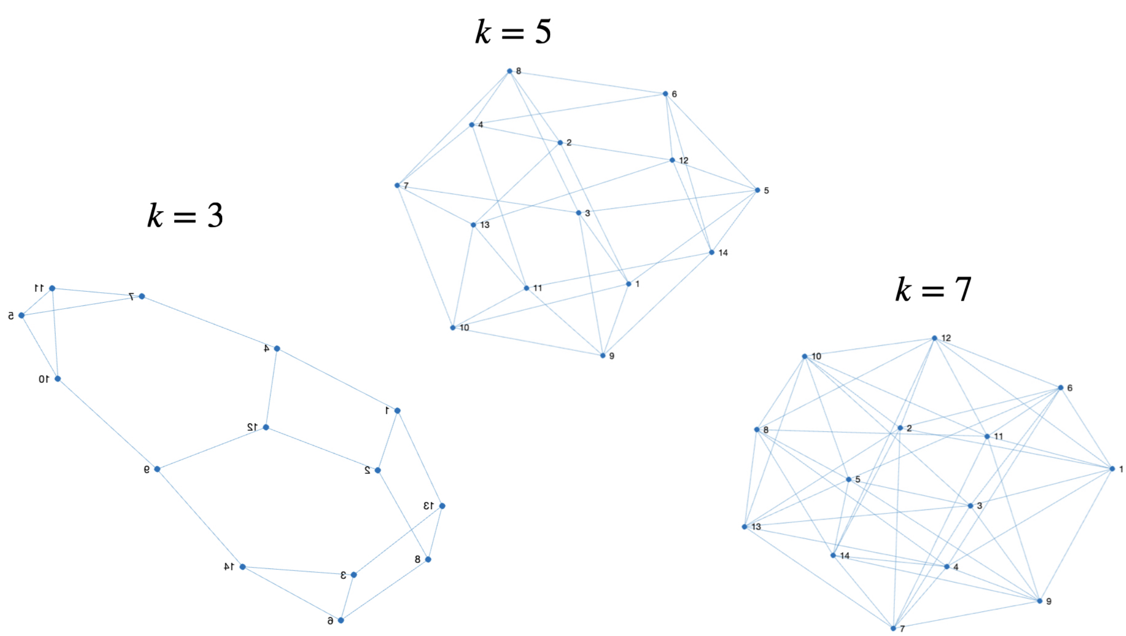


Figure 1: Examples of 3-regular, 5-regular and 7-regular graphs with 14 nodes.

