Enhancing Learning via Novelty Insertion

MITili Learning Effectiveness Research Grant

Final Report

Matt Groth, Research Associate

Annie Cardinaux, Research Specialist

Anna Musser, Technical Associate

Qi Ding, Sohini Kar, Jenny Zhao UROPs

Lara Cavinato, Research Affiliate

Sidney Diamond, Research Affiliate

Riccardo Barbieri, Research Affiliate

Pawan Sinha, Professor, PI

Abstract

Research in the neural basis of learning and plasticity indicates that environmental novelty can play a role in modulating synaptic malleability in specific regions of the adult brain. This finding has potentially important implications for the domain of pedagogy. Might the inclusion of novelty in teaching material enhance neural plasticity and thereby improve information uptake?

We have explored this question by systematically studying the impact of 'novelty insertion' in a lecture video on comprehension and physiological indicators of arousal such as electrocardiogram/heart rate (ECG), and electrodermal activity (EDA). During the study, we recorded participants' physiological data while they watched an introductory psychology lecture on video. This lecture was augmented to include short video clips that either contained novel or familiar content. A comprehension assessment evaluated whether participants possessed a better command of the material that followed "novel" videos compared to material that followed "familiar" videos. Additionally, physiological data were analyzed to determine if

videos with novelty insertions resulted in higher engagement or arousal as indicated by the level of sympathetic or parasympathetic activation.

Encouragingly, while our behavioral results do not quite cross statistical significance thresholds for demonstrating a differential increase in learning following exposure to novel rather than familiar stimuli, they do indicate a trend towards such a pattern. We found that physiological responses to familiar and novel content were highly variable. Hence, no definitive conclusions can be made concerning novelty's impact on engagement or arousal. Neither the sympathetic nor parasympathetic nervous system showed differential activation for novel and familiar videos. We discuss caveats and potential protocol changes if this line of investigation were to be pursued further.

While suggesting a possible linkage between novelty and learning in pedagogy, this project has also had a beneficial secondary outcome in the form of tools that can prove useful for other researchers interested in undertaking experimental studies of learning and comprehension with instructional videos.

Background

Novelty may be defined simply as a characteristic of stimuli that lack pre-existing neural representations. Such stimuli are of great interest in studies of brain function in health and disease. Exposure to a novel stimulus can result in a substantial neural response, distinct from that which occurs when a familiar stimulus is encountered. Novelty processing shares a common pathway with the reward system which may reflect a functional link in which novelty motivates exploration in search of reward (Bunzeck et al. 2012). Differential effects of familiarity and novelty have been observed in distinct parts of the human brain, with familiarity-based activations being observed in the dorsolateral and superior medial cortex, mediodorsal thalamus, anterior cingulate and the left angular gyrus, and novelty activating regions such as the fusiform gyrus, primary and secondary occipital cortex, the PRC and the anterior hippocampus. A key brain structure involved in learning and memory, the anterior hippocampus, communicates novelty detection to the striatum and the dopaminergic midbrain, resulting in dopamine release, which then affects learning (Kafkas A, Montaldi D, 2018). Part of the midbrain, the substantia nigra/ventral tegmental area (SN/VTA), respond strongly and selectively to novel stimuli ((Schott et al. 2004), (Wittmann et al. 2005)). Projections from the hippocampus to the SN/VTA via the subiculum are necessary for this modulation of SN/VTA dopamine release, as evidenced by studies where the release of dopamine is blocked by inactivating the subiculum (Lisman, Grace, and Duzel 2011). It is this hippocampal release of dopamine and norepinephrine that enables learning and the integration of new information. Thus, research has revealed compelling novelty-dependent modulation of the mesolimbic dopaminergic system ((Reed, Mitchell, and Nokes 1996)). Such dopaminergic neuromodulation

has been found to be capable of inducing synaptic plasticity in the hippocampus, which is conducive to learning ((Frey and Morris 1998)).

Importantly, the impact of novelty on learning enhancement transcends the novel experience itself; research has found that novelty-induced reinforcement of long-term potentiation can persist for 30 minutes after the novel experience has ended (Straube et al. 2003). This means that the neuro-modulatory influence of novelty on hippocampal synaptic plasticity is not confined just to the period of the novel experience, but rather transcends it, lasting even after the novel stimulus has been extinguished ((Li et al. 2003); (Straube et al. 2003)). These results have far-reaching implications for pedagogy and form the basis of the project we undertook.

