DISTINGUSHING MISCONCEPTIONS FROM MYTHS USING TRADITIONAL AND CLASSIFICATION ANALYSES TO EXPLORE RESPONSE TIME AS A DISCRIMINATOR

A RESEARCH PAPER

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTERS OF ARTS

BY

NATHANIEL S. RING

DR. DARRELL L. BUTLER – ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2013

Distinguishing Misconceptions from Myths Using Traditional and Classification Analysis to Explore Response Time as a Discriminator

One example of a common myth regarding the psychological and neuro-sciences would be the statement, "*humans only use 10% of the brain"*(see Taylor & Kowalski, 2004, for a review). Alternatively consider the statement, "*drug use leads to violence*". Unlike the previous statement people can directly or indirectly observe the consequences of drug use and violence. Additionally, in the latter case there may also be other political, social, economic, and individual reasons for agreeing or not in the statement (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Both examples statements are false but commonly believed—importantly, they may represent two different types of cognitive processes. For the majority of people, the 10% misconception was likely learned from others, whereas it is possible to infer the relationship between drugs and violence by thinking about what one knows about drugs and about violence. In other words, myths are beliefs about science that are learned specifically and involve direct memory lookup of what has been learned, and misconceptions are beliefs that are inferred from memory.

Myths are a common part of the human experience; indeed nearly every scientific field has recorded many myths that are common among novices and non-experts (Hynd & Guzzetti, 1993; Pine, Messer, & St. John, 2001). This manuscript focuses on those myths and misconceptions that are common concerning criminal justice and the social and behavioral sciences (Lewandowsky, et al., 2012; Taylor & Kowalski, 2004; Lillenfiled, Lynn, Ruscio, & Beyerstein, 2009; Barkan & Bryjak, 2008). Unlike previous literature which has focused on mitigating the general prevalence of misconceptions, this manuscript will focus on experimental methods to distinguish misconceptions from myths. The central purpose of this manuscript is to explore the distinction of misconceptions from myths. There are three aims of the introduction.This introduction will present a brief summary of the cognitive and neurological methodology for investigating long-term memory with a focus on information retrieval, the methodologies most commonly used to measure myths and misconceptions, and research done by Dr.

Butler, myself, and our research assistants at Ball State University between 2011 and 2013 to attempt to discriminate misconceptions from myths. I will then present an argument for using response-times to bridge the gap from survey methods to the neural methods, providing a comparison of the analyses and their interpretations.

Myths and Long Term Memory

Based on cognitive literature, this manuscript will make two assumptions about human memory. First, since declarative memory is distributed (Anderson & Lebiere, 1998; Taatgen, Lebiere, & Anderson, 2006), for efficient information is retrieval, it must be pulled from multiple sources in parallel. Second, declarative remembering is a constructive process (McClelland, 2011), thus each reconstruction is subject to potential retrieval accuracy errors, with increased errors over time following exposures to misinformation (Loftus, 2005), gaining retrieval fluency (Baddeley, 2005; Anderson, Bjork, & Bjork, 1994). However errors may be reduced with retrieval practice and corrective feedback (e.g., the testing effect; Roediger & Karpicke, 2006).

There are many experimental paradigms to test information retrieval from long-term memory in the laboratory. Researchers focusing on myths have largely relied on surveys to assess the prevalence of specific beliefs (see Lillenfiled, Lynn, Ruscio, & Beyerstein, 2009, for an example), secondary memory indicators such as self-report memory strength (see Glovich, 1991), or objective learning measures (for a general review of the literature see Glovich, 1991).

Traditional survey methods for assessing the presence/absence of misconceptions provide information about the prevalence, in general, of those specific ideas within the sampled populations (e.g., college students). However, these methods do not provide detailed information as to how individuals retrieve this information in memory. Specifically, traditional survey methods cannot provide evidence if the response to a given item on a survey is retrieved directly from memory, or if it inferred from other information in memory in a more thoughtful and deliberative process (Taylor and Kowalski, 2004; Gardner and Dalsing, 1986, McCutcheon, 1991, Ruble, 1986).

Confrey (1990) systematically reviewed the literature on student conceptions within mathematics, science, and programming, providing a framework to broadly connect student conceptions and student myths. In this review, Confrey suggests that understanding student conceptions is critical to education, and by virtue of identifying the domains in which students have myths then educational practices can be improved. However, Confrey notes that much of the research to measure and identify (mis)conceptions are domain specific areas with few attempts to integrate cross-disciplinarily. Additionally, the common methodology is to use pre-post designs on surveys/quizzes issued at various times within courses. Finally, much of this research is focused on reducing the number of myths students hold toward a specific aspect of sciences, and less interested in the cognitive basis of the errors.