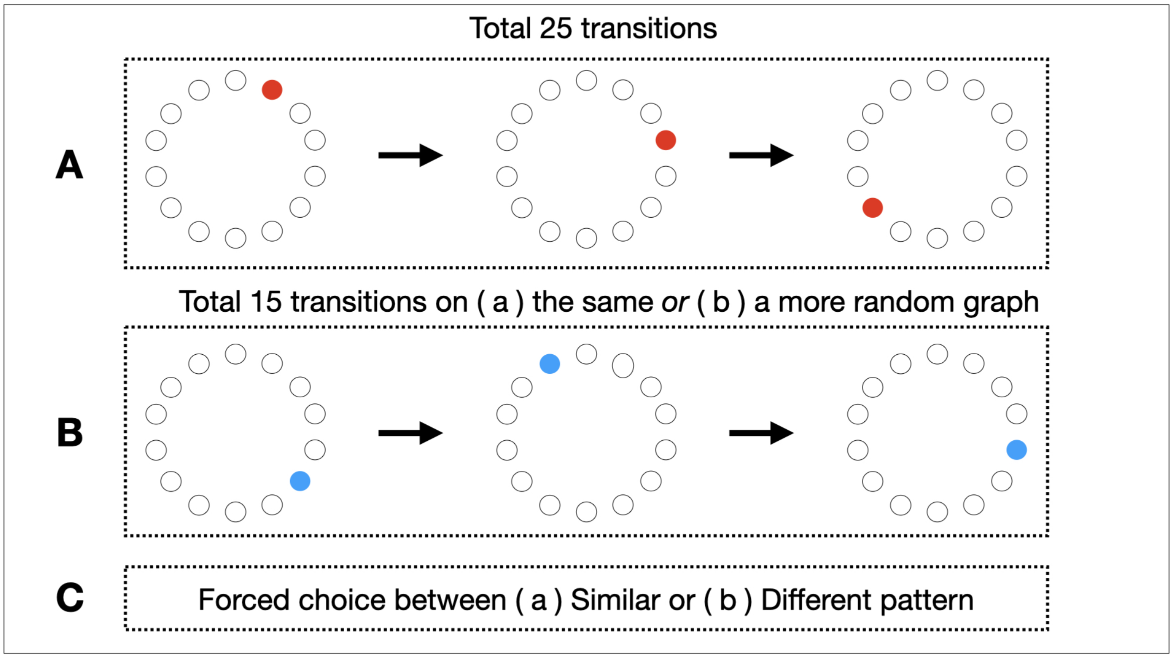


Figure 2: A representation of the experimental set-up. Participants are shown a pattern in the shape of sequentially lit up dots. The pattern changes color from red to blue at some point and participants are asked to spot if there has been a shift in the underlying data generating mechanism by answering if they found the patterns to be similar or substantially different. During the blue part, they are either shown a continuation of the random walk on the same graph, or a random walk on a different, more random graph.

In Part II, participants are shown 20 trials of a total of 40 steps with a 50% probability of transitioning to a more lattice-like graph and a 50% percent probability of continuing unchanged after 25 steps. The 'lattice-like graph' is the 4-regular 14 node network with the shortest Predictive Information Decay time ($\tau = 0.48$). This time, the red-phase random walks take place on ten different 4-regular graphs with Predictive Information Decay times varying between 0.77 and 7.9, see Figure 3. As in part II, the order of the random walks and their underlying graphs shown is randomized, as is the correct response. No two participants therefore get the same order of the experiment and they have no way to know which graph will transition to the lattice-like graph and which will not.

When the random walk does transition from one graph to the lattice-like graphs in the second part of the experiment, it is made sure that this random walk continues from the same node as where the red part ended.