The observation that experiential novelty leads to dopaminergic activation and hippocampal plasticity implies the possibility that the presence of novel stimuli can heighten learning and memory. This leads to the prediction that interspersing a stimulus set with novel items can lead to an overall increase in the comprehension of, and the motivation to learn, the stimulus information. This may sound somewhat counter-intuitive given the increased learning load and potential for distraction due to the novel stimuli, but empirical data suggest that this prediction is worth pursuing (Bunzeck and Düzel 2006). This points to a simple approach for potentially enhancing the effectiveness of pedagogical material, namely, interspersing novel items unrelated to the main topic into the instructional material. We call this approach 'Novelty Insertion'.

Specific Aims

The primary aim of this project was to determine whether a novel stimulus could enhance the learning of lecture material. To this end, video clips with novel or familiar content were inserted into a pre-recorded lecture. Subjects were then tested to determine whether they possessed better comprehension of the lecture material that followed 'novel' video clips when compared to material that followed 'familiar' video clips. Additionally, the effect that novel stimuli had on participants' engagement was explored by examining markers of arousal through the use of electrodermal activity (EDA) and electrocardiogram/heart rate sensors. By measuring both learning performance and physiological arousal, we sought to determine correlations, if any, between arousal and novelty enhanced learning.

Study Design

Participants

Healthy adults between 18 and 35 years old were recruited to participate in this study. This experiment assessed the extent to which one learned basic neuroscience knowledge, thus any

individuals who had taken more than one neuroscience or psychology course were excluded. Additionally, due to the academic caliber of MIT students, they too were excluded from the study so that the result could have greater generalizability to the population at large. With these criteria, we enlisted 47 participants over the duration of the project. The study was conducted at the Martinos Imaging Center of the McGovern Institute for Brain Research. Participants were escorted to the EEG lab where they received an informed consent document to review and sign.

Physiological Sensors

The physiological recording modalities used for this study included Electrocardiogram / heart rate (ECG), and electrodermal activity (EDA). The system we used was a BioSemi ActiveTwo (Fs = 2048 Hz). EDA was recorded by placing two electrodes on the palm of the hand and measuring the conductance between these electrodes. This conductance varies as a function of the subject's perspiration. Such data provides us valuable information concerning the sympathetic nervous system, for it is this system that controls activity of the sweat glands. ECG was recorded using three electrodes (in the Einthoven's triangle configuration) placed on the subject's torso. These sensors measured the fluctuations in the electric field that occur due to the circulatory system. We used this information to calculate the subject's heart rate variability, which is a measure of the fluctuation in the intervals between heartbeats. It is predominantly controlled by the vagus nerve, which interfaces with the parasympathetic nervous system.

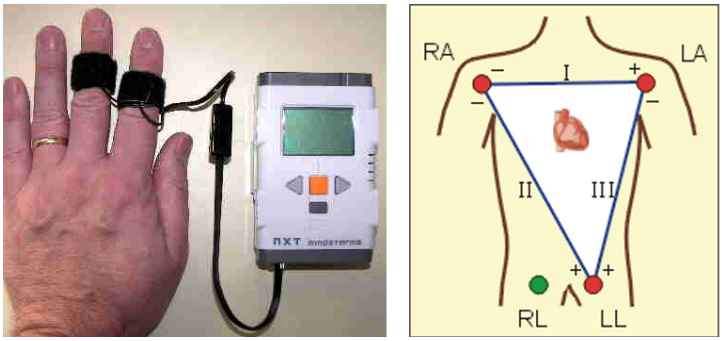

Figure 2. *The two kinds of physiological sensors we used in this study. (Left panel) Placement of GSR sensors. (Right panel) placement of ECG sensors in the Einthoven triangle configuration.*

Video stimulus

We prepared a 70 minutes long video. This consisted of several elements, the first of which was a five-minute recording of bubbles floating against a black background. This value-neutral clip appeared twice, at the start and at the halfway mark of the video. The purpose of this clip was to induce calmness as a precursor to determining participants' baseline measurement of arousal. The bubble clip was followed by the first lecture clip. There were six lecture clips in total, with each lecture clip lasting approximately 8 minutes. The lecture that was selected for this experiment was "Brain 1: Structure and functions", derived from the course 'Introduction to Psychology' by Professor John Gabrieli. Between the lecture clips, we inserted 90 seconds long video clips containing familiar or novel content. Familiar clips showed a woman cleaning a

typical American kitchen, while novel clips presented footage of exotic animals. At the end of the final lecture clip a two minutes long high-arousal clip was shown, which presented a first-person perspective of peering off the edge of a skyscraper to the ground below. The purpose of the rooftop clip was to act as a metric of maximum arousal, to be used in range-corrected indexing (Braithwaite et al. 2013). Figure 3 shows thumbnails from the video clips. All transitions between clips were 5 seconds long and consisted of a fade into or out of black.

