Department of Psychology, University of California, Los Angeles

# Fall 2011

Table of Contents

**Improving Mathematics through Conceptual Learning**

**Britney Zavada**

Eastern and Western values offer two different approaches on mathematics and how to teach the subject. It has been widely known that Americans suffer the status of one of the lowest performers for mathematics. The question is how can we alleviate, alter, and enhance teaching structures to cater to this problem. Although there has been substantial research, no definitive answer has been made. One method that seems to help is through methods of comparing and contrasting problems that target the theory and deep structure.

Past studies have exemplified many processes that help promote efficient and accurate problem solving skills. When children in western education are approached with a completely novel problem, usually they will be completely baffled. Why is this? First, when students are taught a subject they are shown numerous examples; most of these are worked out, and then they must solve a problem set. Although this seems like the right approach, all it teaches students is how to mimic the process for other problems that have similarities. Eastern approaches focus more on the theory behind the problem, and go deeply into teaching only one or two problems. When these students are given a novel problem they can use the skills and analogies from the first problem and succeed. When students categorize problems they will judge them on their surface structure rather than the causal relationships. Instead, teaching should facilitate the process of seeking parallels and points of similarities between the problems. Relevant analogies are key the understanding of complex math problems. These analogies do not have to be related in the surface structure, but only their deep structure. (Needham & Begg, 1991)

These effects are even more apparent with community college students who struggle to get their degree because of the mandatory math classes. Many students can get through high school without the understanding of numerous math concepts that are then encountered in college. One study emphasizes the importance of proofs that are usually avoided in all educational studies because the formalities are believed to be a waste of time to learn. Opposite to this belief, if students are able to understand the idea of a general proof, then the proofs of a specific problem illuminate the apparent deeper structure. Educational systems need to begin by making students familiar and masters of proofs with fundamental problems that then can be built up to complex ideas. When proofs are presented in textbooks they are often presented in a cryptic and unclear format that makes the process even more daunting to students. Instead, these proofs should indicate the facts that validate the theorem, serve as a model for other proofs, and show the developmental steps of the proof. (Gries & Schneider)

Another study emphasizes on the language and manner that mathematics is taught in the classroom from kindergarten till senior year of high school. Often it is hard for one teacher or course to change the effects of teaching structure from past years. A complete revolution of how mathematics is taught is necessary to get the desired results and to be a competitive rank against other countries. The utilization of the terminology gives rise to the meaning and transforms jargon into meaningful representations of the complex problem. Teachers must allow time and analysis for student to build problem solving skills. Symbols and terminology have to be unmystified so that students can access their underlying purpose and use them accordingly. (Rothman, Rosalind, Cohen, Jill)

The research that I had focused on last year sought to gain answers through compare and contrast tests that forced students to recognize the underlying structure of the problem. Even though this was not explicitly said to the participants, they had to answer open ended questions that would provoke this kind of thinking. Also, the test did not have numerous problems, only a few problems that assessed different aspects of the overlying subject, factoring. If students got the first answer correct then they would be able to skip over the comparisons and go to the next problem. The results from the study showed however that the comparisons gave an enormous increase in the post test score. Other similar studies had been performed but not will this kind of effect. This promising evidence provoked further research to see if these results can be applied to the classroom and the computer.

This year I focused on the transition of taking the paper version of the study and formatting it for the computer. The main difficulty is making sure that the test is just as effective and the effects do not get lost within the interface of the computer. Although I have computer programming experience, the goal was to create a website utilizing software that would allow all researchers to manipulate and make changes without a computer programmer. Immediately there were struggles finding software that would cater to our needs dynamically taking in an answer from the participant. The first problem regarded the environment that we tested the trial on. The majority of quiz software will only run on Windows machines, which made us change the location of the testing. We also had to explore our options for software that would allow us to manipulate the feedback that was given and a way for participants to see their past work. I learned the importance of piloting and continuing to make changes to the quizzes as I made them. In order to allow participants to input any answer to a question, our quiz mostly consists of fill in the blanks with a few multiple choice. The reason for the multiple choice questions was because it was difficult to figure out how participants would write powers and square root signs. There are keyboard short cuts but having this option might hinder rather than help. Most of the problems do not require them so this is how we got around this issue. For further research however, it will have to be addressed possibly with a toolbar or add-on that would allow students to plug it into the box. Other problems we faced were attempting to get the quiz online instead of the preview window, and participants only being able to use one computer at a time. When a quiz was presented in the viewer, all the data would be lost if another quiz was opened to view or if the window was closed. This forced us to transfer the data to an excel sheet as soon as the participant was done so that the data would not be lost. Occasionally a student would be late and therefore pushing all participants back because of the issue of one computer. For a few participants who were more than 30 minutes late, we simply asked them to do the pretest. This way at the time of the post test we will have them do the computer quiz and go right into the post test. Many participants thought that the quadratic formula was necessary to solve certain problems and asked how they should input this data. It was difficult because I could not give them the information that the formula was not necessary, but at the same time it had to be consistent. From this I also learned the importance of good instructions and examples before the participant takes the quiz. Furthermore, I made mistakes regarding sign up times and using Sona with the academic calendar creating a date for thanksgiving and the Friday that there was no instruction. In the future I will have to be keener about these dates to avoid confusion and cancelation.

Overall, I thoroughly enjoyed researching with Jessica Walker and learned a great deal about how to create and conduct an experiment. I learned how to appropriately handle obstacles that come with participants and their corresponding data. So far the results have been very promising with students actually willing to write more on the free response questions. The interface of the computer does not seem to impede participants or the input of their answers. For the first participants we asked for feedback regarding the ease of doing the quiz on the computer and if they had any constructional criticism. In general participants had a positive experience and had few suggestions. The quiz is now possible to view online however, there is an image of the brand name that hinders the instruction page that we are still attempting to alleviate. We are also hoping to make it possible so that when a participant gets no feedback, it is impossible to see if it was correct or not. Furthermore, we are limited with the number of acceptable answers. This makes our program less open to errors in spacing, commas and other symbols. If we took this study to a more finished product these problems would be fixed. I hope to stay on the research team and see this project build hopefully into a website or program that can provide assistance for community college students struggling with mathematics.

*References*

Needham, D., and I. Begg. "Problem-oriented Training Promotes Spontaneous Analogical Transfer: Memory-oriented Training Promotes Memory for Training." *Memory & Cognition* 19 (1991): 543-57.

Rothman, W. Rosalind, and Jill Cohen. "The Language of Math Needs to Be Taught." *Academic Therapy* 25.2 (1989): 133-42

Gries, David, and Fred B. Schneider. "Teaching Math More Effectively, Through Calculational Proofs." *The American Mathematical Monthly* 102.8 (1995): 673-763.

**Developing Objective Measurements in Perceptual Causality Studies**

**Jon Lexa**

*Perceptual Causality Yesterday and Today*

Research in the field of perceptual causality has grown since Belgian experimental psychologist Albert Michotte first tested with launching effects. Despite this growth, much of the work has yet to advance like some other fields. For example, it could be argued that contour and shape perception research has moved along quite far in comparison. Part of the problem arises from the field of perception and causality being incredibly complex.

When thinking casually of causality, one can already begin to see the intricacies. How do we infer causality, what cognitive processes are needed to accomplish such a feat, why are we at times incorrect at inferring causality? These are all questions which can be thought of at surface level when pondering causality.

Perception is also a field riddled with complexity. Particularly, the debate that arises regarding the role of the perceptual process. The idea is generally accepted that one can perceive an object to be animate and cause another object's action. The question that sustains is how much of that process is perceptual processing. Because perceptual causality requires a cognitive effort and perceptual processes are usually regarded as low level mechanisms, researchers debate over whether perceptual processes might be involved in higher level cognition.

Apart from figuring out the role cognitive levels play in perceptual processing, controversy is also exhibited when discussing whether this perceptual processing is innate. The difficulty with standing on one side of the argument or the other, is that it is very difficult to prove something as innate. Sure, we can study infants who have relatively little to no experiential knowledge, or study across cultures, ages, and continents to provide results that are free of most cultural constraints, yet these would still not provide concrete evidence of perceptual causality being innate. A study completed by Goodman, et al. (2010) showed that it was possible to build a computational model that had implications for the innateness of causality in cognitive development. Following the theory that the presence of abstract learning is often interpreted as evidence for innateness, they proposed that causality may have abstract frameworks that allow for causal inferences to be made.

These are several areas which make perceptual causality a difficult field to research, and by no means are these the only issues. I will now discuss what I believe to be an area of importance that may be easier to approach than the aforementioned problems.

*Observations of Perception*

Perception is often times our subjective account of the world around us. We tell people what color we think the sky is, we comment to each other about certain perfumes, we describe the weather through the way we feel. Therefore, when studying perception, it is easy to get create experiments which record subjective data points. However, subjective results are often times weak and easily attacked. Henceforth, researchers have noticed the importance of objective results in perception studies, and have made significant efforts towards obtaining them. In a study done by Palmer, et al. (2006), objective responses towards occluded objects was obtained by cleverly asking the participants not if they could recognize the occluded object, rather by instructing them to state where the occluded object ended and began. This gave a quantitative result that was more objective than having the participants say they thought it was a certain type of object.

All the efforts (to my knowledge) in the perceptual causality field of studying inferences of causality examine subjective responses made by the participant. As Kant stated in his Second Analogy, we have subjective observations of an objective sequencing of events. We can often agree on what we perceive as the cause of a ball being launched, like we see in the examples in Michotte's studies (1946), yet the agreements are subjective in nature. As stated earlier, a subjective observation can be problematic for several reasons.

First, it produces a flimsier proof in support of the hypothesis. Subjectivity can be easily argued, whereas objectivity resolves one of having to instantiate one's opinion, gracefully avoiding issues of debate. People may also report different interpretations of the same event. My viewpoint of an action is not the one and only viewpoint. Other viewpoints are just as right as mine, and in some cases may be even more correct.

Second, the popular launching display is a very simple and controlled situation that does not allow much scalability. The subjective reports of causality from launching displays cannot be easily transferred to other situations involving causality, especially real world scenarios. For example, three dimensional scenarios that surround our daily lives complicate the causal relations which are easily recognizable in simple launching events. Subjective judgments have more variability and are harder to contain in an experimental paradigm than objective findings.

*Benefits of Objective Responses*

As one can infer from the problems that arise from subjective responses, obtaining objectivity would allow for sturdier accounts of causal perception, accounts that can hold their own against criticism and possibly be transferred to other causal situations.

Another benefit of creating an objective measurement of causal perception is that this can be used as a standard across the causal perception field. This would allow for researchers to create a baseline, from which they can build their research. A current issue within the field is that researchers are exploring the causal perception phenomenon but not using standard terminology when titling their papers or writing their abstracts. Using a standard method of objective measurement would give researchers now and in the future, the ability to string together the studies with a commonality: the measure used.

### *Attempts at Creating a Standard*

A study done by Tanaka, et al. (2008) used an fMRI to link subjective causality to the objective contingency. The contingency being the rate at which a button is pressed to the amount of money earned. They did find increased activity in the medial orbitofrontal cortex and dorsomedial striatum when the subjects were judging the causal efficacy of their actions. I am not sure whether the purpose of the researchers was to attempt bridging the gap between subjective interpretations and objective measurements; however, the results are only correlations and those brain areas cannot be assumed to be the core of human causality. Although an fMRI does provide us with objective measurements, until we can understand the link between thought and brain activity, we cannot use this as a way to objectively measure perception of causality.

### *Difficulty in Measuring*

We can look at examples of previous work to build an objective measurement of perceptual causality (Palmer, 2006). Perhaps there is a way of producing observations of causality in variations of launching displays in participants through answering an objective test of some other measure. Precaution would have to be taken to ensure that the participant is not explicitly instructed to locate causal events with words such as "hit," "cause," "launch." An interesting method to accomplishing this would have the participants participate in a tracking task with multiple balls being launched. They would then be asked to identify the path they believed the ball would take. By identifying the source and the target, they would be inferring causality, yet their responses would be recorded with objective counts of correctness. Along with balls representing causal events, non-causal events could also be introduced to see if the same findings regarding the transference of causality from a causal event to a non-causal event from Scholl's study can be replicated within the tracking task.