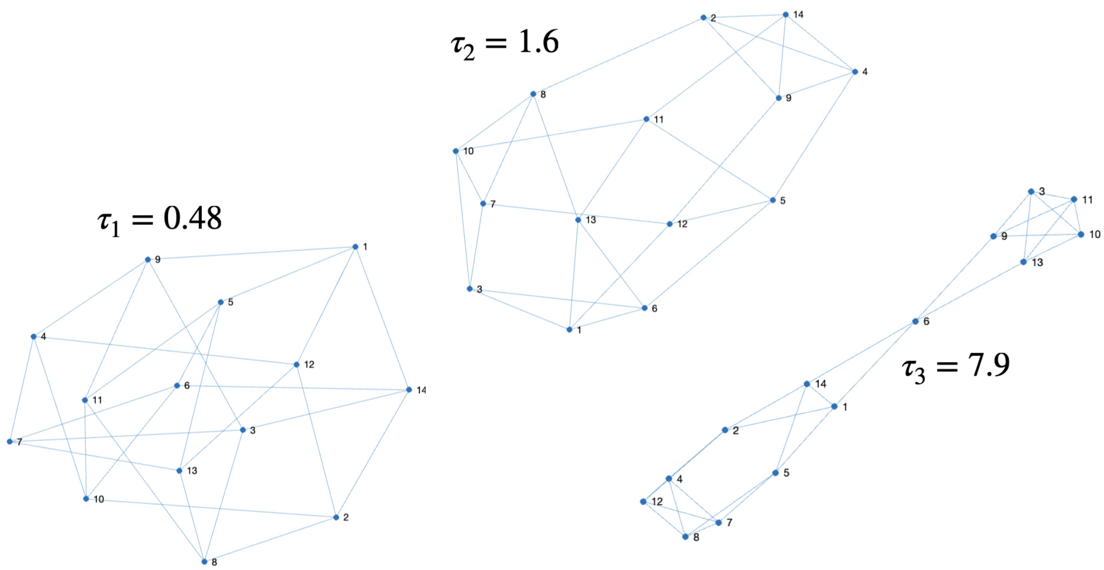


Figure 3: Three examples of 4-regular graphs with 14 nodes. The left graphs has the lowest possible Predictive Information Decay time, the middle has one just above the median and the right has the highest possible Predictive Information Decay times.

This experiment is coded in Psychopy and ran in a Javascript version through Pavlovia. In total, 28 participants have completed part I and 16 participants have completed part II. Due to a fault in the code of the experiment, the amount of data collected for part II was limited.

3.1 Instructions and trials

This experimental design had two trial tasks which were relatively easy, but demonstrated the workings of the rest of the experiment. Participants were shown a random walk on the 4-regular 14 node network with the longest predictive information decay time and in one of the tutorial trials, there was a clear change in pattern whereas in the other the pattern was clearly similar (but not the same). In the instructions, participants are instructed to rely largely on implicit, intuitive judgement instead of explicit memory and problem-solving. They are given a metaphor for the random walk to help with the judgement, see Appendix A.

4 Results

To analyze results, we determine the average correct response rate as a function of degree (in part I) and as a function of predictive information decay time (in part II). The results for part 1 are presented in Figure 4. Both the average correct response rates per degree and those for every graph are shown. The results for part 2 are displayed in Figure 5. Both the average correct response rates per graph and those per individual random walk are shown.

A post-experiment survey was conducted in the early stage of the pilot. After completing the experiment, participants were asked about perceived difficulty and about what strategy they relied on. Results of this survey can be seen here [10]. It is important to note that most of these responses were submitted when the experiment still had a major bug in part II, which made this part impossible to score well on.

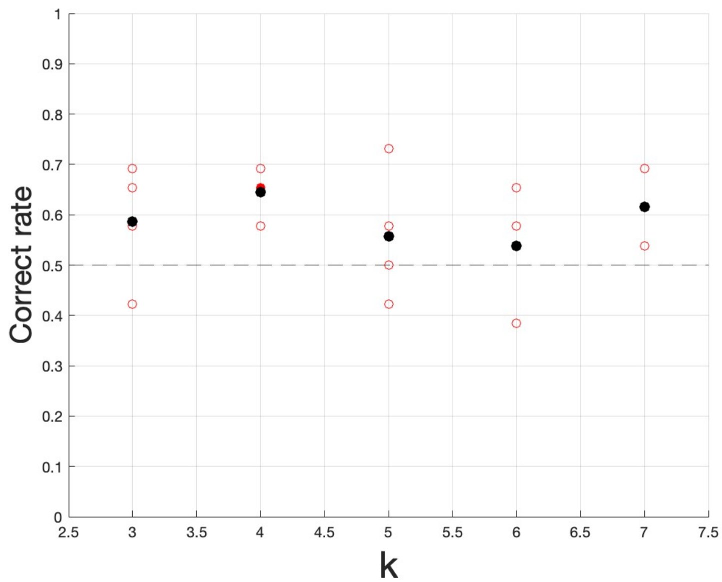


Figure 4: The fraction of correct responses in Part I of the experiment where the graphs varied in their degrees, from 3-regular to 7-regular graphs. The dashed line shows chance performance. Black dots show the average performance per degree and the colored dots show the performance per graph. Filled in colored dots show two overlapping data points.