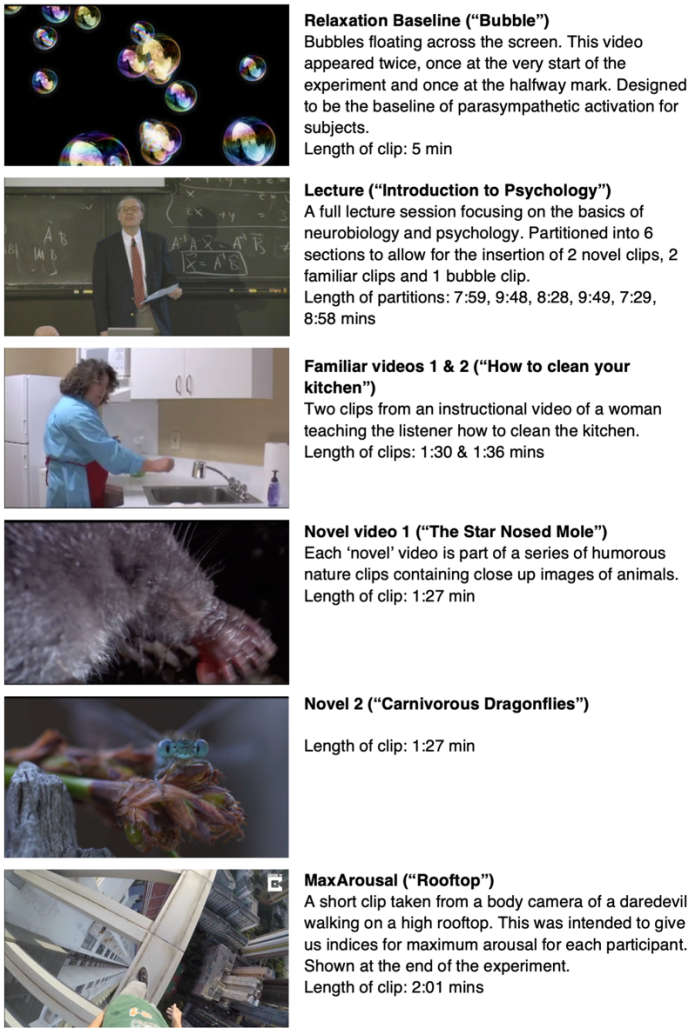

Figure 3. *The different videos we used to create the composite clip for the study.*

Instructions

Participants were informed that they would be watching an introduction to neuroscience lecture and that their primary responsibility was to learn as much as they could from this lecture. They were also informed that other video clips had been inserted into the lecture and they would not be responsible for retaining any information that non-lecture video clips presented.

Conditions

Participants were assigned to one of two conditions, schematized in figure 4. Within condition one, familiar video clips were inserted in the first half of the video (between lecture clips 1 and 2, and lecture clips 2 and 3), while novel video clips were inserted in the second half of the video (between lecture clips 4 and 5, and lecture clips 5 and 6). In condition two, novel clips were inserted in the first half of the video and familiar clips were inserted in the second half. We counterbalanced the positions of familiar and novel video clips through the creation of these two conditions to ensure that any observed difference in lecture comprehension was due to the effects of the novel and familiar clips, rather than any intrinsic differences between the first and second halves of the lecture. Both conditions included the five-minute bubble clip at the start of the video and then again at the midway point between lecture clips 3 and 4. Similarly, both conditions concluded with the rooftop video clip.