Some researchers have turned to neurological methods to as an alternative to avoid the limited characteristics of surveys and self-report methods. For example, Qin, van Marle, Hermans, and Fernández (2011), investigated memory strength using fMRI, linking the subjective aspects of memory confidence and objective accuracy to disparate neural activation. The researchers suggest that a top-down process is related to subjective memory confidence, whereas the objective accuracy was related to medial-temporal binding processes. Their approach was focused on determining the neural structures for image/gist memories, distinguishing the subjective knowledge from the objective knowledge. Their findings support a two-system approach in memory. One system that determines the subjective (gist) strength of a memory, the second more is associated with the accurate and specific recall of information. The findings of Qin and colleagues report are highly consistent with cognitive theory for information retrieval in long-term memory (e.g., Anderson and Lebiere, 1998).

There has been many surveys and inventories developed to understand myths and misconception and in the teaching of sciences particularly in physics (Perkins, Adams, Pollock, Finkelstein & Wieman, 2005), biology (Klymkowsky & Garvin-Doxas, 2008), engineering (Evans, Gray, Krause, Martin, Midkiff, Notaros, Pavelich, Rancour, Reed-Rhoads, Steif, Streveler, & Wage, 2003), and experimental methodology (Meyer, Shanahan, & Laugksch, 2005). Many of these measures were developed to provide information to educators about pedagogical practices (Nesher, 1987) that could correct myths and misconceptions. Moti Nissani (1997) argues while many myths have been documented and educational practices have been developed to mitigate their prevalence, little work has been done to understand the causes within those fields. In this view, myths are a function of individual's traits to learn new information. Further, there are many myths that occur beyond the sciences and possibly have social/political implications. Finally, he argues this prevalence is the consequences of *conceptual conservatism* (Nassani, 1997)—a trait in which individuals will learn new concepts however will fail to integrate the new concepts with old concepts, similar to the Freudian isolation defense mechanism (Baumeister, Dale, and Sommer, 2002). In this view memory is not a significant factor for myths, rather myths and misconceptions are a function of personality characteristics.

Factor analysis has been used to uncover latent traits associated with myths. An unpublished survey (Swhear, Butler, Brown, and Ring, 2012), using principle components analysis suggested a complex system of latent variables for scales measuring specific misconceptions in psychology, criminal justice, and cross-disciplinary misconceptions (those related at the intersection of criminal-justice and psychology). Individuals indicated the degree to which they believed statements from these disciplines were true or not true. The PCA structures revealed six components for the psychology misconceptions explaining 53.19% of variance; five components for criminal-justice explaining 46.97 % of the variance; and three components for cross-disciplinary explaining 51.71% of the variance; and combined scales revealed 14 components accounting for 55.81% of the variance. Using the same data set, when group level characteristics such as major or academic progress are entered into the factor analysis, total explainable variance decreases slightly on each subscale and as a total scale (Swhear, Ring, Brown, & Butler, unpublished data). These results are consistent with research investigating myths following training (McKeachie, 1960; Vaughn, 1972; Gardner & Dalsing, 1986; Kowalski & Taylor, 2001). Researchers argue these beliefs are often strong and can become fixed over time (Otero, 1998; Vosniadou, 2001).

An alternative method is to use pre-post designs on exams to determine if particular concepts remain widely believed following new learning. Research in this manner provides objective concept measurement whereby an increase in the total number of correct responses suggests abating myths. However, individuals may learn new information merely to pass an exam and do not incorporate the new evidence into their belief systems (Otero, 1998; Vosniadou, 2001). Thus when students perform better on exams, they often falsely attribute the increase in their performance. The testing effect demonstrates a fundamental problem of the subjective component of human memory in that most students do not attribute the increased recall accuracy as a function of practicing recall, rather students are more likely to indicate repeated studying would predict improved recall (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2007).

Similarly, individuals may have a subjective experience of knowing or not knowing information, but fail to recall the source event (e.g., Tip of the Tongue effect, Brown and McNeill, 1966) or inaccurately report the source information. Researchers using secondary measures of individuals subjective meta-memory often ask individuals confidence in their response. In some cases, subjective rating is then treated as a covariate to overall performance on a conceptions survey. However, in most cases, the research has shown individuals are often poor at introspection into their own memory processes (Loftus, 2005; Tulving, 1981; Schacter, 1995; Kensinger, 2008). How then can participants articulate situations in which they inferred or directly accessed information stored in declarative memory?

Kahenman (2011) provides a cognitive view of conceptions, describing them as decisions resulting from *fast or slow* processing systems. Kahneman argues that the vast majority of our decision-making and information processing occurs with system one, the fast processing system. Under special circumstances, decisions involve system two. The first system operates automatically and affectively (e.g., using direct memory access) to reduce information loading, or using schemas and heuristics when information/experience is limited (Tversky & Kahneman, 1974; 1982; Gowda, 1999; Alverz & Busenitz, 2001). The second system operates as a more deliberate and effortful process. Unfortunately the benchmark response-time difference between these two systems has not been formalized. Broad relative metrics for determining system one and system two have been developed, but none have sensitivity to the millisecond scale. It is important to note that most models using response time usually consist of two or four keys (Kosinsky, 2008; Donkin and Nosofsky, 2012). Response time has been argued to serve as a potential measure to detect if information is a retrieval strategy or an inference strategy (Kahenman, 2011; Camp, Lehman, & Lehman, 1980; Reder, 1987)—a fast and slow memory processing system (Bergert and Nosofsky, 2007; Kahneman, 2011). From a theoretical perspective, Kahneman (2011) provides a useful framework for which misconceptions can be distinguished from myths.