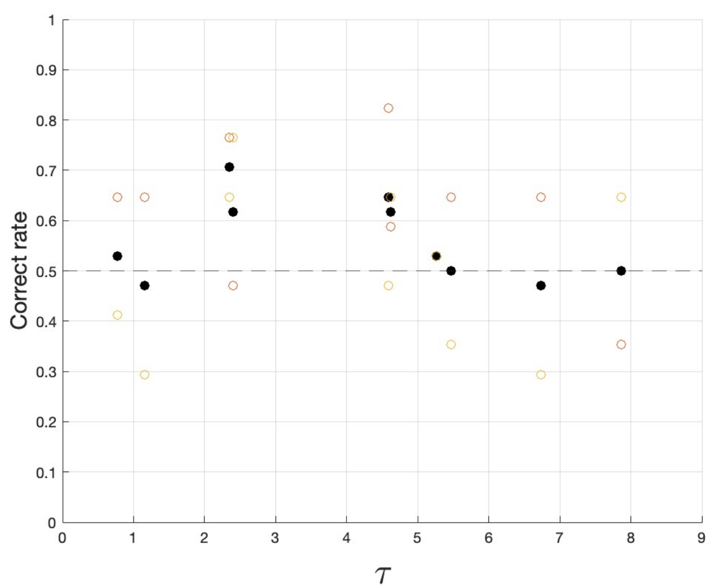


Figure 5: The fraction of correct responses in Part II of the experiment where the graphs varied in predictive information decay time, from the lowest to the highest. The dashed line shows chance performance. Black dots show the average performance per graph and colored dots show the average performance per random walk.

5 Conclusion

From the data generated in this pilot experiment, we are unable to draw conclusions on the relevance of predictive information decay rate for human pattern recognition performance. Participants performed only slightly better than chance throughout most of the experiment, making it hard to distinguish a signal. This seems to suggest that the task was too hard or too unclear.

In part 1, participants were systematically better able to perform above chance than in part 2, but the data do not show a dependence of performance on predictive information. This can be seen as evidence that predictive information alone is not very relevant for pattern recognition.

The results of part 2 show no clear dependency of performance on predictive information decay rate. This seems to suggest that predictive information decay rate does not provide a useful framework for studying pattern recognition. However, the amount of participants is fairly low and performance varied strongly per random walk, even for the same graph. For most of the task, participants were not able to perform better than chance. From the data, it is hard to conclude if the apparent peak in performance for moderate predictive information decay rate (2-5) is significant.

6 Discussion

This project served to create a pilot study for a potential further, more elaborate research. In this discussion, we will mainly focus on giving advice on how the experimental set-up could be improved and what factors to take into consideration when designing similar experimental paradigms. There are many minor aspect of this experiment that could be optimised, such as the stimulus size, the speed of the transitions, the length of the random walks and the graphs with which participants are asked to contrast the data. In this section, we will suggest major revisions that make this optimisation a lesser concern before those are addressed.

Predictive information: The remarks that will follow will mostly be relevant for part 2 of the experiment. The results of part 1 do seem to suggest that predictive information does not strongly influence pattern recognition performance. However, variance in performance varied widely for different graphs of the same degree. More data per degree might be needed to distinguish if there is any effect size. Comparison might be made more insightful by making a more motivated choice for exactly which k-regular graphs are chosen. For example, k-regular graphs very close to the median predictive information decay rate for graphs of that degree might be taken as representative, rather than the random samples used now. The fact that there seems to be little correlation between predictive information and performance is still somewhat surprising, as random walks on 3-regular graphs would move much more back and forth between a couple nodes than the 7-regular graphs. However, the randomly sampled networks used in this part of the experiment are all still very unstructured, which can explain why they all seem equally difficult.

Quality of the data: The current experimental set-up did not only include too little participants, the amount of information collected per participant was also limited. As for the second part, people only gave two answers per graph architecture and corresponding decay time. The random walks on these graph shown to participants were also all the same, instead of generated differently per participant. This allows for significant noise to enter the data, where a particular random walk was particularly hard to judge or evaluate, or it was coincidentally very similar to the random walk on the lattice-like graph that followed. More variations and more data points per graph are needed to draw conclusions to negate these effects.