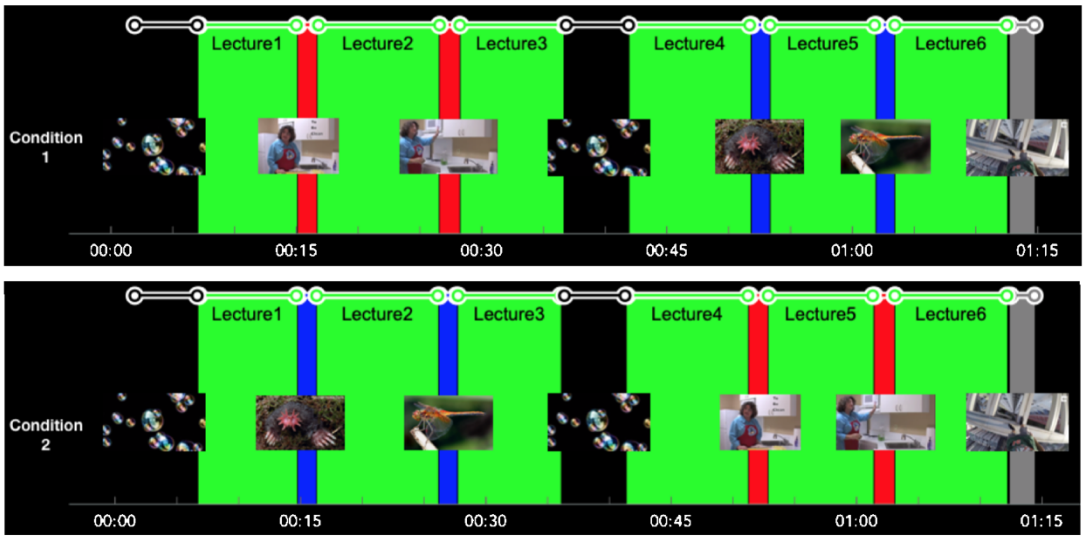

Figure 4. *The two conditions to which participants were assigned. They differed in the order of insertion of the familiar and novel video clips. Condition 1 had familiar insertions precede the novel ones, while condition 2 had the reverse order.*

Behavioral pre-assessment

Once informed consent was completed, participants took a laptop-based pre-assessment test. This comprised 30 questions regarding basic neuroscience knowledge. The purpose of the pre-assessment was to determine each participant's baseline knowledge of neuroscience.

Behavioral post-assessment

Sensors were removed once participants finished watching the video. Subjects then completed a post-assessment which tested their comprehension of the lecture. This post-assessment consisted of 110 questions, all of which pertained to the lecture. This assessment was carefully created by the research team to ensure that we tested as much information as possible from each of the lecture sections.

Results

Pilot Sessions

We conducted a pilot study on 10 participants (5 per group). The purpose of the pilot was to evaluate the experimental design and make adjustments as needed. Thus, there were many differences between the pilot and final iteration of the experiment. For instance, the pilot experiment involved a break in between the first and second half of the experiment, only included 8 pretest and 8 posttest questions (one question per lecture section), a different novelty video was used which explored carnivorous plants, and while this experiment used the same course (Introduction to Psychology (9.00SC)), a different lecture was used (Day 5, Vision 1). Perhaps the greatest difference between the pilot and final experiment was that subjects in the pilot were asked to consume caffeine (a cup of coffee) before the experiment in a bid to boost alertness. Regrettably, however, we found that caffeine interfered with the heart rate data and therefore its use was discontinued in the final experiment.

Comprehension Results

We enlisted 37 participants in our post-pilot study. Of these, 12 individuals fell asleep during the sessions and hence had to be excluded from consideration. Of the remaining 25 subjects, 22 scored a higher post- than pre- score (the covariance of pre- to post- is 0.009 and the Pearson's correlation coefficient is 0.66).

The full post-assessment contains 110 multiple choice questions, but for various reasons some may be better indices of learning than others. We considered three kinds of 'inclusion filters' to restrict the questions for which to analyze responses. These were:

Ceiling/Floor Filter: Questions that have an average score $20\% < x < 80\%$ across all subjects.

On-slide Filter: Questions derived from information that was included on the slides.

Time Filter: Questions pertaining to information presented within 450 seconds after end of insertion.

These results reveal that in every case, the ratio of participants in which Novelty score > Familiarity score is always marginally greater than 0.5. With all three inclusion filters in use, we find that the p-value for the novelty versus familiarity scores (the former being larger than the latter) in the second half of the lecture achieves statistical significance ($p = 0.042$). Also, the time-filter on its own yields statistically significant results (novelty scores > familiarity scores) for the entire lecture session ($p = 0.0419$), and almost so for the second half of the lecture ($p = 0.052$). Conducting this analysis for various time durations after insertion end yields data shown in figure 5. The results suggest that if the novelty insertions are indeed causing an increase in learning, the effect is greatest a short time (within 8 minutes) after the insertion ends.