Pilot Studies

In this section will summarize the pilot research on myths and misconceptions conducted from Summer 2011 to Spring 2013 in preparation of the experiment to be reported.

I have defined the behavior of a *myth* as a fast response to an incorrect statement and a *known fact* as relatively faster response to a verifiable statement. While *misconceptions* and inferred *facts* are relatively slow compared to known statements. A *central idea* is defined as the related collection of statements written as both misconception and fact (i.e., percentage of brain use in humans). I will first describe the pilot studies that focused on developing the stimuli through focus groups and statistical evaluation of the instrument. Second, I will then describe the software pilot study where based on the feedback from focus groups and interviews during the pilot studies, we determined the need for a broader response range and the inclusion of a *I don't know* (IDK) key. Finally, I will then discuss the final instrument and it's reliability and validity characteristics. Following a description of the materials, I will then outline the four statistical approaches used to explore the use of reaction time as a metric for determining if an individual has a misconception or myth.

Instrument development

Participants. 644 Ball State University undergraduates in two departments volunteered and received research credit in their courses. We sampled from introductory and advanced courses in criminal justice (n=182) and an upper level forensic psychology course (n=40), and the Department of Psychological Science human participants pool which broadly represents all majors at the university, but mostly freshman.

Materials. Following good practice for survey development, in the summer of 2011 a research team, of which the author was member, identified 40 central ideas from texts in criminal justice and psychology concerning prevalent misconceptions in those areas (Barkan & Bryjak, 2008; Lilienfeld, Lynn, Ruscio, & Beyerstein, 2009). For each central idea, the researchers created an *incorrect statement* and *fact statement*, therefore a total of 80 individual statements. This initial list was revised by interviewing students to make sure they understood the statements and experts in criminal justice and psychology on the validity of the individual statements. In the Fall of 2011, 561 students were randomly assigned to one of two survey's: either all items were false statements or all were true statements. In Spring 2012 an additional 312 students were randomly assigned to one of two surveys that consisted of balanced statements of each central misconception as to demonstrate content validity of the scale. Results indicated there were no significant differences between the forms on participant's behavior in terms of overall number of misconceptions. Additionally, there were no differences between those central idea statements regarding criminal justice, psychology. Both scales had moderate to high reliability (Chronbach's α Fall 2011 = .773; Chronbach's α Spring 2012 = .667) and while there is a discrepancy in the α coefficient, outcome scores on all three criterion (total score, criminal justice items, and psychology items) were not significantly different with a critical α = .001.

Pilot research also investigated predictors of total number of misconceptions by gathering demographic and self-report behavior data regarding their television viewing (Brown, Butler, Ring, & Swhear, 2011). The researchers gathered a total of 644 participants, 182 from a criminal justice course, a group from a forensic psychology course, and the remainder from the psychological science subject pool. The researchers were interested in predicting the number of correctly identified misconceptions. Results indicated there was an interaction for only students in the psychology program and their overall educational level. These findings suggest as psychology student's progress through their education, they are less likely to endorse a myth statement, there was also a significant effect of education on the endorsement of a myth ($ß_{Psychology*Education}$=.215, $ß_{Education}$=.151). However, the total variance accounted for in this study was relatively low (Adjusted R^2 = .157), suggesting stronger possible indicators or scale redevelopment. No other factors that were evaluated were reliably significant. Therefore, the research team concluded to perform a revision of the scale and explore alternative indices of myths and misconceptions.

Each statement from the pilot studies was reviewed for statement clarity using focus groups and formed the foundation for the second stimulus set. The critical statements were then doubled again to provide additional to verify the response behavior of each central idea. The total test stimulus list consists of 180 statements. In order to provide baselines on direct access memory or inference, a list of sentence verification (40 items) and inference statements (40 items) were collected from the literature (Collins & Quillian, 1968, Camp et al., 1980). Additionally, based on the feedback from individuals in the focus groups in which students indicated they would choose answers random if did not know anything about the critical statements. Therefore, the response scale was refined from a *True* to *Not True* scale to also include and *I don't know* response option. The revised scale reliability increased substantially (Chronbach's α = .884).

Reading and Information Processing

To provide more sensitive analyses of response time, we gathered 80 participants to provide data on reading time for each statement. Participants were told to press any key on the keyboard or mouse as soon as they read and understood the statements presented. See figure 1 for results and comparison against data processing individuals. Results indicate the strongest predictor of reading time is sentence length in total character and spaces (R^2 = .938) and the linear equation is:

Read Time = 300.78 + (35.642) * Number of characters in statement

Using this equation, reading time can be separated from fast-slow decision-making. As each statement increases in total number of characters (including spaces), there is a strong positive relationship to response time. Thus, as individuals deviate from the estimated read time, per statement, an individual is likely using inference processing.