If we look at how often participants answered with ‘Different’ per graph in part 2, we can see evidence that the exact random walk in the red part of the task can already strongly steer the decisions taken by the participants. For one of the random walks, participants only responded with 'Different' in 20 percent of the times, and in others with 80 percent, see Figure 6. As the correct answer was randomly generated, 'Different' was the correct answer around 50 percent of the time. It might be the case that for that graph, the continuation of the random walk on that graph in the blue phase (when the correct answer should be 'Similar') was very different from the red phase, because the pattern jumped to another cluster at around the same time as the pattern changed color. These effects could be accounted for through a more diverse dataset, but also suggest that the task to contrast with a lattice-like graph is too hard and does not accurately test pattern recognition skills.

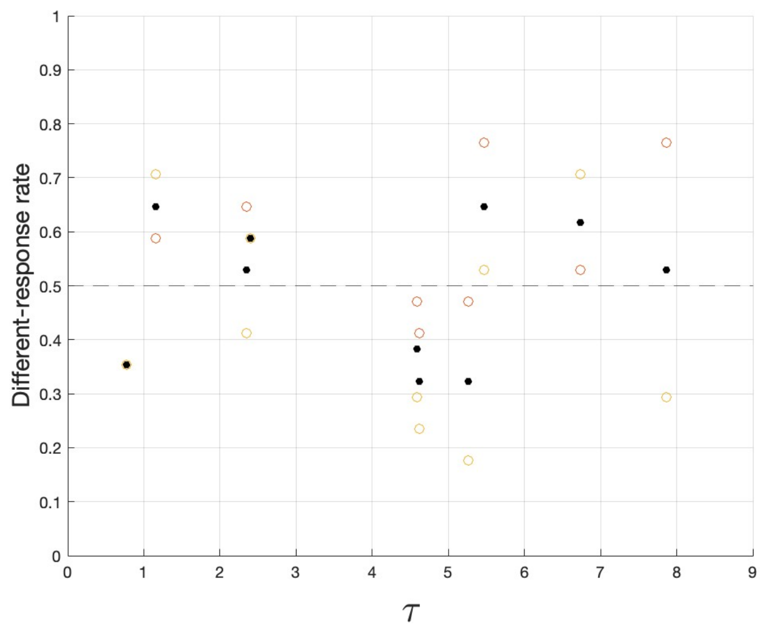


Figure 6: The amount of times participants answered 'Different' to the prompt whether the red and blue data were similar or different, per graph. Black dots are averaged over the two random walks over the same graph, colored points are the individual rates.

Difficulty of the task: In its current set-up, the task is quite difficult and hard to remain focused on. Subjects only performed slightly better than chance overall. More importantly, staying focused on the stimuli throughout the experiment is difficult and it is easy to get distracted, as participants stated in the post-experiment survey. After a small lapse of focus, it is difficult to get back into the task and performing it is generally not very motivating. This might make it difficult to make the task longer when done voluntarily, as suggested by the first point. Arguably, the large number (14) of different states presented is a key driver of the task's difficulty. In k-regular graphs with less nodes, however, predictive information decay rate isn't as widely varying, making it harder to provide contrast in the provided data. Stepping away from undirected k-regular graphs to less regular network architectures might allow for a wide variety of decay rates with less nodes. Likely, other adjustments should also be made to make the task more manageable and more motivating.

Explicit vs implicit approach: Pattern recognition is generally a more implicit skill, a quick evaluation. To test the relevance of predictive information decay in pattern recognition, we aim to query participants more on their implicit evalution of the pattern than giving them a particular task they need to solve by using reasoning and memory. Generally, subjects reported in post-experiment surveys and informal interviews that they tried to find tricks to spot a pattern change based on explicit evaluation. The format of this task, the tutorial and the tutorial instructions should be evaluated to promote implicit judgement.