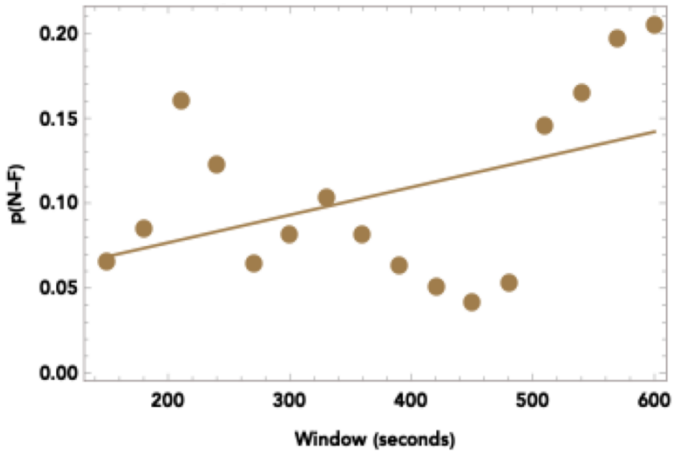

Figure 5. *The p-value of single-tailed, paired t-tests between familiar and novel scores across the entire lecture session, as function of the window of time after each insertion from which assessment questions could be derived.*

Physiological Results

GSR signals were preprocessed using continuous decomposition analysis in Ledalab. The tonic and phasic components of the signal were decomposed, and then peaks indicative of skin-conductance responses (SCRs) were detected and counted along the phasic component using a 0.01 micro siemen threshold. Metrics derived from this include non-specific (NS)-SCR rate, NS-SCR Amplitude, and NS-SCR standard deviation (STD).

In the HRV signals, QRS complexes were detected using the pan tompkins detection algorithm, and RR intervals were calculated. Next, heart rate variability (HRV) indices were calculated using a Point Process model of HRV developed in Riccardo Barbieri's lab.

In order to reduce between-subject effects due to variations in absolute measures of GSR and HRV metrics, we calculated range-corrected indices. In preparation for calculating range-corrected indices, minimum-activation and maximum-activation phases were designated. maximum-activation phases included the bubble videos and the first lecture phase. The physiological results from the bubble phases do not follow the same trend for all subjects however, and we have found that some subjects showed increased sympathetic activation in such videos. For the maximum-activation baseline, the rooftop video proved to be reliably higher for most subjects in both sympathetic and parasympathetic measures. (This direct correlation of sympathetic and parasympathetic arousal is unusual, but we think it may be due to a vertigo or freeze response that participants feel as they are watching the first-person view of the person in the video dangling off the rooftop.)

Of the 24 potential participants, data from 18 were usable. Five subjects were 'GSR-non-responders' (did not exhibit skin conductance response (SCR) in one or more of the relevant stimuli) were removed from physiological analyses. One participant's data was also considered

an outlier because their overall average GSR peak rate and high frequency HRV was over two standard deviations away from the group mean.

Figure 6 summarizes our physiology data. Interestingly, and contrary to our expectations, the 'True Facts' (novel) videos did not elicit significantly different physiological responses from the 'Cleaning the Kitchen' (familiar) videos. A single-tailed paired t-test of the hypothesis that sympathetic activation as measured by the GSR peak rate was higher during the novel videos was just shy of statistical significance ($p = 0.0503$, $n = 18$, std = 0.074). A single-tailed paired t-test of the hypothesis that the parasympathetic activation as measured by the high-frequency HRV power was higher in the familiar videos also did not disprove the null hypothesis ($p = 0.4123$, $n = 18$, std = 32.7).