Experiment I

Participants

The sample consisted of 79 Introductory Psychology students from the Ball State Psychology Subject Pool with corrected to- or normal eyesight and no severe reading disability. Students were given research credit for their participation.

Procedure

Participants were run in groups of two with a researcher outside the room to record if any distracting noises occurred which could influence reaction times. Participants were instructed how to use the software by a researcher and prompted at the beginning of each section. Participants responded using a three-key scale (*I believe this statement is True*, *I believe this statement is not true*, or *I do not know*; F, J, and spacebar, respectively; Gardner and Dalsing, 1986). The research session was broken up into 6 blocks to reduce fatigue and increase attentiveness during trials (Sanders, 1998; Whelan, 2008). In order to provide individual baselines on direct access memory and inference, a list of sentence verification (23 items) and inference statements (23 items) were collected from the literature (Collins & Quillian, 1968, Camp et al., 1980). The first two sessions serve as the base-line conditions for within-individual reading speed and inference processing and to practice familiarizing with the software. These conditions were counter balanced between participants.

Following the two practice sets, participants were then randomly assigned to one of four blocks of central idea statements. The total test stimulus list consists of 180 statements. The remaining sessions will consist of 45 statements, a randomly selected from the list of 180 statements. Participants will be asked to respond as accurately and quickly as possible to each statement. Upon completing the experiment, participants were debriefed, thanked and awarded course credit for their time. The data set is created as person-period for means of classification, and RT comparisons are based on normalized data (c.f., Tabachnick & Fidell, 2013, see Whelan, 2008 for a review) each participant was assigned a random participant identification number.

Analyses

Based on previous literature in response-time studies (e.g., Whalen, 2008), several analyses were conducted to determine the relationship between judgments, RT, statement-ground truth, guessing behavior, and deviation from reading time. Specifically, the analyses focus on the predictive value of models using RT as a metric. Thus, two general approaches are used. First, comparing overall model instance classification; second, using a computational time metric to determine which analysis is most efficient.

To determine *fast-slow* judgments, a variable was created by taking the average read time latency from a baseline-comparison group for each statement (see figure 1), and compared to each participant's reaction time per statement. Because each participant also performed a reading latency and inference latency, data was analyzed to detect if participants were responding faster than 2.5 standard deviations from the benchmark reading-times (e.g., 1250ms slower than the Read Time for each statement) and were excluded from results. Response times to statements that closely approach average read time suggest little processing therefore direct access, whereas reaction times with longer latency suggest inference processing.

Model comparisons are performed by comparing effect sizes and/or correct classifications given. This is possible due to the person-period structure of the data, treating each statement, and participant as a unique function of variance for each model. Analyses are conducted using Weka Data Mining for the CART and Logistic Regression Classification analyses (Hall, Frank, Holmes, Pfahringer, Reutemann, and Witten, 2009). Analysis of variance is conducted using R (R Development Core Team, 2008). For all data mining analyses, parameters are simply the variables being evaluated in the dataset. For example, a two-parameter model would consist of just two variables—a predictor and criterion.

Analysis of Variance

The analysis of variance is a common inferential statistic used in response-time experiments (Whalen, 2008), however, results should be interpreted with caution for at least two reasons. First, data likely violate normality and independence, so data are likely transformed; commonly by taking the natural log. Second, because the data is then on an $e^y = x$ scale, providing a meaningful description of the differences in averages becomes difficult. Therefore, in terms of *fast-slow* performance, I averaged each participants read time and process time derived from their baseline conditions. I then determined within

2.5 standard deviations, if a response time was either faster than average, average, or slower than average. The analysis investigates their overall performance on the scale and two what extent individuals deviate in total scale score based on Fast-Slow processing.

Logistic Regression Classification

Unlike the ANOVA, Logistic regression can provide odds ratios and the probability of a discrete outcome. Because overall, I am interested in the response characteristic, (e.g., belief); this is a traditionally optimal predicative analysis. Three analyses were performed. The first model predicted choice (True, Not True, and I don't Know) from overall scale performance (e.g., number correct), ground truth, Participant, order of stimulus block, and individual stimulus presentation order. The second model included response time as a continuous variable. The third model included response time as a discrete variable.

CART

CART (Classification and Regression Tree; see Lewis, 2000 for an introduction) is a nonparametric statistical method that can select from among a series of variables those which provide the most predictive strength on the explanatory variable. This method also accommodates the interactions between variables (Brieman, Friedman, Olshen, and Stone, 1984). This approach is desirable because of the assumptions (and lack there of) towards the analysis. Yohannes and Hoddinott (1999) listed seven desirable characteristics of CART (listed in Appendix A). For this analysis, two approaches were used. First, I conducted the analysis classifying response (true, not true, I don't know) by item ground truth and participant as predictor variables (see Appendix B; Table 1). The second analysis included participant response time and within-participant *fast-slow* nominal variables. All data output may be found at https://docs.google.com/file/d/0B1-Z92O8Q1afdndCRTk5TmVPNU0/edit?usp=sharing

Results

ANOVA

A 3 (belief: Not True, True, I don't know) x 3 (Response Time: Fast, Average, Slow) x 42 (Misconception: 1-42) was performed. Results indicate an interaction of Belief x Misconception Number

[F (187, 17,471) = 1.273, p = .007; η^2 = .014]. Some individual's responses to statements are more likely to predict overall performance and this makes sense. Individuals who indicate IDK are more likely to have an overall decreased number of correct responses (see Figure 2 and 3) due to nature of the scale score. There is also a main effect of response time on total number of correct responses [F (2, 17,471) = 43.044, p < .000, η^2 = .005]. Students who responded slower, more frequently responded correctly (Mean Slow = 102.535; Mean Average = 99.874; Mean Fast = 96.821).