*Conclusion*

Hume states that causality is unobservable, and what we perceive between two events are assumptions. If this holds true, then any kind of objective measurement that we develop will be of an unobserved cause, and all efforts to strengthen the findings of the causal perception field will be built on assumptions. The thought that an entire field of study would be based on an assumption is slightly concerning, yet it is not the first time that science has progressed on such a foundation.

However, if Hume's idea of causation does not hold true, then the benefits of creating an objective measurement of causal perception are numerous. To summarize some of the benefits, simply creating an industry standard and allowing for stronger results would help in advancing the field. Unless cognitive neuroscience develops to the point where we can draw more conclusions then just correlation, the method of objective measurements will probably take form through clever experimental manipulations.

Because of my limited knowledge in the field, this is just a mere presupposition at the possibility of an objective measurement in the field of perceptual causality. Despite my efforts to push the field forward, much of the work completed has tackled the difficulty of this nature head on. I would like to thank Dr. Phillip Kellman and Everett Mettler for their insight and expertise in this matter.

*References*


Hume, D. (EHU) (1777). *An Enquiry concerning Human Understanding*. Nidditch, P. N. (ed.), 3rd. ed. (1975), Clarendon Press, Oxford.

Hume, D. (1740). *A Treatise of Human Nature* (1967, edition). Oxford University Press, Oxford.

Kant, Immanuel. *Critique of Pure Reason*. Transl. Paul Guyer and Allen W. Wood. New York: Cambridge UP, 2009.

Michotte, A. (1946/1963). *La Perception de la Causalite*. Louvain: Institut Superieur de Philosophie, 1946. [English translation by T. Miles & E. Miles, The perception of causality, Basic Books, 1963.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory object perception. *Journal of Experimental Psychology: General, 135* (4), 513–541.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science, 13*, 493-498.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4* (8), 299 - 309.

Tanaka S.C., Balleine B.W., & O' Doherty J.P. (2008). Calculating consequences: brain systems that encode the causal effects of actions. *Journal of Neuroscience; 28*, 6750– 5.

**Judgments of Learning and Other Metacognitive Measures**

**Maxwell Mansolf**

*Background*

A fundamental mystery in cognitive psychology, albeit one that has only recently been studied extensively, is how the human brain monitors its own activities. The Latin prefix "meta" meaning "in the midst of, among, between, with" is used extensively in the literature relating to this topic, which has aptly been dubbed "metacognition." Intimately related to the concept of metacognition is that of "metamemory," the memory of memories. Research in metamemory is of critical importance in the field of education, as the ability of students to judge the extent of their learning governs their study behavior. If a student is aware that, for example, they know the material in one chapter of a textbook much more than the material in another chapter, both of which will be tested on, the student will typically favor studying the less well-known chapter.

Obtaining information on the complex mental representation of memory strength remains a difficult task for any cognition researcher. Many measures have been formulated to tap this resource, the most frequently used of which are feelings of knowing (subjects are asked whether they would recognize a study item, if it were presented to them), judgments of learning (subjects are asked how likely they are to remember a study item later), and confidence judgments (subjects are asked, post-test, how confident they feel about the answer they provided). These resources all have their strengths and weaknesses as predictors of various dimensions of memory performance, though the measurement that is currently the focus of most metamemory research is the judgment of learning, or JOL.

The focus of this paper will be on judgments of learning: their origins, their strengths, their weaknesses, and how they can be made more accurate. The motivation for increasing the accuracy of judgments of learning is simple, and relates back to the application of metamemory research to education: the more accurately a student can distinguish what is known and what is unknown, the more efficiently the student can allocate study resources.

### *Judgments of Learning: Foundations*

An obvious method of determining how much an individual knows about their memory is to ask them. As such, a considerably body of cognition research has been devoted to a single type of measurement: the judgment of learning (JOL). The method of taking a JOL is simple: a subject is asked how likely they are to remember an item at a later time, usually a later memory test. These measurements are usually collected on a 0-10 or 0-100 scale, where a 0 indicates that an item will certainly not be remembered, a 10 or 100 indicates an item will surely be remembered, and a 5 or 50 indicates an item has about a 50% chance of being remembered at later test ("remembered" relative to the specific test method, typically recall or recognition).

These judgments of learning have several important properties, each of which reveals a different aspect of metamemory. The *accuracy*, or *calibration*, of JOLs compares, across an entire study-test period, the predicted (based on JOLs) and actual (based on performance at test) memory performance on studied items, and is typically reported by comparing the mean JOL score across an item list to the mean probability of accurate performance on the proceeding memory test. JOL accuracy is used as a measure of overall monitoring ability; if a subject is very accurate in making their JOLs, they think they remember as much information as they actually do remember. The *resolution* of JOLs measures how well subjects can distinguish which items are later remembered or not remembered, and is typically measured via the Goodman-Kruskal gamma correlation G, which is calculated according to the number of concordant (ranked high and recalled, or ranked low and not recalled) and discordant (ranked high and not recalled, or ranked low and recalled) items in the study; G varies from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. JOL resolution measures the overall ability of a subject to be aware of what they will and will not remember, and while accuracy and resolution are often correlated, resolution offers a more functional representation of memory monitoring ability.

In the first research on judgments of learning, researchers found that yes-no post-study judgments of knowing reliably correlate with later memory performance (Arbuckle & Cuddy, 1969), demonstrating that subjects can monitor their own knowledge; however, as research expanded and the limits of these abilities began being tested, less encouraging results emerged. For example, in the first publication to address the resolution of judgments of learning, researchers found that the mean gamma correlation between JOLs and recall performance was 0.30 (Leonesio and Nelson, 1990) and in a recent unpublished meta-analysis of JOL research, the mean gamma correlation for immediate JOLs across many studies was 0.42 (Rhodes & Tauber, in press). While these correlations are reliable and demonstrate that humans are capable of monitoring their own learning immediately after study, they are also fairly small and demonstrate that this monitoring, while present, is fairly poor.

*JOL: Theories and Shortcomings*

In early JOL research, JOLs were believed to directly tap into the strength of a memory trace, a theory called the "direct-access hypothesis" (King, Zechmeister, and Shaughnessy, 1980). In 1997, Asher Koriat of Haifa University in Israel formulated a new framework for JOLs, called the "cue-utilization approach". According to this theory, JOLs are not made by judging the strength of a memory trace directly, but by utilizing intrinsic (item-specific), extrinsic (context-specific) and mnemonic (akin to a feeling-of-knowing judgment) cues, and depending on whether these cues actually correlate with recall performance, judgments of learning could range from very accurate (if the cues used correlate well) and very poor (if the cues correlate poorly) (Koriat, 1997). The cue-utilization view of JOLs agrees with most current theories of metacognition in that metacognitive judgments are not made directly, but indirectly through heuristic methods.

The cue-utilization approach predicts a fundamental flaw in how JOLs are made: because JOLs are based on cues, some of which are not present at test and some of which correlate poorly with memory performance, the JOL values themselves are sometimes poorly correlated with, or even negatively correlated with, memory performance at test. For instance, in one study, subjects were presented with words in varying font sizes, and gave higher average JOLs to words in larger font sizes, despite font size being uncorrelated with memory performance (Rhodes & Castel, 2008). Another factor that can bias JOLs and lead to "metacognitive illusions" is associative strength; when a word pair is presented that has a high associative strength of the first word to the second word (ex. "kitten-cat" has a high associative strength), subjects are confident that they will later recall the second word when presented with the first (ex. recall "cat" when given "kitten"), but when the words are switched such that the associative strength is weaker (ex. "cat-kitten" has low associative strength), subjects will believe they are just as likely to remember the second word when presented with the first (ex. recall "kitten" when given "cat"), and give high JOLs to both normal and reversed pairs, while recall performance is much worse for the reversed pairs than for normal pairs (Koriat & Bjork, 2005). These and other cues which subjects use to make JOLs are believed to lead to the poor correlation between JOLs and memory performance, and much work in the field of metamemory has been in pursuit of methods by which to improve the resolution of JOLs; two of these methods are described below.

### *Improving the Resolution of JOLs – The Delayed JOL Effect*

In 1991, Nelson and Dunlosky at the University of Washington experimented with a method of improving JOL resolution, which while effective, has been met with much controversy since its discovery (Nelson & Dunlosky, 1991). Nelson and Dunlosky presented subjects with word pairs to study, similarly to other JOL research, but while some JOLs were made immediately after study, some JOLs were made after a delay; for instance, a participant would see a word pair (cue-stimulus) and after studying several other word pairs would be presented with the cue word, and asked to give a JOL on the word pair associated with that cue. Subjects performed dramatically better for these word pairs than for immediate JOL pairs (gamma of 0.90 compared to 0.38), and Nelson and Dunlosky dubbed this the “Delayed JOL Effect.” In the original paper, the theoretical explanation for the delayed JOL effect, dubbed the “monitoring dual memories” hypothesis, is that when subjects make delayed JOLs, they utilize primarily long-term memory (LTM) cues which correlate well with memory performance, whereas when subjects make immediate JOLs, information from short-term memory (STM) interfered with access to LTM cues, resulting in a dramatic improvement in JOL resolution when JOLs are made after a delay.

Soon after this paper was published, however, Spellman & Bjork published a short paper attempting to debunk the delayed JOL effect; they reasoned that when subjects are presented with the cue word and asked to make the delayed JOL, they are in fact attempting to recall the stimulus word (Spellman & Bjork, 1992). If the recall attempt is successful, the word pair is given a high JOL, and the successful recall attempt begets improved recall performance at test; likewise, if the recall attempt is unsuccessful, the word pair is given a low JOL, and the failed recall attempt begets diminished recall performance at test. This hypothesis, dubbed the “self-fulfilling prophecy” hypothesis, has been widely accepted as the definitive explanation for the delayed JOL effect. However, a recent large-scale meta-analysis of 98 effect sizes from 42 delayed JOL studies (published and unpublished) reported that while there is a large effect on monitoring in delayed JOL conditions compared to immediate JOL conditions across the studies, the effect of the delayed JOL condition on memory performance is much smaller, albeit still reliable (Rhodes & Tauber, in press). Furthermore, even when subjects are presented with cue-stimulus when delayed JOLs are made, the delayed JOL effect reliably appears, albeit not with the same magnitude as in cue-only conditions. These results, while not ruling out the self-fulfilling prophecy hypothesis completely, demonstrate that it fails to explain all of the data, and some combination of the two hypotheses, along with other unaccounted-for factors, can potentially explain the delayed JOL effect completely, though the degree to which these factors influence the effect is still a mystery.

*Improving the Resolution of JOLs – The Gambling Task*

Another method that has been proven to improve the resolution of JOLs, pioneered at UCLA by Shannon McGillivray and Alan Castel, is the gambling task (McGillivray and Castel, 2011). In this task, instead of making JOLs on items, participants place a point value on each item, from 0 to 10, and if the word is later recalled, those points are added to their final score, otherwise the points are subtracted from the final score. All bets are made immediately after studying each item. When subjects are new to the task, they tend to perform very poorly, but across multiple study-test trials, when given feedback in the form of their overall score at the end of each list, participants perform increasingly well. In later research comparing the resolution of these bets to memory performance, this "betting" manipulation has been proven to be almost as effective as the delayed JOL condition in improving JOL resolution. The current explanation of these betting effects on JOL resolution is that once subjects are familiar with the task, they tend to choose to give high scores to easy-to-remember items and put forth effort to remember those items for the memory test, and if an item seems difficult, it is not bet on (i.e. given a point value of 0) and no more attention is paid to it. Because the task is so new, very little work has been done to examine the interactions between this gambling condition and other factors known to improve or reduce JOL resolution; I propose a new line of research below with this intention in mind.