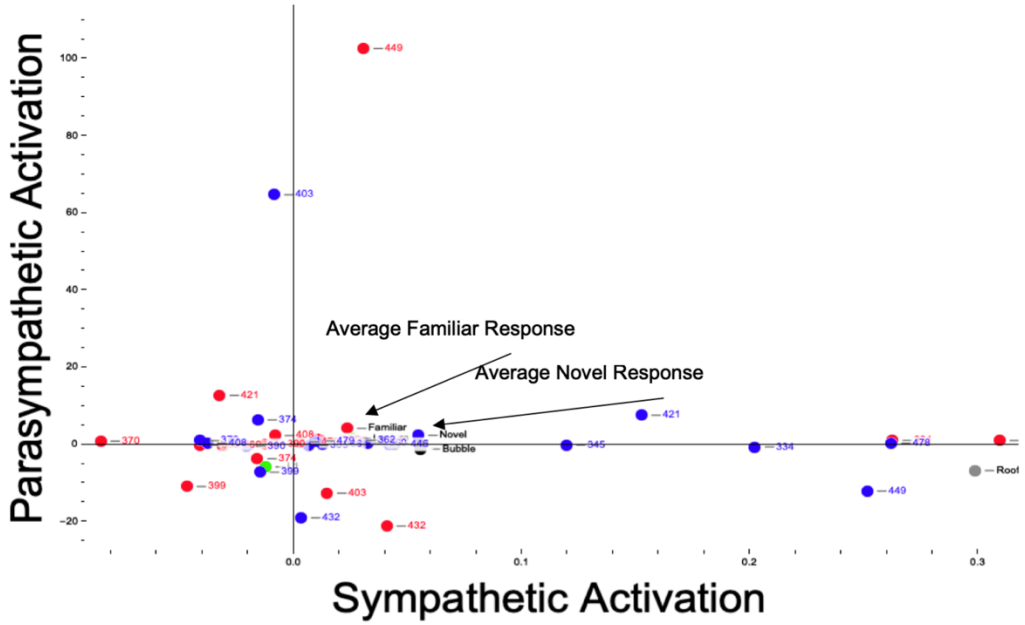

Figure 6. *The average and individual sympathetic and parasympathetic arousal indices for 18 participants. The novel insertions (blue) trend towards having increased sympathetic activation (p=0.0503) and do not have decreased parasympathetic activation (p=0.4123) relative to the familiar (red) insertions. It is also notable that there is high variation between subjects.*

Discussion

Our behavioral analyses show a trend towards an enhancement of learning due to inserted 'novel' stimuli. The 'Time-filter' yields a significant difference in learning following the familiar versus novel insertions. A similar outcome is observed will all of the inclusion filters considered together. Past research suggests that the neural consequences of novelty-exposure last 15-30 minutes, but our results indicate that 'novelty-enhanced' learning may have a more compressed timeline of roughly 8 minutes.

Physiological data were intriguingly less discriminative. Using metrics derived from GSR and ECG, we found no significant difference in arousal levels during the familiar or novel stimuli.

The sympathetic and parasympathetic nervous systems usually are mutually inhibitory, meaning that typically, one activates when the other deactivates. However, this is not always the case and we have found specific stimuli that can cause both systems to activate in many subjects. This argues against the prospects of developing a unidimensional overall arousal/engagement index; a multi-dimensional metric for arousal that incorporates both sympathetic and parasympathetic measures is an important goal for future research.

It is important to consider some of the caveats associated with this work. Although they limit the ability to draw scientific conclusions from our data, they serve as important points of information for future study designs.

Subject enrollment

Initially, this study was designed so that subjects could participate from the comfort of their own dorm rooms. Students were to be given smart-phone compatible skin-conductance and heart-rate sensor boxes to allow the recording of Galvanic Skin Response and heart rate data while they watched the modified lecture video. This design was appealing in that it would have allowed multiple participants to complete the experiment simultaneously while eliminating the need for researchers to test each sequentially. Regrettably, the sensor box proved unreliable, producing signals of poor quality due to a low sample rate and susceptibility to electrical noise. Consequently, the EEG lab at the Martinos Imaging Center became the new home of this experiment. In such an environment, high-quality physiological data could be collected, but four hours and two researchers were now needed to run a single subject. This reduced the number of subjects we were able to run during the course of this experiment and the attendant statistical power of the data collected.

Subject drowsiness

A pilot study of this experiment was run with ten subjects whereby participants had coffee prior to watching the introductory psychology lecture with familiar and novel clips inserted. This study proved useful for it clearly illustrated that while caffeine did boost alertness, it also had a dramatic impact on the electrocardiogram/heart rate signals, compromising their interpretability. As a result, caffeine was eliminated from the subsequent experiment. Unfortunately, without caffeine many subjects fell asleep while watching the lecture video, rendering their data unusable. Specifically, of the 37 participants run in the second version of the experiment, 12 became too drowsy to complete the experiment. To counter this issue, additional light was added to the EEG room and a button press task was implemented. This task occurred once after each familiar and novel video clip and required participants to press a button once if the letter 'A' appeared on the screen and twice if the letter 'B' appeared. The purpose of the button press task was to help subjects stay awake for the duration of the video. These augmentations mitigated, but did not eliminate, subject drowsiness.

Lecture Variability

The lecture we used has highly variable content, with the first half of the lecture employing a different teaching style than the second. For instance, the first half of the lecture presents a higher amount of factual information, while the second half uses story-telling to take a detailed

look at a few case studies. Such differences could impact the difficulty level of questions, the participant's ability to engage with the material, and their physiological arousal. These variabilities of content can be accommodated., and fruitfully used, if the participant pool is large. A large pool will provide enough statistical power to analyze the first half of the lecture and the second half of the lecture separately. Doing so could answer additional interesting questions like: Is novelty insertion more effective when presenting dry factual information when compared to information that is presented through stories?