A second 3 X 2 (Response Time: Fast, Slow) X 42 ANOVA was also performed to evaluate the effect of the clustering used in the first analysis. The only parameter that changed was response time clustering. Overall model fit, as measured by R squared, decreased (R^2_{Model2} = .123; R^2_{Model1} = .140). However it is questionable if the loss of the additional parameter significantly increases model fit. Computational time for more sophisticated analyses deemed impractical given the weak effect sizes.

Logistic Regression

To evaluate the notion that response time can provide useful information to response characteristics. Model one results indicate a significant model fit and correctly classified 58.79% of the instances in the data set. Model two out performed the initial model correctly classifying 71.19% of the instances. Model three performed slightly better than model two and overall classified the most number of instances correctly (72.81%).

It is important to note processing time for each model as a metric of their general feasibility. The initial 9-parameter model took 215.68 seconds to generate each fold plus the initial model. Whereas the 10-parameter continuous response time model took 311.01 seconds and the 10-parameter *Fast-Slow* model took 244.57 second to generate each fold plus the initial model. Because each model uses 10 folds to classify instances, this is a substantial increase in computational time. Thus, in this case the best performing model both for computational time and increase in classification strength is 10-parameter *Fast-Slow* model.

CART

Results suggest CART is able to correctly model misconceptions and myths with fairly high reliability. Model comparisons suggest that classifying data as *Fast-Slow* improves modeling responses as they deviate from reading time. Thirteen models were performed (see tables 2-5 in Appendix B). Predicting beliefs (True, Not True, I don't Know) were significantly enhanced at two stages. First, when including the *Correct* variable which indicates if the response was correct; Second, when including response time characteristics. However, there were differences in the degree of classification based on the type of Response Time metric. While CART models are able to deal with a wide range of data types robustly, not having a specific anchor point to define *Fast-Slow* decreased performance of the models.

Discussion

Overall, these analyses suggest the addition of a *fast-slow* parameter improves the ability of a scale to distinguish misconceptions from myths. As Taylor and colleagues have pointed out, previous scales lacked this level of sophistication and prediction of responses. Further using response time provides critical information about the cognitive mechanisms at play when individuals are recalling information from long term memory.

Findings suggest while traditional analyses may be useful in terms of assessing scale characteristics, they do not provide the level of predictive performance given high dimensional datasets. The ANOVA provides evidence to suggest individuals whose latency is over 2.5 standard deviations from any given read time, are more likely to correctly respond to items. However, given both model fit and small effect sizes, other factors may be account for the overall performance. More importantly, the ANOVA only provides limited data that individuals who are more methodical and thoughtful in responses are more likely to correctly identify which statements are true or false. Weakly suggesting individuals are either in a system 1 or system 2 states as outlined by Kahenman (2011).

Findings from the logistic regression provide respectable information, however there is a big trade off in computational time required for each analysis. Data support the two retrieval characteristics hypothesized and are further supported when treated as a hypothetical construct for prediction.

Findings from the CART analysis provide the most useful information for predicting response behavior when including item response-times for two principle reasons. First, interpreting results is substantially improved given the decision tree output. By virtue of treating the dataset as a decision tree for response characteristics, specific characteristics are modeled and outlined. Second, the confusion matrix (see Table 3, Model Comparisons), quickly show when the additions of new or different predictors influence classification of the criterion responses. For example, I conducted a Naïve Bayesian Tree analysis which produce over 2,000 pages of output, with a great degree of independent trees and leaves per tree (1025 and 963, respectively). While the model was highly successful (87.6% correctly classified instances), and fast (19.17 seconds per fold), model interpretation is highly difficult given the sheer complexity of the results. It is for that reason, CART provides the most functional alternative approach. Standard analysis like logistic regression or ANOVA does not provide computationally or resource advantages unlike data mining approaches. In particular, both Logistic and ANOVA took significantly longer computer processing time than CART and several other methods not discussed in this manuscript.

Additionally, unlike the logistic and ANOVA models, the CART models can be deployed to predict individual's response characteristics without the use of group-level data. In other words, CART is able to provide a person-level prediction model unlike the traditional model. Future researchers will need to determine how many stimuli are needed to build a classification model that can help determine which ideas students have learned incorrectly and which ideas they are incorrectly deciding on based on other things they know. This has important consequences to how one goes about changing an incorrect idea. Discovering the ideas upon which students base an inference can be quite difficult, but may be critical to correcting an error.