Spatial patterns One important factor driving the strategy used by participants is the format in which the data is presented. A circular format was used to try and minimise the urge to look for spatial patterns in the data. However, when asked what strategy participants used, they often reported looking for 'how much the dot jumped around the circle', 'how many dots it skipped', 'if the dot went in a circular motion clockwise or anti-clockwise'. Ultimately, the spatial location of the nodes is completely unrelated to the underlying network, as the position of nodes on the circle could be switched. Hence, the tendency of participants to interpret the data in a spatial way and look for spatial patterns rather than purely transitions potentially drives them to perform the task in a certain, likely more explicit way. The format of the data presentation should likely be revised, either by changing modalities or changing towards number or symbol-based data presentation.

Sound based task: A potentially promising way to alter this experiment is by replacing the visual with an auditory stimulus. This might help prevent subjects from trying to perform the task explicitly and help take a more intuitive approach, as we normally do when listening to 'melodies'. It is also easier to display auditory data with a high variety without increasing the complexity of the task. Playing 14 different frequencies is easier than displaying 14 different visual stimuli while keeping all possible states easier to discriminate. A potential problem with this is that just like subjects might try to seek spatial regularities in the visual data, they might look for other spurious relationships in the auditory data, such as harmony or closeness in pitch.

Sensitivity to phrasing: As test subjects do not have insight in the underlying data generating mechanism, the differences between the trials and the theoretical background, it is difficult to instruct participants well. Many words, such as *pattern* and *probability* will have a certain connotation that will prime participants to do the task in different ways. The best way to solve this difficulty is probably to make a more extensive tutorial, such that the text in the beginning of the experiment can be limited to a minimum and interpretations of phrases will not matter as much. The tutorial text and the hints given (see Appendix A) should be and tested in trials before used widely.

Response time as metric: In earlier studies on predictive information by Bassett et al. (2020), response time based tasks were used instead of forced choice tasks used here. This allows for lots more data to be generated during the experiment. A downside is that it might make the task even more overwhelming, that the data is more noisy and that it only probes prediction for one state ahead, rather than more long term patterns. Still, it might be possible that there are superior alternatives to the current forced choice tasks, especially by making it possible for participants to look at the data for longer if they wanted and coupling a scoring mechanism to how fast they find the right answer.

A Tutorial Texts

A.1 Before Tutorial 1

"In this experiment, you will look at 14 circles. One after another, they will light up, in a chance-driven, but not completely random way. Your goal for this task is to recognize regularities in the transitions.

The data will be presented to you in two phases, a 'red' and a 'blue' phase. As the lit up circles turn from red to blue, you wil have to determine if

(i) the chances to go from one circle to any other have stayed the same

OR

(ii) the chances to go from one circle to any other have changed.

Your task is not to see if the subsequent blue and red sequence are *literally* the same. So when you look at the data, do not try to remember it, but just look for regularities.

You will now be presented with a trial run which should be relatively easy. For the remainder of this experiment, you will only need to use the 'D', 'S' and 'R' keys. It might help to press the 'S' and 'D' keys with different hands, if this is comfortable. Press 'R' to continue."

A.2 Before Tutorial 2

"I hope you saw that the pattern changed when the colours switched from red to blue! During the red part, the dots stayed in a bit of a 'cluster', as only a small part of all the dots were lit up and some transitions were repeated. During the blue part, the pattern became much more random, jumping all over the screen.

Note that there is no perfect way to assess this task! Since the underlying data is chance-driven, your answer will be a guess.

Let's try another one, and then we will start with the real experiment.

Press 'R' if you are ready."

A.3 End of the Tutorial

"I hope you saw that the pattern remained similar when it switched from red to blue! Some parts of the pattern repeated itself and the dot jumped around in a somewhat similar fashion.