*Delayed bets: Theory and Predictions*

In my proposed research, the effects of betting vs. JOL and immediate bet/JOL collection vs. delayed collection, including a novel delayed gambling condition, will be examined, with the ultimate goal of discovering a more precise and useful metacognitive measure than delayed JOLs or immediate bets. The proposed pilot experiment will duplicate the method of Nelson & Dunlosky (1991), but with a separate condition in which bets are made instead of JOLs; in other words, immediate vs. delayed JOL/bet collection will be a within-subjects condition, and JOL vs. bet will be a between-subjects condition. In theory, the two manipulations should both facilitate higher JOL resolution; the delayed JOL effect, as conceptualized in Rhodes & Tauber (in press), is primarily a function of cue utilization, and even if a covert recall attempt is made, subjects will be better at determining which words will be recalled at test after a delay than immediately after study. Subjects will then use this information more efficiently when making bets than JOLs after the delay, because the cues utilized to determine which items will be easy or difficult to remember (cues drawing from LTM) will better match conditions after a delay than immediately after study. However, because both the gambling task and the delayed JOL task produce resolution values close to the maximum value of 1, these effects will not be strictly additive, and a ceiling effect will likely emerge; one proposed correction to this study is to present cue-stimulus instead of cue-only at time of delayed JOL, to elicit the weaker delayed JOL effect, and perhaps avoid ceiling effects. For the delayed JOL condition and immediate bet conditions, we expect to reproduce the results of Nelson & Dunlosky (1991) and McGillivray and Castel (2011). Because the pilot study will only use one study-test trial, however, we don't expect robust effects of the gambling condition to appear in the pilot study, though future work will likely incorporate multiple study-test trials.

Another prediction for the delayed gambling condition is that these bets will be more polarized than either delayed JOLs or immediate bets, which have been shown to elicit either very high or very low point values, an effect not seen when immediate JOLs are taken. Because the values in the two original conditions are already very polarized, it is likely that ceiling effects will emerge, but because polarization of point values is a function of value choice, and not monitoring ability as with JOL resolution, and because the true ceiling (eliciting only values of 0 or 10) hasn't been reached in either of the original conditions, it is expected that this increased polarization will occur, and it would be very peculiar if the effect did not emerge.

Because the effect of the gambling condition is primarily one of study strategy, of particular interest is how subjects will allocate their bets between delayed and immediate bets. If subjects believe that after a delay, these bets will be more accurate, they will give higher bets in the delayed condition than in the immediate condition, and vice versa if they believe that bets will be more accurate immediately after study. If there is no correlation between the presence of a delay and the bets subjects give word pairs, this will demonstrate that bets are made on an item-by-item basis and that subjects cannot observe the delayed JOL effect as it is occurring (an interesting post-test question would be whether subjects intentionally gave higher bets immediately after study or after a delay). In terms of the two theories of the delayed JOL effect (monitoring-dual-memories or self-fulfilling-prophecy), it is currently unclear how this research will help to discriminate between the two, as both seem to predict the same results in this model, but hopefully as more data is collected and the next steps are taken in the research, the effects of the conditions will be clearer, and these fundamental aspects of the delayed JOL effect will become easier to tease apart. As it currently stands, however, this research is still promising in that it could provide insight into the cues utilized in making JOLs when given motivation to perform accurately, as well as create a metacognitive measure with better resolution than delayed JOLs or immediate gambles, demonstrating that subjects can be very capable of monitoring the reliability of their own memories given a specific study format.

*References*


Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. Journal of Experimental Psychology, 81, 126 –131.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects-of memory? Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 464-470.

King, J. E, Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. American Journal of Psychology, 95, 329-343.

Koriat, A.  (1997).  Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning.  Journal of Experimental Psychology: General, 126, 349-370. doi: 10.1037/0096-3445.126.4.349

Rhodes, M.G., & Castel, A.D. (2008). Memory Predictions are Influenced by Perceptual Information: Evidence for Metacognitive Illusions. Journal of Experimental Psychology: General, 137(4),615-625.

Koriat, A., & Bjork, R.A. (2005). Illusions of Competence in Monitoring One's Knowledge During Study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 187-194.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The 'delayed-JOL effect.' Psychological Science, 2, 267-270. doi: 10.1111/j.1467-9280.1991.tb00147.x

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. Psychological Science, 5, 315-316. doi: 10.1111/j.1467-9280.1992.tb00680.x

McGillivray, S., & Castel, A.D. (2011). Betting on Memory Leads to Metacognitive Improvement by Younger and Older Adults. Psychology and Aging, 26(1),137-142.

Rhodes, M.G., & Tauber, S.K. (in press). The Influence of Delaying Judgments of Learning (JOLs) on Metacognitive Accuracy: A Meta-Analytic Review. Psychological Bulletin.

**An Investigation on Causal Reasoning**

**Alice Merrick**

*Abstract*

The present study tests the *SS power model*, a Bayesian model of causal learning that has been extended to sequential learning. In contrast to previous models, the SS power model hypothesizes that learners are guided by priors that favor a minimal number of causes (sparse), each individually of high strength (strong). Previous research has shown that a stronger cause will tend to reduce the apparent strength of a weaker cause when the two co-occur. We are currently testing the model in a paradigm in which participants learn causal strengths based on sequential presentations of data. The task involves judging the causal strengths of foods that may produce allergic reactions. Foods are always presented in pairs, after which participants judge the strengths of each individual cause. These judgments establish the strengths of the causes for the subsequent generation of examples, so that strengths change over the course of the learning process. The SS power model predicts that over the course of several generations of examples, participants' judgments of causal strength will converge towards their sparse and strong priors.

*Introduction*

People have a remarkable ability to learn causal relationships, but how does a reasoner come to know what one thing cause another? How can people acquire causal knowledge from limited observations? As philosopher David Hume pointed out, our observations don't include any causal information. So how do people determine causal relations from inherently non-causal data?

A naive solution would be that people try to learn all possible covariations among events. That would be an intractable problem, and especially unlikely because people can learn causal relationships from only a handful of examples. This suggests that causal relationships must be inferred.

One argument is that human induction is guided by prior assumptions people make about the nature of causal relationships. One principle that people might use to to make inferences about causal relationships is simplicity. This idea is familiar to scientific inquiry as Occam's razor: According to Isaac Newton's first Rule of Reasoning in Philosophy, "We are to admit no more causes of natural things, than such as are both true and sufficient to explain their appearances." Simplicity, however is not only difficult to quantify, but it's also unjustifiable to *assume* that it's used in induction.

Nevertheless, simplicity is a naturally appealing heuristic, and we do find evidence for it. In trying to answer why, given that data are inevitably consistent with multiple explanations, are some causal explanations chosen over others, Lombrozo (2007) explored the role of simplicity in evaluating competing explanations. Lombrozo found that simpler explanations are assigned higher prior probabilities. Her findings suggest that simplicity *is* used as a basis for evaluating explanations and for assigning prior probabilities (2007). In other words, people prefer explanations that require fewer causes.

The goal of this article is to test a model that has formalized the role of simplicity in causal reasoning, the *SS power model*, using the method of iterated learning. The model works within a Bayesian framework for causal learning (Griffiths & Tenenbaum, 2005) and models simplicity using "*generic priors*—systematic assumptions that human learners hold about abstract properties of a system of cause–effect relations" (Lu et al., 2008).

A number of previous models have tried to explain how humans make inferences about causal relationships. The two that the SS power model directly stems from are the power PC theory (Cheng 1997) and Griffiths' and Tenenbaum's causal support model (2005). The power PC theory is based on cause effect probabilities, and reconciles the idea that causal relationships are not directly observable with the idea that people hold prior beliefs about the power causes have to produce or prevent their effects (Lu et al., 2008).

According to the power PC theory, human judgments of causal strength are equal to the probability of the cause producing the effect in the absence of all other causes. The power PC theory provides a better fit for some human data than its precedent, the ΔP model, in which call the strength is based on the difference between the probability of the effect occurring when the cause occurs and the probability of the effect occurring when the cause does not occur (Cheng, 1997), but neither model provides a full account of human performance.

Griffiths and Tenenbaum proposed a model based on Bayesian statistics that used causal graphical models. The use of causal graphs made it possible to distinguish between causal structure and causal strength. Griffiths and Tenenbaum (2005) gave an account of learning causal structure while Lu et al. (2008) showed how this approach could be extended to infer causal strength (Griffiths and Tenenbaum 2011).

Although the graph structure specifies the causal relationship among variables, a function is necessary to specify the strengths of those relationships. This model, like previous models uses Noisy-OR (for generative causes) parameterizations to characterize the function form of causal relationships in which causes are assumed have the power to cause or prevent the effects independently (Griffiths & Tenenbaum, 2011). The Noisy-OR function determines that the probability of observing the effect *e* is:

where $w_0$ and $w_1$ are the strength of the background cause and candidate cause respectively, and *c* is a binary value representing the presence or absence of the candidatecause. For our purposes, $c = 1$. The probability of getting the observed data from a causal relationship with weights $w_0$ and *w,* $P(D|\, w_0,w_1)$, or the *likelihood* of the data, is given by the product of observing (or not) the effect, given $c, w_0,$ and $w_1$ over $e^+$ and $e^-$ and $c^+$ and $c^-$.

By applying Bayes' rule, we can then use the likelihood function to compute the posterior distribution over $w_0$ and $w_1$:

$$P(e^+|c;w_0,w_1) = 1-(1-w_0)(1-w_1)^c$$

where $P(w_0,w_1)$ is the prior on $w_0$ and $w_1$. But in order to do this, first we must estimate the prior on $w_0$ and $w_1$. Because people have a preference for causal models with fewer causes (Lombrozo, 2007), and for causes that minimize complex interactions (Novick & Cheng, 2004), the priors we will be using are the sparse and strong priors (Lu et al., 2008).

$$P(w_0,w_1|D) \; \square \; P(D|w_0,w_1)P(w_0,w_1)$$

In previous work, Lu et al. evaluated different kinds of priors on causal strength by testing predictions of different models implementing those priors (2008). Here, we are using an experimental method based on iterated learning to test if people are using generic priors. Iterated learning can be explained as having a chain of agents in which one agent answered data that was generated by the previous agent and then forms a hypothesis about how that data was generated and then generates new data for the next agent (Griffiths & Tenenbaum, 2011). A classic example of this is language learning, and in fact, iterated learning was first used in studying cultural transmission.

If the agents are using a Bayesian inference then as the chain gets longer the probability that an agent picks a particular hypothesis converges to the prior probability of the hypothesis (Griffiths & Tenenbaum, 2011). For each iteration the agent's response depends only on the current data, so we don't even need multiple agents to simulate iterated learning in a laboratory setting. The response that an agent gives in one trial can be used to generate the data they will see in the next trial.

Before actually running the SS power model, (which we will leave for next quarter) our aim is to show that over several generations of iterated learning, participants' responses will converge to the generic priors they use to estimate the strength of causal relationships.

*Methods*

*Participants.* The participants were University of California, Los Angeles, undergraduates. They were recruited from the Psychology Department's subject pool and received course credit for participating.

*Stimuli and Procedure***.** Similar to the previous version of the experiment, the cover story concerned a doctor trying to determine how likely it is that certain food(s) are causing his patients' allergies. The experiment was run in MATLAB with an interface created with Psychtoolbox. The participants were asked to answer a question, which was suggested by Patricia Cheng, to test if they were using causal inference.

> We conduct a study of medicine X, and find that: 50% of the participants who received medicine X (those in the experimental group) have headaches. Likewise, 50% of the participants who did not receive medicine X (those in the control group) have headaches as well. Recall that participants were randomly assigned to the two groups. Can the headaches in the experimental group be attributed to medicine X?