References

Alvarez, S. A., & Busenitz, L. W. (2001). The entrepreneurship of resource based theory. *Journal of Management, 27,* 755-775.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing geffect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861-876.

Anderson, J. R., & Leibere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Earlbaum.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning Memory, and Cognition, 20,* 1063-1087.

Baddeley, A. (2005). Short-term and working memory. In Tulving, E., and Craik F. I. M. (Eds.). *The Oxford Handbook of Memory.* (pp. 77-92). Oxford: Oxford University Press

Barkan, S. E., & Bryjak, G. J. (2008). *Myths and Realities of Crime and Justice: What Every American Should Know.* Jones and Bartlett Publications, LLC.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning Memory and Cognition. 33,* 107 – 129.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1983). *Classification and Regression Trees.* Wadsworth.

Brown, R., & McNeil, D. (1966). The "Tip of the Tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325-337.

Camp, C. J., Lachman, J. L., & Lachman, R. (1980). Evidence for direct-access and inferential retrieval in question-answering. *Journal of Verbal Learning and Verbal Behavior, 19*, 583-596.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. *Review of Research in Education, 16,* 3-56.

Donkin, C., & Nosofsky, R. M. (2012). The structure of short-term memory scanning: an investigation using response time distrubition models. *Psychonomic Bulliten & Reivew, 19,* 363 – 394.

Evans, D. L., Gray, G. L., Krause, S., Martin, J., Midkiff, C., Notaros, B. M., Pavelich, M.,

Gardner, R. M., & Dalsing, S. (1986). Misconceptions about psychology among college students. *Teaching of Psychology, 13,* 32-34.

Glovich, T. *How We Know What Isn't So: The Fallibility of Human Reason in Every Day Life.* NY: The Free Press.

Gowda, M. V. R. (1999). Heuristics, biases and the regulation of risk. *Policy Science, 32,* 59-78.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An update, *SIGKDD Explorations,* Volume 11, Issue 1.

Hynd, C. R., & Guzzetti, B. J. (1993). Exploring issues in conceptual change. In D. J. Leu & C. K. Kinzer (Eds.), *Examining central issues in literacy research, theory and practice* (pp. 373-381). The National Reading Conference.

Klymkowsky, M. W., & Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed's tools and the Biology Concept Inventory. *PLoS Biology, 6*, *1*, e3. doi:10.1371/journal.pbio.00600003.

Kahneman, D., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus, and Giroux.

Kosinksy, R. J., (2008). A literature review on reaction time. Last Updated September 2012. Retrieved June, 2013 from: biae.clemson.edu/Bpc/Bp/Lab/110/Reaction.Htm

Kowalski, P., & Taylor, A. (April, 2001). *An exploratory study of developmental misconceptions.* Paper presented at the American Educational Research Association Conference, Seattle, WA.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N. & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106-131.

Lewis, R. J. (2000). An introduction to classification and regression trees (CART). *Avaliable at: http://www.saem.org/download/lewis1.pdf.* Access May 1, 2013.

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2009). *50 Great Myths of Popular Psychology: Shattering Widespread Misconceptions about Human Behavior.* Hoboken: John Wiley & Sons.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and Memory, 12,* 361-366.

McCelland, J. L. (2011). Memory as a constructive process: The parallel-distributed processing approach. In S. Nalbantian, P. Matthews, and J. L. McCelland (Eds.), *The Memory Process: Neuroscientific and Humanistic Perspectives.* Cambridge, MA: MIT Press, pp.

McCutcheon, L. E. (1991). A new test of misconceptions about psychology. *Psychological Reports, 68,* 647-653.

McKeachie, W. J. (1960). Changes in scores on the North-Western misconceptions test in six elementary psychology courses. *Journal of Educational Psychology, 51,* 240 – 244.

Meyer, J. H. F., Shanahan, M. P., & Laugksch, R. C. (2005). Students' conceptions of research methods I.: a qualitative and quantatitive analysis. *Scandinavian Journal of Educational Research, 49,* 225-244.

Nesher, P. (1987). Towards an instructional theory: The role of student's misconceptions. *For the Learning of Mathematics, 7,* 33-40.

Nissani, M. (1997). Can the persistence of misconceptions be generalized and explained? *Journal of Thought, 32,* 69-76.

Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science texts. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice.* (pp. 145 – 164). Mahwah, NJ: Erlbaum.

Perkins, K. K., Adams, W. K., Pollock, S. J., Finkelstein, N. D., & Wieman, C. E. (2005). Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey. Marx, J., Heron, P., & Franklin S. (Eds), *Physics Education Research Conference.* American Institute of Physics.

Pine, K., Messer, D., & St. John, K. (2001). Children's misconceptions in primary science: A survey of teachers' views. *Research in Science & Technological Education, 19,* 79-96.