The rest of the tasks will be more subtle, so you might make many mistakes. In about 50% of the cases, the pattern stays the same and in the other 50% it changes. This task will require a lot of focus, but do not try and memorize all you see. It is fine to give your answer based on subjective judgement.

Tip: you could think of this task as looking at a school class playing 'catch'. Any kid could throw the ball at any other kid, but in reality, some classes are not so egalitarian. Every kid has a couple favorite friends among which they will pass the ball, leading to regularities. You will look at school classes with a varying degree of 'in-groupness', and your task is to find out if you are looking at footage from one class, or footage from two classes with totally different children.

We will reset your score. Press 'R' if you are ready to start the real experiment!"

References

1. Voss, J. L. & Paller, K. A. Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. *Neuropsychologia* **46,** 3021–3029 (2008).
2. Schacter, D. L. Understanding implicit memory: A cognitive neuroscience approach. *American psychologist* **47,** 559 (1992).
3. Knowlton, B. J. The relationship between remembering and knowing: A cognitive neuroscience perspective. *Acta psychologica* **98,** 253–265 (1998).
4. Srivastava, N. & Vul, E. A simple model of recognition and recall memory. *Advances in Neural Information Processing Systems* **30** (2017).
5. Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M. & Barense, M. D. Integration and differentiation of hippocampal memory traces. *Neuroscience & Biobehavioral Reviews* **118,** 196–208 (2020).
6. Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal* **27,** 379–423 (1948).
7. Cover, T. M. & Thomas, J. A. Elements of information theory second edition solutions to problems. *Internet Access,* 19–20 (2006).
8. Lynn, C. W., Papadopoulos, L., Kahn, A. E. & Bassett, D. S. Human information processing in complex networks. *Nature Physics* **16,** 965–973 (2020).
9. Bialek, W. & Tishby, N. Predictive information. *arXiv preprint cond-mat/9902341* (1999).
10. Klop, M. https://airtable.com/apppXM0iTdIWpT2Ea/shrD1t192JLrxIZa6.

Layman summary:

This study focuses on understanding how humans memorize and recognize patterns in sequential data. The process of recognition is intricate and depends on largely unconscious and fast judgements. For instance, recognizing a familiar melody without hearing it in its entirety suggests that humans store information in a shorthand manner. This recognition might depend on an internal model that compares a stimulus for familiarity. However, the specifics of how humans effectively recognize patterns remain unclear.

To explore these questions, the study turns to information theory, which provides a scientific framework for analyzing information and communication. It quantifies the amount of information transmitted through a channel by analyzing it in the context of a chance driven signal-generating process following a particular distribution. Information theory depends on establishing a sender, receiver, information channel, and syntax, which states all possible messages that could be sent from the sender to the receiver.

In this study, data takes the form of a random walk on a network with fixed transition probabilities between its discrete states. These random walks are Markovian, meaning the information observable at the current time constitutes the full past of the system, which fully characterizes the probability of transitioning to any possible next state. This setting is simple but allows for specific manipulation of information theoretical features of the data.

The study aims to investigate the usefulness of Predictive Information Decay in understanding human performance in information processing and pattern recognition. Predictive information is the amount of information that the past yields about the future in an ongoing data sequence. The study will extend this to Predictive Information *Decay* by examining differences in predictive information that the states of the random walks provide about states further in the future, instead of looking only one step ahead. The motivation for this is the assumption that when humans try to recognize patterns in data, they look for longer reoccurring trajectories, not only at frequent transitions from one state to another.

The study is a pilot for an experiment that could lead to a full-scale study, providing insight and advice on how to proceed. The preliminary results suggest a major revision of the experimental set-up and corresponding hypothesis is needed, as little signal of a correspondence between predictive information decay and performance on recognition tasks is found. Ultimately, finding links between information theoretical concepts and human performance on pattern recognition tasks could help develop models of human recognition memory that can make falsifiable predictions. This report also discusses some of the difficulties of designing a psychophysics experiment to verify a certain model of human cognitive capacities.