The participants clicked a "yes" or "no" button to respond. If a participant answered "No" to the above question, they were considered to be using causal inference. Then the participants were given the following scenario:

> Recently a large number of tourists have visited an island in the South Pacific, and soon afterwards developed an unusual skin rash. A doctor has narrowed down the possible causes to three exotic fruit, called cherimoya, melona, and Buddha's hand. [Figure 1]

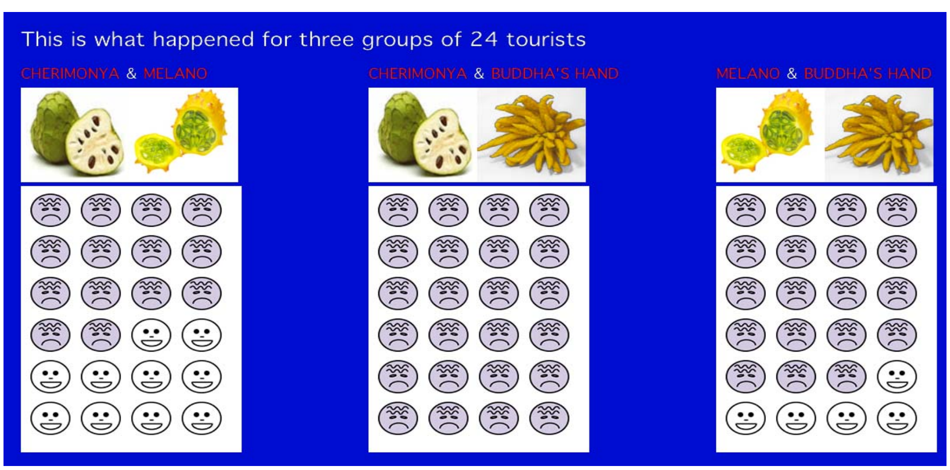

Figure 1: A screenshot of the initial data presented to each participant. The frowning face represented the occurrence of the rash.

> Now the doctor is trying to determine how likely it is that eating each of the fruit will cause the rash. He also knows that the skin reaction is caused by individual fruits, not special fruit combinations. However, this task is more difficult because it turns out that the fruits are always served as a salad, in which exactly two fruits are mixed together.
>
> Your job is to play the role of the doctor trying to assess the effect of each fruit on the rash. You will see a series of samples of tourists who ate various fruit combinations, and will be able to see how many came down with the rash, and how many did not. You can assume that the unusual skin rash has no possible cause other than one or more of the three exotic fruits. Based on this information, you will have to answer a series of questions about the how likely it is that each individual fruit causes the rash.

At the start of the first generation the fruits were assigned probabilities of causing an allergic reaction. The initial parameters were 0.2, 0.5 and 0.8. These probabilities were randomly assigned to each food at the beginning of each trial. For each pair of foods, the probability of either food causing an allergic reaction was calculated on the basis that the two were independent events, using the Noisy-OR function given by (1). The data were presented graphically using pictures that showed the total number of tourists in each group as well as the number that developed a rash. The total number of tourists remained constant throughout the experiment at 24, the idea being that sparse data will make participants more likely to rely on priors (Lu et al. 2008).

The participants were then asked:

> Suppose that there is a sample of 100 tourists without the rash. If these 100 tourists ate [FRUIT] alone, how many of them would have the rash?

All three data summaries remained on the screen while the participants responded to alleviate demands on memory. Participants entered their responses via keyboard. Originally we used a within-subjects iterated learning design. In this design, a participant's responses were converted to probabilities and used to generate the data in the next generation for a total of 10 generations per participant.

*Results*

The results from using initial probabilities of 0.2, 0.5 and 0.8 were mixed at best. There seemed to be little consistency in participants' responses, with answers changing dramatically from one generation to the next. We next tried using initial probabilities of 0.1, 0.5, and 0.9 or 0.1, 0.1, 0.9. Although there seemed to be a convergence towards final probabilities of 0, 1, 1, seemed to be the trend, responses nevertheless suffered the same problems as before.

*Discussion*

Although we eliminated the memory demands of the last version, the task seemed to be still too difficult for some participants to continue answering reliably for all ten generations. In some cases their responses might have converged to the prior probabilities, but then diverged afterwards. Because we used so little information from one response to generate the next set of data, the chains were sensitive to mistakes that might be made when entering responses. In order to get more reliable responses, we decided to try a between-subjects sequential design in which each participant responded to one set of data, creating the data for the next participant. This will, of course, require a much larger number of participants.

Our plans for next quarter include running the SS power model on the data and redesigning the experiment to better suit running one generation per participant. To make the chains more robust against errors, we will present the data several times using different sample sizes (e.g. 16, 24, 32) and then average the responses to generate the data for the next participant.

# References


Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1119–1140.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405.

Griffiths, T. L., & Tenenbaum, J. B. (2005). *Structure and strength in causal induction. Cognitive Psychology, 51,* 334–384.

Hume, D. (1739). *A Treatise of Human Nature*.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P., & Holyoak, K. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115(4),* 955–984.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*(3), 232–257.

Newton, I. (1687, 1729) *The Mathematical Principles of Natural Philosophy*, trans. A. Motte.

Novick, L., & Cheng, P. (2004). Assessing Interactive Causal Influence. *Psychological Review, 111(2),* 455–485.

Yeung S., & Griffiths T. L. (2011). Estimating human priors on causal strength. Proceedings of the 33rd Annual Conference of the Cognitive Science Society.

## The Discrimination of Sinusoidal Contours and the Level at Which Shape Processing and The Recognition of Global Form Occurs.

**Sophie Gerrick**

Though much research has been done regarding shape perception and discrimination, the level at which shape processing and the recognition of global form occurs remains unclear. Many theorize that contours can be detected in part by the primary visual cortex (the V1 area). However, further research has demonstrated that the V1 region alone cannot account for all of contour detection and recognition. Thus, the current belief is that form is further determined by the V4 region (Loffler, 2008). As the V1 region focuses on localized line segments, the recognition of full, geometric shapes requires an additional system. In order to recognize these complex shapes and discriminate transformed versions of the contour, a more global representation must exist.

My research project has been focused on the discrimination of sinusoidal contours and determining the level at which this discrimination occurs. Specifically, we are interested in establishing the time necessary to go from local to global processing. To identify this, we will show subjects a series of sinusoidal image pairs separated temporally by a mask. The second image in each pair will either be identical to or a morphed version of the first image. The second image will have undergone a slight transformation, randomly chosen between rotation and scaling. After each image pair, we will ask participants if the two contours were identical (albeit with the transformation), or if the second stimuli was a different, morphed image. By only showing the two images for short amounts of time, we can learn more about how the contour is processed and discriminated.

And, by varying the exposure time, we can identify the time necessary for global rather than local processing.

I plan to use transformations on the images to ensure that the shape recognition occurs on a global level. By displaying the two stimuli in randomized locations on the screen, we apply translations to reduce the subject's ability to use local (V1). Additionally, to reduce the participants' ability to adjust to the translation, we will use additional transformations. By randomizing the application of the scaling and rotating transforms, we can ensure that the participants' ability to recognize the contours is dependent on the global, V4 representation.

The V4, intermediate representation mechanism has several known qualities. For instance, it has been shown that such shape processing does not change with size or position, but can be influenced by orientation (Loffler, 2008). By transforming our stimuli using either scales or rotations, we can determine more precisely where and when the intermediate shape processing takes place.

In order to establish the exposure time necessary for V4 processing, we must make certain that the participants are not able to continue processing the first image after the exposure time. To interrupt the image processing, we display a mask, a randomly generated screen of black and white pixels (Tse, Martinez-Conde, Schlegel, & Macknik, 2005). Although previous studies indicate that the mask would have an effect whether it was displayed within the contour or occluding it (Loffler, 2008), we plan to mask the whole screen to ensure that the image processing is interrupted as thoroughly as possible. Thus, any recognition should be due to the exposure time alone.

We generated the set of 100 contours and their respective morphs by generating random sinusoidal contours and alterations. We chose to use closed contours due to the increased sensitivity to closed shapes (Dumoulin & Hess, 2006). Additionally, as researchers at McGill Vision Research Unit found that a patient with a lesion around the V4 region of his brain had difficulties processing concentric and closed sinusoidal contours, it is likely that these stimuli will induce a strong, observable effect on the V4 region (Dumoulin & Hess, 2006). If so, it should be possible to use these contours to determine the point at which the processing switches from V1 to V4.

Current research indicates that the V4 region of the brain responds more significantly to areas of high curvature (Feldman & Singh, 2005). The researchers note that negative curvature is more readily recognized and that concave extrema "carry the most information about the local shape of an object", but other parts may be more important in other aspects. Due to the range of sinusoidal contours that we generated, it should also be possible to determine if the concavity and complexity had a significant impact on participants' ability to recognize shapes with various exposure times. If the strength local curvatures affect performance, this may well indicate that the results are still significantly affected by the V1 response. It may be necessary to compare the integrals of the concave and convex regions of different stimuli and determine whether these had an effect on the recognition of the transformed contour.

Researchers at the Glasgow Caledonian University looked into the effect that radial frequency has on object discrimination. They found that the subjects' ability to discriminate between contours was significantly lower at higher radial frequencies. They asserted that local cues cannot sufficiently explain this effect. As the shapes with higher radial frequency should have higher magnitude concavities, the local (V1) response should be stronger for such shapes. However, the observed performance could be explained by a more global recognition mechanism (Loffler, 2008), supporting our decision to utilize closed sinusoidal contours as the stimuli in our experiment.

The researchers also noted that at high radial frequency (more than 8 nodes), the participants' ability to recognition closed contours was roughly equivalent to their ability to distinguish between small portions of the forms. This implies that for radial shapes with more than eight projections, the global system of shape recognition is less effective (Loffler, 2008). As our research requires participants to process the stimuli globally at longer exposure times, we needed to make certain that the stimuli would induce global representations. And so, when creating the contours, we restricted the number of projections that our stimulus sinusoids contained.

As attention has been shown to influence the firing rate of neurons in the V4 region (Dumoulin & Hess, 2006), it may be necessary for us to test the participants' ability to recognize the contours under different conditions. We could test their performance both with and without a distracting task to determine whether selective attention plays a role in the time necessary for global processing of an image. Alternatively, the distracting task could be replaced with distracting noises which would draw the subject's attention away from the screen.  After determining the required exposure time under different conditions, we can compare the results and determine whether attention has an effect on this aspect of the perception and try to establish the strength of this possible effect.

Previous research indicated segregation between different visual functions. Early studies by Harvard Medical School imply that two primary structures, the magnocellular division and the parvo-interblob system are responsible for the different vision pathways (Livingstone & Hubel, 1987).  The researchers explain that the magnocellular division appears to be responsible for faster brain processing, while the parvo system works at a slower pace.  From their studies, the researchers believe that perceptual processing of form occurs in both regions, with each system performing a different part of the processing.  This parallel processing of form would allow the mind to perform calculations and interpret different aspects of the vision efficiently.  As the slower, parvo system is primarily responsible for gathering information for static object recognition (according to the two streams hypothesis), our study is primarily interested in this structure.  As the parvo system requires more time, by altering exposure times, we should be able to determine more about these two structures, as well.

Once we establish the time necessary for the transition from local to global image processing, it will be possible to determine more precisely the mechanisms involved.  By looking at previous research regarding Electroencephalography studies of image perception, we can then establish the brain regions involved.

*References*


Dumoulin, S. O., & Hess, R. F. (2006). Cortical specialization for concentric shape processing. *Vision Research* , 1608-1613.

Feldman, J., & Singh, M. (2005). Information Along Contours and Object Boundaries. *American Psychological Association* , 243-252.

Livingstone, M., & Hubel, D. (1987). Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth. *The Journal of Neuroscience* , 3418-3488.

Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research* , 2106–2127.

Tse, P. U., Martinez-Conde, S., Schlegel, A. A., & Macknik, S. L. (2005). Visibility, visual awareness, and visual masking of simple unattended targets are confined to areas in the occipital cortex beyond human V1/V2 . *Proceedings of the National Academy of Sciences of the United States of America* , 17178-17183 .

**Phoneme Features Project**

**Zachary Sandoval**

This project is identical to MacKay, D.G. (1970). (Phoneme repetition in the structure of languages. Language and Speech, 13(3) 199-213) except that the unit of analysis is phonological features. It consists of two studies.