Qin, S., van Marle, H. J. F., Hermans, E. J., & Fernández, G. (2011). Subjective sense of memory strength and the objective amount of information accurately remember are related to distinct neural correlates at encoding. *The Journal of Neuroscience, 31,* 8920-8927.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rancour, D., Reed-Rhoads, T., Stief, P., Streveler, R., & Wage, K. (2003). Progress on concept inventory assessment tools. *Proceedings of the 33rd ASEE/IEEE Fronters in Education Conference,* pp. T4G1-T4G8.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19*, 90-187.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249 - 255.

Ruble, R. (1986). Ambiguous psychological misconceptions. *Teaching of Psychology, 13,* 34-36.

Taatgen, N. A., Lebiere, C., & Anderson, J. R. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation.* Cambridge University Press: 29-52

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*, 6th ed. Boston: Allyn and

Bacon.

Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevelance, strength, and sources of misconceptions. *The Psychological Record, 54,* 15-25.

Tulving, E. (1981). Similarity relations in Recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 479-469.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science, 185,* 1124 – 1131.

Sanders, A. F. (1998). *Elements of Human Performance: Reaction Processes and Attention in Human Skill.* Laurence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Schacter, D. L. (1995). Memory distortion: History and current states. In Schacter D. L. (1996). *Searching for Memory: The Brain, The Mind, and the Past.* New York: Basic Books

Swhear, J. M., Butler, D. L., Brown, M. P., & Ring, N. S. (May, 2012). Myth modifiers: Do behavioral science courses counteract students' myths? Poster Presentation at the Society for the Teaching of Psychology Conference. Chicago, IL.

Vaughn, E. D. (1977). Misconceptions about psychology among introductory college students. *Teaching of Psychology, 4,* 138 – 141.

Vosniadou, S. (2001). What can persuation research tell us about conceptual changes that we did not already know? *International Journal of Educational Research, 35,* 731-737.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Report, 58,* Article 9. Available at: http://opensiuc.lib.siu.edu/tpr/vol58/iss3/9

Yohannes, Y., & Hoddinott, J. (1999). *Classification and Regression Trees: An Introduction.* Washington, D.C.: International Food Policy Research Institute.

Appendix A.

Figure 1.

Mean natural log transformed reaction time for each statement. The first group represents direct memory look-up stimuli in the baseline or pre-experiment condition, the second group represents inference statements in baseline or pre-experiment condition, and the third group are the stimuli studied in the experiment.

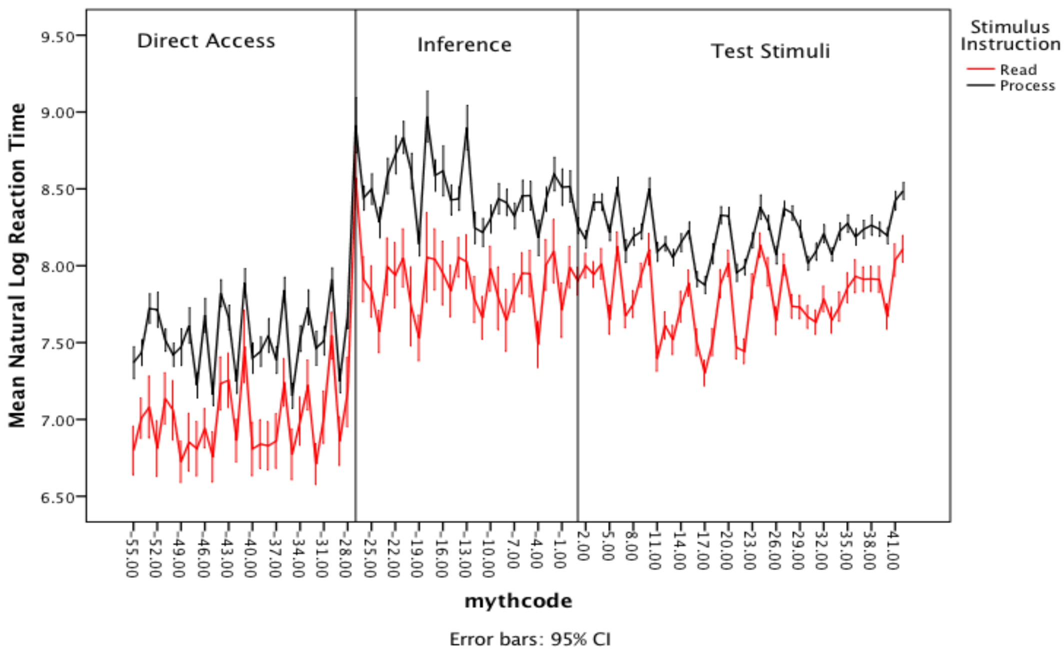
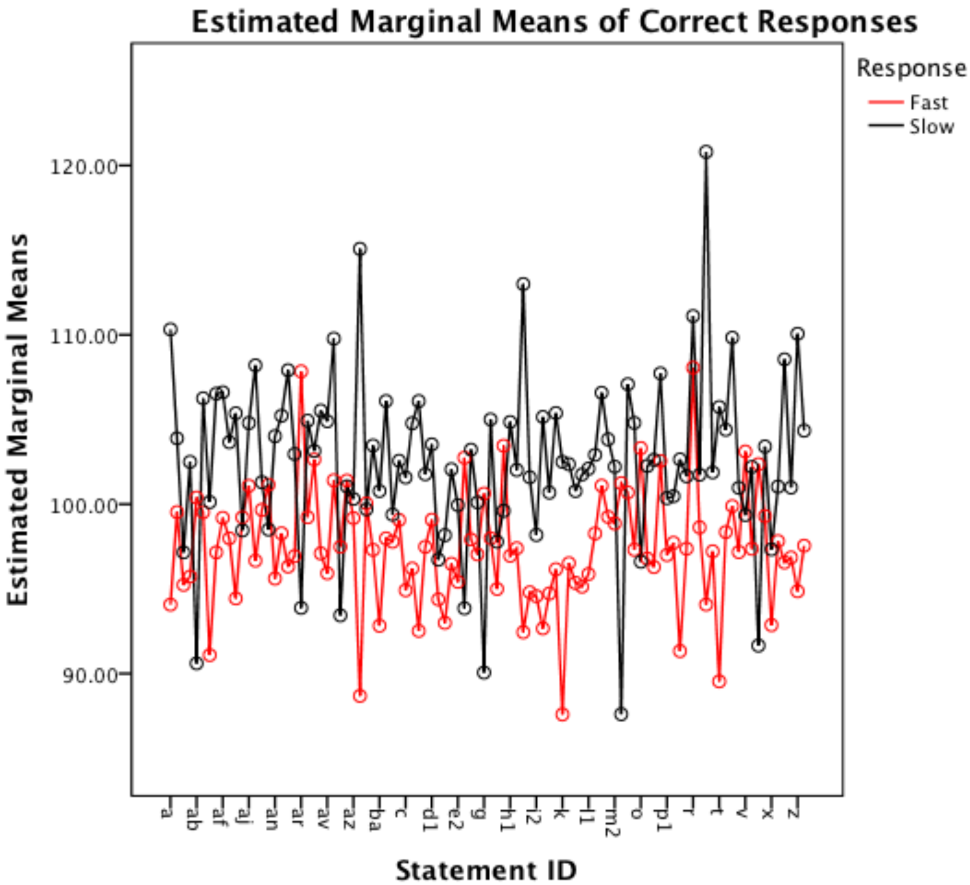