Choice of Novel and Familiar Stimuli

Our novel stimuli, 'True Facts' videos, and our familiar stimuli, 'Cleaning your Kitchen' videos, both contain elements of novelty. While it can be assumed that most adults would be familiar with kitchens and how to clean them, watching a tutorial on how to dust may in itself be a novel experience. Additionally, the woman who hosts the 'Cleaning your Kitchen' videos is not a celebrity and therefore would be unknown to the subjects that watched her. Research on facial recognition has indicated that familiar and novel faces are processed in different regions of the brain and thus including an unfamiliar actress in our familiar stimuli may well have inadvertently made them appear novel, lessening the induced differences on comprehension and physiological signals.

Scope of assessment battery

Our study was limited to exploring the short-term impact of novelty insertion. The results with the time-filter indicate that the impact of novel experiences may last for a narrow temporal window. However, it is unknown whether the consolidation of the information acquired during this window manifests later in time. Hence, studying the impact that novelty insertion on long-term memory may be a promising avenue of inquiry. Incorporating an additional comprehension assessment that would take place a few days after the experiment could help address this issue. It may also be advantageous to incorporate additional measurement modalities to enhance the ability to examine arousal and engagement mechanisms at play. Such mechanisms would also allow for both real-time and post-hoc assessment of subject drowsiness.

Notwithstanding these caveats, this study has the potential for its results to translate into pedagogy in online and conventional learning environments. The notion that novelty plays a role in learning and student engagement, while intuitively appealing, needs to be backed up with more scientific research of the kind we have undertaken here. This will not only help to raise awareness of the need for new protocols for effective instruction, it will also pave the way for linking insights from neuroscience to education.

Acknowledgments

We wish to express our gratitude to the MIT Initiative in Integrated Learning for having supported this research.

References

Braithwaite, Jason J., Derrick G. Watson, Robert Jones, and Mickey Rowe. 2013. "A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments." Psychophysiology 49 (1): 1017–34.

Bunzeck, Nico, Christian F. Doeller, Ray J. Dolan, and Emrah Duzel. 2012. "Contextual Interaction between Novelty and Reward Processing within the Mesolimbic System." Human Brain Mapping 33 (6): 1309–24.

Bunzeck, Nico, and Emrah Düzel. 2006. "Absolute Coding of Stimulus Novelty in the Human Substantia Nigra/VTA." Neuron 51 (3): 369–79.

Frey, Uwe, and Richard G. M. Morris. 1998. "Synaptic Tagging: Implications for Late Maintenance of Hippocampal Long- Term Potentiation." Vol. 21. https://doi.org/10.1016/S0166-2236(97)01189-2.

Li, Shaomin, William K. Cullen, Roger Anwyl, and Michael J. Rowan. 2003. "Dopamine-Dependent Facilitation of LTP Induction in Hippocampal CA1 by Exposure to Spatial Novelty." Nature Neuroscience 6 (5): 526–31.

Lisman, John, Anthony A. Grace, and Emrah Duzel. 2011. "A neoHebbian Framework for Episodic Memory; Role of Dopamine-Dependent Late LTP." Trends in Neurosciences 34 (10): 536–47.

Reed, Phil, Chris Mitchell, and Tristan Nokes. 1996. "Intrinsic Reinforcing Properties of Putatively Neutral Stimuli in an Instrumental Two-Lever Discrimination Task." Vol. 24. https://doi.org/10.3758/BF03198952.

Schott, Björn H., Daniela B. Sellner, Corinna J. Lauer, Reza Habib, Julietta U. Frey, Sebastian Guderian, Hans Jochen Heinze, and Emrah Düzel. 2004. "Activation of Midbrain Structures by Associative Novelty and the Formation of Explicit Memory in Humans." Learning and Memory 11 (4): 383–87.

Straube, Thomas, Volker Korz, Detlef Balschun, and Julietta Uta Frey. 2003. "Requirement of β-Adrenergic Receptor Activation and Protein Synthesis for LTP-Reinforcement by Novelty in Rat Dentate Gyrus." The Journal of Physiology 552 (3): 953–60.

Wittmann, Bianca C., Björn H. Schott, Sebastian Guderian, Julietta U. Frey, Hans Jochen Heinze, and Emrah Düzel. 2005. "Reward-Related fMRI Activation of Dopaminergic Midbrain Is Associated with Enhanced Hippocampus-Dependent Long-Term Memory Formation." Neuron 45 (3): 459–67.

Software Tools

In order to execute this study, we created a number of software tools that may be of use to the rest of the MIT research community.