As stated in the original study, the structure of phoneme repetition in Croatian and Hawaiian was found to be remarkably similar. In both languages, immediate repetition of phonemes as in AACHEN was very infrequent, but phoneme repetition after some degree of separation as in PROPER was significantly more frequent than chance expectation.

Also from the original study, the degree of separation for maximum probability of repetition was slightly different for vowels and consonants in both languages. This pattern of phoneme repetition was unrelated to syllable length word length or word frequency in these languages. The hypothesis was advanced that this pattern of repetition resulted from an evolutionary process, reflected not only in recorded phonological changes in the history of languages, but also in errors in speech, and phonetic changes at rapid rates of speech, all of which frequently involve repeated phonemes. Some of the questions considered were as follows: do vowels and consonants have identical patterns of repetition? Is the structure of phoneme repetition in different languages similar, i.e., is the structure of phoneme repetition a linguistic universal (as defined in Greenberg, 1963)? Does the probability of phoneme repetition depend on either word frequency or word length? Do syllabic factors influence the structure of phoneme repetition in a language? Is there evidence for evolutionary changes in the history or languages, which mould the structure phoneme repetition? Do repeated phonemes present a problem in natural speech production?

In this study, the frequency of feature repetition was calculated separately for Croatian consonants versus vowels in a way that allowed us to determine if repetition was greater or less than chance at different lags or degrees of separations between the repeated features. For example, immediate repetition of a feature in adjacent segments in a word was counted as lag 0.

Over 258 words with a phoneme length of 10 segments were selected from Bogadek (2000) using the following sample procedure: A set of five chapters was selected at random. The chapters were A, B, D, L and S. All words of length 10 in chapters A, B, D, L, and S were tabulated. We then tabulated repetitions of the features [+voiced], [-voiced], [+nasal], [-nasal], [+bilabial], [+alveolar], [+velar], [+stop]. [+fricative], in consonants in those words. For example, with # representing any segment, the repetition of [+voiced] in [gd##] was counted as separation 0; whereas, the repetition of [+voice] in [g#d#] was counted as separation 1, and so on. The frequency of feature repetition was then calculated for each gap length and averaged across all consonants in the sample as in Table 1b in MacKay (1970) in two analyses. In analysis 1, vowels were counted as a featureless gap of 1 and in analysis 2, vowels were counted as [+voice].

These are the consonant features of interest:

[+voiced]

[-voiced]

[+nasal]

[-nasal]

[+bilabial]

[+alveolar]

[+velar]

[+stop]

[+fricative]

These are the vowel features of interest:

[+voiced]

[+front]

[+back]

Similar procedures were then used to calculate the frequency of repetition for two vowel features: [+front] (e.g., i, e) and [+back] (e.g., u, o, a) for each gap length with averaging across these features as illustrated in Table 1a in MacKay (1970). However, only one analysis is necessary because consonants are always counted as a featureless gap.

The mean frequency of the features [+voiced] and [-voiced] and [+nasal] and [-nasal] etc. for consonants and [+front] and [+back] for vowels in Croatian was then calculated separately based on the frequency of consonants and vowels containing those features in the overall sample.

These possible subsidiary analyses were the same as the analyses above except that we selected only the words in chapters A, B, D, L, and S that did not contain either repeated consonants (N=?) or repeated vowels (N=?). We then tabulated repetitions of the features of interest, first for consonants in the first set of words, then for vowels in the second set of words. For example, with # representing any segment, the immediate repetition of the [+voice] voicing in [gd##] was counted as separation 0, whereas, the repetition of the voicing in [g#d#] was counted as separation 1, and so on. The frequency of feature repetition was then calculated for each gap length and averaged across all consonants in the subsample as in Table 1b in MacKay (1970). Similarly, for vowels, the frequency of the feature repetition was then calculated for each gap length and averaged across all vowels in the subsample as in Table 1a in MacKay (1970).

To determine whether the pattern of repetition as a function of lag differs for feature repetition versus segment repetition, we compared our feature repetition results with the frequency of segment repetition independently determined in MacKay (1970). That is, we produced tables of each feature type for vowels versus consonants as in Tables 1a and 1b in MacKay (1970). We then translated the tables into figures resembling Fig. 2 in MacKay (1970) except with separate figures for vowel features versus consonant features. We then translated the vowel and consonant frequencies in Table 1ab in MacKay (1970) into two separate Figures, one for vowels, the other for consonants.

The second part of this study was the same as the first part except that the sample consisted of at least 2015 Croatian words selected as follows: four words from each page of Bogadek (2000) were chosen at random using a random number table. The words were then sorted by length, and the frequency of a feature repetition as a function of gap length was calculated for consonants and vowels separately for each word length and each feature type .The goal was to produce tables for each feature type resembling Tables 2 and 3 in MacKay (1970). We then translated the tables into figures resembling Fig. 5 in MacKay

(1970) except that vowel feature figures were separate from the consonant feature figures. Finally, for comparison, we translated Fig. 5 in MacKay (1970) into two separate figures, one for vowels and another for consonants.

The original study of phoneme repetition in Croatian revealed that immediate repetition of phonemes as in AACHEN is less common than would be expected by chance. However, with wider degrees of separation as for the repeated r's and p's in PROPER, the probability of repetition significantly overshot chance expectation. This overshoot can be viewed as a law of latent alliteration or phoneme harmony in the structure of languages. Of course this is by no means a deterministic law, but only a statistical one, applying in general but not in every word.

These findings were replicated in Hawaiian—an unrelated language with a completely different phonological structure from Croatian. The corroboration suggests that the pattern of phoneme repetition is a universal feature of human languages.

*References*

Bogadek, F.A. (2000). Croatian-English dictionary

MacKay, D.G. (1970). Phoneme repetition in the structure of languages. Language and Speech, 13(3) 199-213

**Moral Psychology**

**Gautam Patel**

*Background*

The study of morality has generally been left to the discipline of philosophy, as it was in that area where philosophers discussed the nature of humans, proposed hypothetical "right" and "wrong" situations, and made how we come to know of things as morally correct or incorrect. In general, these "truths" regarding morality were simply based on their own perception and limited interactions. They were difficult to generalize, and unscientific in nature (Moral Psychology, 2006).

The study of morality within psychology is focused not on the absolute nature of actions (determining exactly what is right and wrong), but more on the development of morality, or the idea of whether morality is universal (Rai & Holyoak, 2010), or, as the research of the UCLA Reasoning Lab works on, how can we change similar situations to change perceptions of morality. It also has a more rigorous, empirical method when compared to philosophy.

*Introduction*

One of the more interesting questions in moral psychology is to understand whether or not there are certain moral principles which are innate, or if morality is learned. In life we can see both types of situations. There seems to be an innate way of knowing that killing another human is wrong, or that taking things which do not belong to you is wrong, but many other ideas are not so natural. For example, I grew up in a Hindu, vegetarian household, and the idea that killing an animal to eat it feels wrong. Not just wrong, but naturally & fundamentally wrong. But this is clearly not the case with most of the world. This example shows how morality can be different & learned for others.

Many moral psychologists believe that much of morality is innate and consistent across the human species. Similar to the Universal Grammar theory of language acquisition developed by Noam Chomsky, which proposes that all humans are born with all of the necessary tools to learn any language (Chomsky, 1965), and do so automatically and similarly, many psychologists believe that there is a similar Universal Moral Grammar (Rai & Holyoak, 2010). This Universal Moral Grammar has normative moral principles, and all humans reason and assess morality in relation these innate principles, but with different parameters (determined by culture). So a large part of morality is innate, and a significant amount is also learned from society.

*Previous Research (UCLA Reasoning Lab)*

A lot of the research regarding Moral Psychology has been done using hypothetical problems such as the infamous Trolley Problem (Thompson, 1985). Using this model as the basis, and making minor variations has allowed researchers to tease apart the effects of potential moral principles through paired problems designed to isolate that specific moral principle.

The generic Trolley Problem is a situation in which there are 5 people tied down to the train tracks a train is about to go through. On another track, there is a single person tied down to the track. You are by the switch which determines which track the train will go through. What should you do? When looking at the traditional utilitarian point of view, it makes sense to flip the switch because the loss of 1 life is less than the possibility of 5 lives lost. Flipping the switch is not only ok, it is actually the morally correct action according to this view. Alternatively, by flipping a switch, you are directly “causing” the death of someone, in a situation you don’t have to participate in. The reasoning behind this situation is that flipping the switch makes you a participant in the person’s death, while not flipping the switch, and letting 5 people die is actually ok because you are not participating directly.

Much of the research done by the UCLA reasoning lab is related to the presentation of the question as opposed to the question itself. Previous research has indicated that framing effects and other small biases can make large differences in judgments (Rai & Holyoak, 2010). For example, the sin of omission is viewed much more favorable than the sin of commission. Also, when assessing moral dilemmas, the dilemma that appears later is often viewed more unfavorably, even though it should be the same. Another example is that judgments regarding the problem are different when the question is posed as number of people dying versus number of people living.

The Alternative Framings of the Trolley Problem study by Rai & Holyoak (2010) tried to find out if the methods of presenting paired moral dilemmas to identify domain-specific principles from a domain general bias can actually support the claims of domain specificity in moral judgment. It is essentially trying to see if the experiment is really measuring what it claims to be measuring, or if framing has a large part to do with it. In the first experiment of that study, participants were separated into two groups, one which was directed to list 7 reasons for the employee to take the proposed action, and another which was directed to list 2 reasons. The group which listed more, was, paradoxically less likely to to agree with the action. In the second experiment, the experiments were interested if relative saving was an important factor. Previous research (Kahneman & Tversky, 1984) found that people were more willing to drive 20 minutes to save $5 on a $15 calculator than $5 on a $125 jacket, even though the absolute value is the same. They wondered if this effect had a parallel in the Trolley Problem. They created two separate problems in which a car of two people could stop the trolley and allow 8 of 10 or 8 of 40 people to be saved. The results show that there was a significant portion of people who chose 8 of 10 people, showing that they value to relative value, not just absolute value.

In general, the study found that it was not possible to empirically distinguish whether influences on moral judgments are based domain specific principles or on the biases find in the contexts of the question.

*Current Research*

The study that I helped work on at UCLA was related to moral psychology because it was evaluating how much of a quantity would be required to make two immoral actions equal. It was mostly interested in the idea of relating different types of immoral actions to one another to see if there is proper reasoning behind their decisions. For example, a question would be John cheated on his medical boards. Allen was at a jewelry store, how much jewelry would Allen have to steal for his actions to be like John's?. The study was already planned, and researched. What I helped out with was to creating and counterbalancing the survey portion of the experiment. The survey was created using the Qualtrics online survey program. I created a multiple choice survey, and counterbalanced the questions using the randomization and block features of the Qualtrics software. I also hard coded the responses to numerical values to keep a consistent measuring system for the results.

My research apprenticeship was actually far more extensive than just this experiment. In fact, the majority of the time was spent doing similar tasks with other experiments, like the Analogical Reasoning experiment. I also helped create and set up other testing materials like monitors and testing software.

*References*


Chomsky, N. Aspects of the Theory of Syntax. MIT Press, 1965.

Moral Psychology. (2006.) Stanford Encyclopedia of Philosophy. Retrieved December 8th,2011, from http://plato.stanford.edu

Thomson, J.J., 1985, “The Trolley Problem”, Yale Law Journal, 94: 1395-1415

Rai, T. S. and Holyoak, K.J. (2010). Moral Principles or Consumer Preferences? Alternative Framings of the Trolley Problem. Cognitive Science, 34, 311–321

**Plus Or Minus 7? Is That All The Information We Can Hold?**

**Mark Corre**

Abstract

Miller reasoned that 7 plus or minus 2 is all the information we can store in our short-term memory but is our short-term memory just related to a simple set of numbers. There is a complex relationship between our short-term and long-term memory. It is possible that long-term information helps us decrease search time by showing us more available representatives of a category. Here I purpose an experiment where participants are given stimuli in which they are asked to group into a category. Those categories which more representatives should reduce the amount of search time one has to do to find the correct solution.

Miller purposed that short term information is stored in discrete units called bits (1956). A bit is enough information to determine what is needed to make a decision between two equally likely outcomes. Miller showed that in short term memory the maximum amount of bits we can reliably make a decision on is 7 plus or minus 2. Is this all there is to memory though? Miller also showed that those who are adept at music score much better than the average person. So is 7 plus or minus 2 all there is to memory?

Stewart, Brown and Chater (2005) claim that individual items in a category are just the differences between the category itself. If the category is simple then there could be a lot of items because there is only a few dimensions for them to vary. Stewart et al. also find it strange that when trying to identify stimuli that only vary along one stimuli is difficult, but if we know the category it is very easy to tell what goes it in the category.

Goldstone (2000) suggested that we are more efficiency in memory when we are able to categorize items. Goldstone presented participants with an lines that only varied in the curvature of and showed that once people learned what the categories where they were able to decide which group each line belonged to easier. If the spatial distance between the stimuli increases performance is decreased. Goldstone suggests that this is because it is harder to group the lines into one coherent category. Goldstone also showed that we are able to more efficiently learn a category if there is a gap in the spacial distance of each category. Thus, the quicker we can group objects together the quicker we can develop a mental representation and store each object. So it is not just our short term memory but our long term memories that are categorized. If this true, then if certain items are well categorized in our long term memory shouldn't it be easier to memorize short term information that can be easily grouped into the long term categories? Here I purpose a test to show that long term memory has a great influence on our short-term memory.

*Discussion*

An example purposed being asked to remember the presence of two animals. One is easily recognized as fitting into the “bird” category another is an invertebrate see dweller. There are reasons for each representatives of the categories to be more easily memorized. The see dweller might be more easily memorized because it is an new category and thus more salient. A new category would take more processing to make and thus would have a longer effect in the brain. This could be a proper conclusion but then why are subjects more easily able to discriminate specific characteristics between a more established category(Goldstone 2000). It may be possible that the processing not used to find and categorize the item is used to store additional information, in other words, to fill our 7 plus or minus informational hold. It may also be possible that a well established category has more candidates, meaning, more features that are able to distinguish the group from other equally likely groups. Since there are more bits already stored for birds it becomes easier to store more bits for new birds(or more knowledge on being able to distinguish this new bird from more separate categories). It also should become faster to group this bird into the bird category. If a more established category allows us to categorize faster, then it is possible that search time is related to bit search time. More established categories should have faster search time. I purpose we show a series of stimuli and ask a participant to group them into one of 4 possible categories. The more we show representatives of these four categories as stimuli the faster it should become for individuals to categorize new stimuli. As the amount of stimuli increase it should also be easier for subjects to match stimuli more correctly to more categories. In a second part of the experiment with 4 more categories that change the groupings of the stimuli, individuals should also be able to match new stimuli more easily to the new categories then they would old stimuli. Even though old categories are more established in memory because there is +1 representative for each category when a new stimuli is presented they should be easier to group. If data confirms that new categories are easier to group it may be possible that we categorize based on individual indicators of the category rather than how well a stimuli relates to a strong representative of the group.

*References*


Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26* (1) 86-112.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Stewart, N., Brown, G.D.A., & Chate, N. (2005). Absolute Identification by Relative Judgment. *Psychological Review, 112* (4), 881-911.

## Spatiotemporal Boundary Formation: Theory and Modeling

**Maxwell Mansolf**

### *Background*

The method by which the brain composes a functional spatial representation of the environment, given visual information that is often sparse, incomplete, and ambiguous, is one of the largest theoretical and computational mysteries in the field of cognitive science. Optical illusions such as the Kanisza triangle (Kanisza, 1955) provide evidence that the human visual system can perceive figures that, while not present in the perceived image, are partially defined by other figures. Many of these illusory figures are perceived through the brain's ability to compensate for partial occlusion: even if one figure is in front of another (such as a picket fence in front of a dog, for example), and the viewer can only see a part of the occluded figure (in this case, the dog), if enough visual information is present, the viewer perceives the occluded figure as a whole. With illusory figures, such as the Kanisza triangle, the brain perceives the edges that are present, and we perceive a complete figure, with illusory contours, which we don't see but nonetheless are aware of, completing the parts of the figure that are unseen.

While these illusory contours were first described in the literature by Schumann (1904) and further investigated and described by the Gestalt psychologists in the early 20$^{th}$ century, an explanatory framework for why the brain perceives these illusory contours did not arise until much later, with Shipley and Kellman (1991). Integral to this framework is the concept of edge completion: if an edge (defined by a sudden change in color, intensity, texture, or some other perceivable quality of the figures) begins at one point, and ends at another point, it can be perceived even if a part of the edge is not present in between those points. The edge can be a straight line, or a curved line, but as the edge becomes more complex, perceiving it requires that more information be present to the observer.

### *Problems with Edge Completion*

The theory that illusory contours are primarily perceived through the continuation and completion of partially defined edges tends to hold very well for most static illusory figures, and remains a powerful theory for illusory figure completion, but there are cases in which it fails to explain what is perceived. For example, in experiments in the Kellman lab at UCLA, subjects were presented with a white screen with black dots scattered throughout it. In the original version of the presentation, a white line, of thickness equal to that of the dots, length such that it stretched from the top of the screen to the bottom, and slope slightly offset from vertical, was drawn on one side of the screen (right or left) and moved across the screen (right to left or left to right, respectively) until it reached the other side of the screen. Because the line was the same color as the background and had no border, it was indistinguishable from the background. However, as the line moved across the screen, it partially or completely occluded the black dots that were on the screen, and through this limited and scattered visual information, participants could determine the slope of the line fairly accurately. These results agreed with Shipley and Kellman (1991), because some of the dots were partially occluded by the line, and an edge with the same slope as the illusory line was formed between the line and the partially occluded dot. From this information, participants could determine the slope of the line through the edge completion paradigm outlined in Shipley and Kellman (1991). However, another version of the experiment was conducted, in which if a dot intersected the line at all, the entire dot became white and invisible, and otherwise the dot would appear in its normal visible black form. In this experiment, no edge data was provided, but participants could still determine the slope of the illusory line, although with less accuracy than in the original experiment. These results did not agree with Shipley and Kellman (1991), as local edge information was not necessary for participants to perceive the illusory line, and the Kellman lab attempted to define a new model for illusory contour perception, one that does not require that local edge information be present.

*Shipley & Kellman (2004) – The New Model*

In the newest model for the illusory line phenomenon in these experiments, edge orientation information is not required, and the slope of the illusory line is instead calculated using local motion signals (read: the timing and sequence of sets of dots appearing and disappearing). More specifically, Shipley & Kellman theorize that the slope of an illusory line in such an experiment can be calculated from 3 dot occlusion events (for example, first dot 1 is occluded, then 2, then 3), by first calculating the local velocity of the line between the two consecutive pairs of dots (ex. the distance between 1 and 2 divided by the time between the 1 and 2 occlusion events [$v_{12}$], and likewise for dots 2 and 3 [$v_{23}$]) and the angle between each of these dot pairs relative to horizontal (ex. the angle of the line from 1 to 2 [$\varphi_{12}$], and likewise from 2 to 3 [$\varphi_{23}$]), and from these values calculating the slope of the illusory line via the formula

$$\Theta = \tan^{-1}\left(\frac{v_{23} * \sin\varphi_{23} - v_{12} * \sin\varphi_{12}}{v_{23} * \cos\varphi_{23} - v_{12} * \cos\varphi_{12}}\right).$$

where θ represents the slope of the illusory line. This formula was derived through simple trigonometry, and for the sake of brevity, the proof is not included here. Note again that this model doesn't require any edge information at all, and simply approximates the slope of the line via the locations and timings of the three dot occlusion events.

*Applying the Model: Preliminary Research*

The model predicts that, given only three dot occlusions, the slope of the illusory contour should theoretically be calculable. However, research by Gennady Erlikhman in the Kellman lab has shown that this is not the case in human subjects. In the standard experimental procedure, participants are presented with a field of dots over which an illusory line passes from left to right or right to left, causing the non-partial occlusion events as it travels. The line is offset slightly from vertical, and participants are told to determine whether the line is tilted slightly to the left (counterclockwise from vertical) or to the right (clockwise from vertical). Ease of perception of the line was measured by calculating the angle of offset required for subjects to accurately determine the slope of the line 75% of the time. The adaptive staircase method used was the PSI method (Kontsevich & Tyler, 1999). For the sake of convenience, I will use the terms "easy to perceive" and "hard to perceive" to refer to conditions in which the threshold value estimated by the PSI method is high and low, respectively.

In one incarnation of this experiment, to be henceforth referred to as the "density experiment", the size of the display and the speed of the line remain constant across experimental conditions, and the number of dots on the screen varies between subjects (numbers of dots used: 9, 16, 25, 36, 49, 64) perform at levels greater than chance (p(correct)>0.5) when 9 dots are present, and 16 dots are required for participants to perceive the slope of the line reliably (p(correct)>0.75). Once there are at least 16 dots present, including more dots improves performance, but the change is never as stark as that from 9 to 16 dots, leading to the conclusion that as long as there are at least 16 dots present, subjects can usually perceive the line, whereas with fewer dots the line is sometimes not perceived at all, and with more dots the slope of the line is more easily perceived.

In the other two experiments, instead of manipulating the number of dots on the screen, the area of the viewing aperture (with 100% area being the same condition as the 64-dots condition of the density experiment, and the screen shrunk horizontally for the other conditions) and the speed of the line (with the fastest speed of the line being the same condition as the 64-dots condition of the density experiment, and still frames added between each movement frame in the other conditions) were manipulated. The results for the area experiment were that, as long as a sufficient amount of area was shown, such that the number of visible dots was above 16, the participants' perception of the line wasn't improved by increasing the size of the viewing aperture; in other words, if participants had enough information to perform the task, adding more information in the form of more dots in the same density failed to significantly improve performance. The results of the speed experiment were similar to those of the density experiment; slowing down the movement speed of the line reliably decreased the perceptibility of the line. However, as the line moves faster, performance at this experiment increases in a more linear fashion than in the density experiment, demonstrating that in general, as the local motion signals become more frequent, the percept of the line becomes stronger. It is worth noting that at very high speeds, performance drops off dramatically, due to problems with high-speed stimuli that aren't specific to this experiment, so very high speeds were not tested in the speed experiment.

### *Modeling the Phenomenon*

Overall, the results of the three experiments agree with the newest Shipley and Kellman model; speed of the line and distance between the dots are important factors affecting the perceptibility of the illusory line, but with those being equal, and as long as enough dots are present, adding more dots does not improve performance. The next step in the project was to model the phenomenon computationally, and see if the model can predict human performance at a computational level. The modeling was done in MATLAB.

In the original model, the slope of the line can be calculated perfectly for every dot triplet; for humans, this is not possible, due to noise in the visual input to the nervous system and within the nervous system itself. As a result, the MATLAB model adds error to the velocity of the line (by adding normally distributed error to the time offset between dot events, proportionally to this time value and a constant free parameter) and to the angle between the dots (by adding normally distributed error to the distance between the dots, proportionally to this distance value and a different constant free parameter). To determine how much error should be added, because results from trial to trial varied dramatically in the simulations (due to the low-level implementation of the model and the flexibility of the adaptive staircase procedure), modeling involved fitting the error coefficients by hand. The slope of the illusory line was calculated, as per the Shipley and Kellman model, for each dot triplet, and then these estimates were averaged across all dot triplets to determine whether the slope of the line was positive or negative. Because when many dots are present, the large sample size renders this estimate very accurate even for high error values, only 16 consecutive dots are taken from the list of dot events in the final versions of the model (this number was taken from the number of dots a human requires to perform the task reliably).