The OCW Assessment Designer software tool could be useful for an instructor or researcher who wants to create an assessment closely tied to the material in a lecture video. The user sees a side-by-side view of the lecture video and a live transcript, which highlights the currently spoken words. The user may mark parts of the transcript as being question-worthy. Next, the user sees a list of all of the markings, and is prompted to write a multiple-choice question for each. The output of this tool is not only a list of multiple-choice questions, but a file mapping each question to the specific time point that the question is from. This mapping could be important for reasons such as determining specifically which parts of a lecture video a student was paying attention to. Other features include the possibility for questions to be derived from information at multiple time points and the ability to 'tag' questions with arbitrary values. For example, in our experiment, we tagged questions according to the amount of visual information they contained. This tool can help the user make an assessment that tests the maximum amount of information possible from a lecture video. It will work with any OCW lecture that provides a downloadable video and a .srt transcript file.

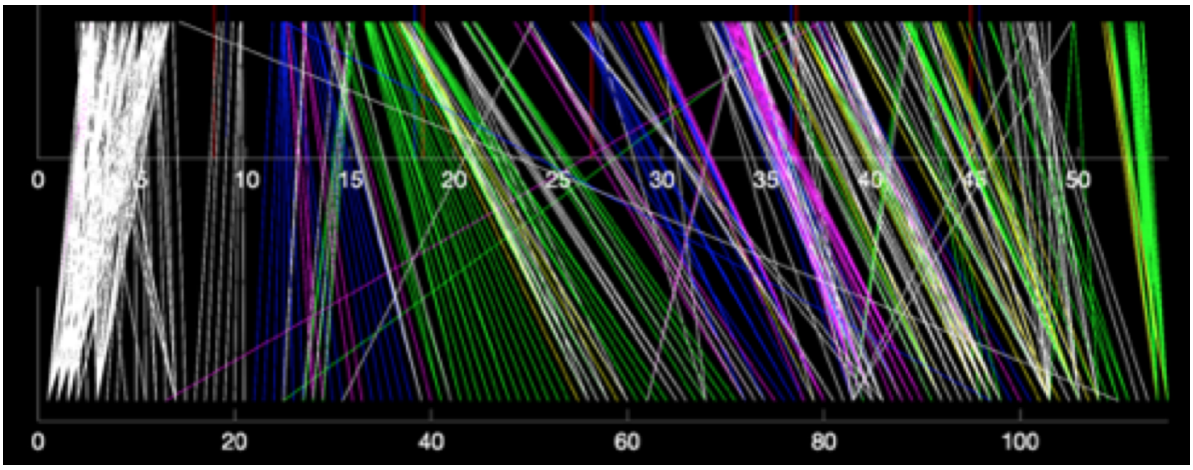

Figure 7. *In this visualization, we have a series of points plotted along two plots and lines connecting those points. The top plot is a timeline of the lecture, with the tick marks representing minutes. Each point on the top plot is the exact moment when the speaker gave information pertaining to questions used in the post-assessment. The bottom plot is simply a representation of each question plotted at x=1, x=2, etc. By looking at the lines coming out of a point on the bottom, one can see where the information for that question was derived from. Some questions are derived from multiple points. The color of the line represents the way in which this information was presented (see legend below).*

Color Key for Knowledge Point Plot (Figure 7)

None: The information was only spoken	White
Text Exact: The information was written	Blue
Graphic Exact: The information was displayed in a plot or diagram	Yellow
Text Extra: The information was related to something written	Magenta
Graphic Extra: The information was related to something displayed graphically	Green
Miscellaneous Visual: Miscellaneous visuals, for example the speaker crumpling a paper to discuss the cortex	Grey

The Data Collection software tool could be useful for researchers who wish to conduct remote behavioral experiments. This tool is comparable to existing online platforms such as Amazon Turk or survey-based websites. However, by virtue of being a java app made from scratch, it has increased flexibility and simplicity compared to those existing technologies. For example, it would be easy to extend this app to add a game or psychological task in addition to multiple-choice and short answer questions. This tool has the capability of uploading resulting data to Wolfram Cloud, but could feasibly be developed further to use other online databases. In our experiment, the participants used this tool to take the multiple-choice quizzes. It also allowed the researchers to note any observations with timestamps during recordings.

We have also developed data pipelines that analyze behavioral and physiological data and correlate the two. Finally, we have developed a Wolfram Cloud database for subject data that interfaces with the Data Collection tool and analysis pipelines.

We are happy to make these technologies open source and available to the public as needed.