*Modeling: Results and Future Directions*

Through much tedious testing and retesting, suitable constants were found such that the human data from the area and density experiments was duplicated almost perfectly by the MATLAB model. This is remarkable and promising because the model operates at a very low-level (calculating slope estimates of the line individually for each dot triplet), and can predict human performance (a high-level phenomenon) to a high degree of accuracy, as shown by the very close proximity of the threshold values generated by the model to those from the human data. However, there is one caveat; using the same constant parameters as were used in the density and area models does not reproduce the human data for the speed experiment. In other words, the model overperforms in the speed experiment, and the threshold estimates are lower than those in the human data. The modelers believe that this is because in the area and density experiments, a black aperture was drawn around the dots and the line began from the edge of the screen (either left or right). In the speed experiment, there was no such aperture, as it allowed heuristic methods to be used to determine the slope of the line by attending to the dots at the edges of the screen, and the line began its movement at a random position within the screen, for similar reasons. The next step in this project, and the step that will be taken next academic quarter, will be to run new versions of the area and density experiments on human participants using these same limitations, such that the data between the three experiments are comparable, and the model can be calibrated such that the same set of parameters can predict all three data sets.

*References*


Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. Rivista di Psicologia, 49 (1). 7–30.

Kellman, P. J., & Shipley, T. F. (1991). A theory of Visual Interpolation in Object Perception. Cognitive Psychology, 23, 141-221.

Schumann, F. (1904). Einige Beobachtungen iiber die Zusammenfassung von Gesichtendrucken zu Einheiten. Psychologische Studien, 1, 1-32.

Shipley, T. F., & Kellman, P. J. (2004). Spatio-temporal Boundary Formaiton: the Role of Local Motion Signals in Boundary Perception. Vision Res. 37(10), 1281-1293.

## Neuroscience and Metamemory

**Shruti Ullas**

Some of the research being conducted in the Bjork lab is focused on metacognition, particularly how people perceive and judge their own learning. Related research has explored the neural underpinnings of these judgments of learning, but more specifically, judgments of memory, or metamemory. In one study, by Chua et al. (2008), the researchers used fMRI to pinpoint active areas while subjects made prospective and retrospective judgments of their memory. In another study, Yokoyama et al. (2009) found that levels of activity varied with how reliable a subject's confidence rating was. Both studies identified neural activity associated with specific metamemory-related processes and made insights into the mechanisms required for these tasks.

Chua et al. (2008) investigated neural correlates of retrospective and prospective judgments. The study began with face and name memory task, where subjects learned faces paired with a name. While being scanned, when presented with faces, they made prospective, or feeling-of knowing judgments (FOK) of how well they believed they could recall the name in the future. This was followed by a forced-choice recognition task, where the faces were presented again but with three name options that the subjects had to choose from. The subjects had to choose the name they believed to be associated with the face presented and then provided a rating of how confident they felt about their choice (CONF). Between the recognition task and confidence rating, subjects also gave attractiveness ratings, so that there would be a clear boundary between the two tasks and overlapping activity would be reduced, since subjects could have been thinking about how confident they were while choosing the name.

The authors found several areas that showed an increased amount of activity during both tasks, others that showed a decrease, and other areas that were specific to either CONF or FOK judgments. During both CONF and FOK judgments, there was increased activity in the medial prefrontal ration, mid-posterior cingulate, lateral parietal, and lateral temporal regions. The occipital region, lateral inferior frontal and dorsal medial prefrontal regions showed decreased activity during both tasks. The authors noticed that regions related to visual processing showed a decrease in activity, which could represent how the memory tasks required shifts away from external attention to an internal focus. They also observed that regions associated with stimulus-driven attention also showed decreases in activity, such as the right inferior prefrontal and dorsal medial prefrontal regions. This observation also supports their conclusion that these metamemory tasks direct attention away from the external stimulus to internal processes.

The authors also found that during FOK judgments, some regions showed activation that did not during CONF judgments. These regions were the medial parietal, fusiform, right superior temporal, and hippocampal formation regions. Even though both tasks were related to metamemory, they were still different types of judgments, so it made sense that there were areas that were active during one task but not the other.

The experiments by Yokoyama et al. (2009) also studied metamemory and confidence judgments, but additionally looked into the types of activity related to more reliable confidence ratings. In this study, subjects were given a metamemory task and a control task. They were asked to first memorize a an image of a dot-pattern, followed by a delay, and were then asked to choose the image they memorized from six image options. They were then asked to give a confidence rating of either "high", "middle", or "low". As a control, they were asked to complete brightness-discrimination trials, where instead of providing a confidence rating, they were asked to choose the brightest word out of three options; either "left", "center", or "right". These trials elicited baseline activity and controlled for language requirements, visual input, response choosing, and motor processes.

Overall, they found there were greater amounts of activity in certain areas while making confidence ratings than during the control. There areas included the right frontopolar cortex, bilateral insula, left temporo-parietal junction, right anterior cingulate, right superior frontal regions, and right superior temporal regions. They believed that these results were consistent with evidence showing that the anterior cingulate is recruited during conflict monitoring and that temporo-parietal regions are also active during monitoring and thinking about cognition as well. They also found that subjects who provided more reliable ratings of confidence showed differences in activity of the posterior-dorsal region of the frontopolar cortex. They concluded that overall, the right frontopolar cortex is important in making reliable confidence ratings.

Both of these articles investigated the brain areas required in making metamemory judgments. They were able to connect activity in areas of the brain with judgments of memory, and the areas implicated were consistent with previous research findings. The articles noted that other researchers have connected some of the same areas with other cognitive functions that are related to making judgments, such as attention and monitoring. They also observed that areas showing decreases in activity which were related to functions that are less important while making judgments, such as visual processing. Overall, the findings of both articles were able to make insights as to what our brains are doing when we make judgments of memory.

Although these articles utilized brain imaging techniques, their research is still relevant to the research being conducted in the Bjork lab. Studies in the Bjork lab have also investigated how people judge their own memory, by asking subjects to provide judgments-of-learning, or JOLs. The use of JOLs is very similar to the ratings given by the subjects in the Yokoyama and Chua studies. Essentially, an FOK, as described in the Chua article, and a JOL, are both asking subjects how well they feel they have learned or remember something. The key difference may be that a JOL is usually on a numerical scale, while an FOK is a categorical measurement. The Chua article also looked at the differences between retrospective and prospective judgments. Similarly, in some studies at the Bjork lab, subjects are sometimes asked to give JOLs during the task, or are asked about their overall impression of their learning during post-test questionnaires. Overall, the Chua article, the Yokoyama article, and the current research in the lab share many similarities, as they are all trying to understand metamemory, and the tools and strategies involved.

After working in the Bjork Lab as a research assistant, I feel that I have learned a great deal about conducting research. I had been an intern in other labs prior to working here, but never had the opportunity to work with the subjects. It was eye-opening to observe how other people approach the tasks, and to hear about their thoughts about the research as well. When skimming post-test questionnaires, it was interesting to see other people's thought processes behind their learning strategies compared with what is expected to occur, or even any strategies I might have used.

I also enjoyed the lab meetings. Every week during Minifog, getting the opportunity to hear from Bob and Elizabeth Bjork was very special, since they are such important contributors to memory research. It was also great to be a part of the behind-the-scenes of research, when graduate students brought in the results they had found so far. When other lab members gave input or ways to improve their design, it made research seem a little less daunting. It was nice to see how it takes time and many attempts before the design of a study is perfected. Before coming here, when I read papers or articles, it always seemed as if the researchers came up with great ideas so quickly and easily, but the lab meetings showed how much more effort really goes into coming up with studies. I enjoyed attending the Knowlton Lab meetings as well. I found the research being talked about was particularly interesting because of the emphasis on neuroscience. I really enjoyed hearing about the fMRI, TMS, and EEG studies because of how they tied psychological processes and neural processes together.

In the upcoming quarters, I hope that I can assist with the fMRI research beginning in the Winter. I hope to do research using fMRI/MRI one day, and I would like to learn more about using imaging tools and how to design experiments that make use of these tools.

**Metamemory: What Do We Know, What Do We Not Know,**

**and How Can We Tell the Difference?**

**Zachariah J. Merrill**

Being able to precisely monitor what you are able to remember is an essential component of memory and your ability to learn. It's because of this that metacognition and, more specifically, metamemory is such a significant field of study in psychology. It's important to know what you *actually* know to better direct future practice. Much research has gone into deciphering just how good our metamemory is, with mixed results across the board. A common measure of one's memory monitoring is through the use of JOLs, or Judgments of Learning, which ask the participant to express how likely they are to remember something (Sungkhasettee, Friedman & Castel, 2011). This can be done on an item-by-item basis or through bigger chunks of information at a time. Ideally, this will give insight to how well people can accurately observe their memory performance. Accurate judgments of learning, like people predicting higher levels of recall for items more likely to actually be remembered, and inaccurate, can speak wonders about an individual's interpretation of their learning strategies. With greater knowledge about how peoples' memories work, it's possible better strategies can be implemented for better memory performance.

A paper focused on participants' judgments of learning when faced with desirable difficulties, illustrates an example when our metamemory isn't as accurate as we would hope (2011). In the study, participants studied upright or inverted words for later recall and also made judgments of learning for each word specifying how likely they would remember each word. Some words were displayed in inverted fashion to create a desirable difficulty. A desirable difficulty essentially occurs when the target that's trying to be remembered is made more difficult to process, thus leading to more time and deeper processing of the item, and in turn making the object more likely to be remembered. So in essence, by making the word harder to process, one increases the likelihood that word will be remembered. Since the inverted words were used as a desirable difficulty, participants should remember the inverted words more often than the upright words, which ended up being the case (2011). However, participants' judgments of learning for the item did not show any discrepancies between the two word presentations. So even though we are significantly better at learning those words with desirable difficulties, why don't we have this knowledge as we go through the test, and use it to more accurately make predictions?

It's possible judgments of learning rely more on a feeling of fluency, defined as the relative ease of processing and retrieving information, far more than any actual knowledge of our memory capabilities. However, if that were solely the case, one might predict the upright words would be given higher judgments of learning than the inverted words due to a higher perceived fluency, and this was not the case (2011). The study goes further to show that even after repeated trials of the task, when subjects would consistently remember the words presented in inverted fashion more frequently than upright words, judgments of learning never increased for inverted words across trials. Even after the participant is given proof that one type of presentation is better than another we do not rectify the error in our predictions. Are subjects not noticing the difference? Is this knowledge obtained and then again forgotten when the next set of words and judgments of learning are required? Humans are fairly apt at inductive logic so it would make sense that after performing consistently better with one presentation type of the word, a person can reasonably induct that this trend will continue for future trials and adjust their predictions accordingly.

Judgments of learning and knowledge about peoples' comprehension of their own memory capabilities are extremely important because this can help identify people's *choices* about how to remember something or what they feel most confident in remembering. The practical applications of this are obvious. Figuring out the most effective implementation of educational programs or job training plans would significantly improve the quality of teaching and the efficiency of learning.

There are more examples of research which show conflicting evidence on people's knowledge of their memory expressed by peoples' preferences in how they remember material they will be tested on later. One study presented individual words in either a massed or spaced presentation, and asked subjects to rate the likelihood they would remember that word (Zechmeister & Shaughnessy, 1980). To elaborate, a massed presentation would show the word, then immediately show the word again, and conclude by asking how likely they were to remember the word. Whereas a spaced presentation would show the word, then show the word again 3 to 5 words later, and ask how likely subjects were to remember the word (1980). Numerous studies have shown, this one included, that spaced practice leads to better long term retention of target material than massed practice, and in turn better performance on later recall tests. However, in this study, subjects judged massed items as more likely to be remember than spaced items, a poor prediction of their actual performance. Is fluency again a factor? The word is shown twice in a row, giving more than enough time for the participant to fully process the item and increase the ease of processing. However, after repeated trials and continual evidence of those words in the spaced condition being recalled more accurately, subjects still don't adjust appropriately.

To further complicate the issue there is evidence illustrating that people can recognize the benefits of a tool like spaced practice, and implement this tool when the target to be remembered is more difficult, or more valuable for them to remember (Toppino & Cohen 2010). This study showed participants target words to be remembered, and gave the option of restudying the word immediately (simulating massed practice), restudying the word later (simulating spaced practice), or not restudying the word. Across multiple experiments the finding was replicated that as word difficulty increased, people's preference for spaced practice increased, insinuating that the more effective strategy was utilized for the words that it was necessary for (2010). This is a confusing result considering the previously presented research that suggested we have very poor knowledge of the benefits of spaced versus massed practice, and may even prefer massed more. The result is nicely explained through the *agenda-based regulation model*, which proclaims that study regulation is goal oriented, which makes sense that a more effective study schedule would be applied for more difficult words *and* more valuable words because that is the best way to achieve your goal (2010). The hitch is in the idea that we are choosing the spaced study because we know its better, since more evidence supporting this claim in diverse domains need be explored.

So where does that leave the issue? Admittedly cloudy. While evidence presented here probably asks more questions than gives answers, it portrays how complex and difficult a judgment of people's metamemory can be. To make matters worse, metamemory is not something that can be judged explicitly by just asking participants, as shown through poor judgments of learning. The many complications and confusing results can lead to a variety of opinions on what we know about our memory and how well we can use that knowledge to our benefit, but because there is so much more to uncover, it leaves researchers with a multitude of avenues to proceed in for the future. Part of what makes the field so exciting is there is still so much more to figure out, and determining the best way to do so will be crucial in order to succeed.

I can only begin my reflection of my time in the Bjork lab with a feeling of sadness that after this quarter I will be leaving. I joined the lab a mere nine months before and it feels as if I was just getting started. I can honestly say my time working with the lab has significantly enhanced my stint in the cognitive science department and thoroughly enriched the way I think about prominent issues in my major. Before I joined the lab, I was merely in the cognitive science major, learning about different theories and memorizing results of different psychology papers. After joining, I was actually a part of the major. I was no longer just a student but actively involved in new cutting edge research that I can say I was a contributor to and the value I place on that experience cannot be expressed.

Journal Clubs for the lab forced me to read source material with a greater understanding, and critically think about important issues going on in the field around me. It challenged me to not only read the material, but use what I had learned and apply it in a new and insightful way that benefited other group members as well. I was able to lead a discussion on a difficult psychology paper, and actually do pretty well! I can, without hesitation, say that I am a far superior psychology student than I was before thanks to these weekly meetings.

As for any future research assistants you may have, I encourage you to include them and ask for opinions as much or more than you already do. That was one of my favorite things working for you in the lab was the thought that I too was bringing something to the table. It's rather intimidating as an undergrad student to make suggestions on a research project run by graduate students and mentored by memory experts like Bob Bjork, but I never felt ashamed giving my opinion to you and I feel that's an invaluable experience for all undergrads to have for them to truly discover if they want to continue on in this line of research. I think you realize that everyone can bring something new and insightful to the table which is important and hopefully beneficial for your progress as well. I really enjoyed my time working with you and I can't thank you enough for the opportunity!

*References*

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychon Bulletin Review*, *18*, 973-978.

Toppino, T.C., & Cohen, M.S. (2010). Metacognitive control and spaced practice: clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1480-1491.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*(1), 41-44.

**Implicit Verb Causality Inferences from Covariance Information**

**Zenon Q. Anderson**

The implicit verb causality phenomena is the concept that the type of interpersonal verb used in a phrase may determine the causal responsibility to either the grammatical subject or the object. Bye's study (2011) focused on two verb types: action verbs and state verbs. Action verbs are typically voluntary behavioral or physical actions (i.e. punched, praised). State verbs on the other hand are typically involuntary and are usually emotional or mental states (i.e. impressed, believed). The form of these verbs can indicate whether the subject or the object of a phrase involving the verb is causal. For action verbs there are two subgroups: agent-patient (AP), and agent-evocator (AE). Agent-patient verbs place the causal role on the agent such as "Mary punched Sue". Agent-evocator verbs place the causal role on the object such as "Mary praised Sue". A similar rule applies to state verbs, but the two subgroups are instead: stimulus-experiencer, and experiencer-stimulus. The stimulus-experiencer state verbs, like the agent-patient action verbs, place the causal role on the subject (i.e. Mary impressed Sue). The experiencer-stimulus state verbs are similar to agent-evocator action verbs because they both place the causal role on the object (i.e. Mary believed Sue). In Bye's study (2011), the experimenter examined where causal responsibility lies in the verb.

There are two hypotheses as to how people learn through implicit verb causality. The first hypothesis, the lexical hypothesis, is that dispositional adjectives are drawn from verbs, which influence the inherent causal weightings by semantic activation. For example, for the given phrase "John influenced Jane", we are able to conclude that John is the agent and Jane is the patient. We would be able to make this inference due to the fact that "influence" may implicitly draw out the word "influential", which is attributed to John. The phrase "John influenced Jane" would therefore be semantically identical to "John was influential to Jane". We would then be able to identify the relational roles involved in the sentence by looking at the reconstructed phrase. The implication of this hypothesis is that verbs themselves influence the causal weightings of the relational roles.

The second hypothesis, the covariation hypothesis, is that "it is the information pattern… that give[s] rise to causal weightings of the various verbs" (Brown & Fish 1983). For example, for the give phrase "Ethan admires Emily", we can conclude that Emily is the stimulus and Ethan is the experiencer. We are able to make this conclusion because of the covariation between "admires" and "admirable features". The phrase "Ethan admires Emily" would be then semantically equal to the phrase "Ethan admires Emily because of her admirable features". The implication of this hypothesis is that verbs do not influence the causal weightings of the relational roles, but rather draw their causal attribution from the covariance of features in the object being examined.

In Bye (2011), the researcher attempted to address this problem by looking at both hypotheses. The experiment sought to use cognitive psychology methods to test whether covariation of information influences causal weighting. There have been prior experiments that studied implicit verb causality, but the main difference in Bye's experiment is the use of novel pseudo-words in place of known verbs. There are two standing hypotheses as to why implicit causality occurs: the lexical hypothesis, and the covariance hypothesis. In order to disprove the lexical hypothesis, Bye chose to use novel pseudo-words, such as "glorp", in order to eliminate any priming of the word that subjects may have already carried. By using these novel pseudo-words, the chance that priming may have influenced weighting of relational roles is eliminated, which had not been done in any prior literature.

In the experiment, subjects were asked to take a seat in front of a computer where they were asked to follow the instructions on the screen and to answer various yes or no questions. The subjects were told that they were "Earth's leading anthropologist, sent to a recently discovered planet to study the language and culture of its alien inhabitants." Their task was to examine the use of an alien word, and then subsequently decide in what scenarios the new word was applicable. During the first phase of the experiment, the subjects were presented with two aliens side by side. Each of the aliens had one feature from each of the four dimensions (height, weight, number of eyes, and color). 24 different aliens were created using variations of these dimensions, two types for height, weight, and number of eyes, and three types for color. An alien verb was shown in between the aliens to indicate a causal relation. For example, in between the aliens would be the word "glorps", which the subject was to interpret as the alien on the left "glorps" the alien on the right. There implied causal relation of the alien verb was altered in several slides by changing "glorps" to "is glorped by". The alien verb presented to the subjects covaried with a single feature in a single role. The participants were to learn which of these features for what role (stimuli vs. experiencer) covaried with the verb's usage.

For each subject, the experiment was divided into three separate phases: the observation phase, criterion phase, and tests of implicit causality phase. In the first of these phases, the observation phase, the subjects were presented with examples of the verb's usage as described earlier. The second phase, the criterion phase, tested whether or not the subject learned the verb correctly. This was done by presenting the subject with two aliens side by side similar to the observation phase, but instead of simply stating that the alien on the left "glorps" the alien on the right, the subject would be asked if the left alien "glorps" the right alien. The purpose of the criterion phase is to determine whether or not the subject correctly learned the definition of the alien verb. After ten correct uses of the alien verb, the subject moves onto the final phase.

The tests of implicit verb causality, or the final phase, presented the subject with questions to determine if causal weightings had been learned. Bye focused on the testing of correct causal attribution and correct pronoun disambiguation. In this phase, subjects were no longer shown images of aliens, but were instead presented a statement with novel aliens names used with "glorp" (Bykkyl glorps Vortrax). To test for learned causal attribution, the subjects were asked whether the given statement was more likely true because the prior alien "is the kind of alien that glorps other aliens", or if it was more likely true because the latter alien "is the kind of alien whom other aliens glorp". To test for learned pronoun disambiguation, the subjects were presented with a causal statement followed by a reason with an ambiguous pronoun. An example of this statement would be "Vortrax glorps Flovon because that's the kind of alien he is". The subjects were then asked who the word "he" was more likely referring to. Both of these likeliness judgments were measured on an 11-point Likert scale. If causal weightings were indeed learned, this would imply that an abstraction from exemplar features was made to produce a verb schema.

The results for this study were constructed from the data of the two tests, causal attribution and pronoun disambiguation, with two conditions: agent condition and patient condition. In the causal attribution test, subjects on average rated the likeliness of the causal role being placed on the agent higher than on the patient in the agent condition, but found the reverse to be true for the patient condition. In the causal attribution test, causal attribution for the patient condition was static, but the attribution to the agent dropped slightly from the agent condition to the patient condition. A similar result was found for the pronoun disambiguation test. In the pronoun disambiguation test, subjects on average rated the likeliness of the pronoun being attributed to the agent as being higher than the likeliness of the pronoun being associated to the patient. Unlike the causal attribution test however, subjects' likeliness ratings for the pronoun referring to the patient were much lower than the likeliness ratings for the pronoun referring to the agent. The results showed that subjects were in fact able to learn causal attribution given covariation information, even when the verb involved is completely novel to the subject. One peculiar finding in this study was that subjects tended to attribute the causal role to the agent more often than the patient. The subjects' bias towards the agent however, is due to a priori linguistic knowledge. This shows that some terms in our language have predefined causal associations due to a prior knowledge such as priming. The results found in this study are the first empirical support for the covariation hypothesis for implicit verb causality.

The conclusions drawn from this study may lead to many interesting future experiments. One possible future experiment could look at the implicit causality inferences drawn from the English language, and compare that with other languages. Are other languages more clear than English in terms of implicitly inferring a causal role? Could this difference be the root of some specific cultural difference?

Another interesting study could examine whether or not the same learning patterns that occur during implicit causal verb learning, occur in morality problems such as the infamous trolley problem proposed by Judith Thomson. The trolley problem is set up as follows:

> "A trolley is heading towards five workmen who work on the tracks. The brakes fail so that the five workers are about to be killed. However, one can flick a switch to redirect the trolley onto another track, where there is only one worker, so that only one person will be killed. Should one flick the switch?"

Given the problem, most people say "yes" that they would flick the switch. An alternate version of the question changes the situation slightly so that the flicking of the switch would push the one worker in front of the other five workers to save the five workers. In this version, most people said no. The change in attitude towards relatively the same question is caused by a change in the locus of intervention. When the locus of intervention is changed from asking the subject to kill one worker as a side effect of saving the five workers, to sacrificing the one worker to save the five workers, the subjects' opinions change as well. If we use the same tactic in Bye's study of using alien words to eliminate any a priori knowledge, we could see if people would still change their opinions in the trolley problems when the locus of intervention is made to be ambiguous. With the locus of intervention clouded, the subject would need to independently determine the meaning of the locus of intervention, and then determine the weight of both choices. There could potentially be some implicit inferences embedded into the syntax of the trolley problems that affect subjects weighting because of a priori information, which may be a confounding variable of Thomson's morality study.

*References*

Brown & Fish, The psychological causality implicit in language*, Cognition*, Volume 14, Issue 3, November 1983, Pages 237-273

Rudolph & Forsterling, “The Psychological Causality Implicit in Verbs: A Review,” *Psychological Bulletin,* 1997, 121, 192-218

Judith Jarvis Thomson, “Killing, Letting Die, and the Trolley Problem,” *The Monist*, 1976, 59, 204-217.