Figure 2.

The mean number of correct responses showing fast and slow responses by each statement.



Appendix B: Tables

Table 1.

CART Models Without Classified Response Times Predicting Belief

Classification and Regression Tree	Models								
	Null Model	6 Parameters	7 Parameters	8 Parameters	9 Parameters	10 Parameters	11 Parameters	12 Parameters	13 Parameters
Correctly Classified	10235 (58.6%)	9952 (57%)	10022 (57.4%)	9926 (56.8%)	10018 (57.3%)	10023 (57.4%)	10049 (57.5%)	15127 (86.6%)	15109 (86.5%)
Incorrectly Classified	7237 (41.4%)	7520 (43%)	7450 (42.6%)	7546 (43.2%	7454 (42.7%)	7449 (42.6%)	7423 (48.5%)	2345 (13.4%)	2363 (13.5%)
Kappa	0.3309	0.3024	0.3113	0.3012	0.3093	0.3086	0.3121	0.7816	0.78
Root Mean Squared Error	0.4279	0.4264	0.4275	0.4263	0.3146	0.4261	0.4282	0.2537	0.2542
Relative Absolute Error	78.19%	81.41%	80.55%	81.87%	81.62%	81.71%	81.03%	29.74%	29.97%
Number of Leaf Nodes	765	443	413	323	219	184	189	29	176
Size of Tree	1529	885	825	645	437	367	377	57	351

Model Variables RT not coded								
Null Model	**6 Parameter**	**7 Parameter**	**8 Parameter**	**9 Parameter**	**10 Parameter**	**11 Parameter**	**12 Parameter**	**13 Parameter**
Ground Truth	Ground Truth	Ground Truth	Ground Truth	Ground Truth	Ground Truth	Ground Truth	Ground Truth	Ground Truth
Belief	Belief	Belief	Belief	Belief	Belief	Belief	Correct	Correct
Myth Code	Myth Code	Myth Code	Myth Code	Training	Training	Training	Belief	Belief
PID	PID	PID	PID	Myth Code	Myth Code	Myth Code	Training	Training
Block	Block	Block	Block	PID	LogRT	LogRT	Myth Code	Myth Code
	Block Order	Block Order	Block Order	Block	PID	PID	LogRT	LogRT
		Reading Time (statement)	Reading Time (statement)	Block Order	Block	Block	PID	PID
			Deviation from reading time RT	Reading Time (statement)	Block Order	Number Correct (scale)	Block	Block
				Deviation from reading time RT	Reading Time (statement)	Block Order	Number Correct (scale)	Number Correct (set)
					Deviation from reading time RT	Reading Time (statement)	Block Order	Number Correct (scale)
						Deviation from reading time RT	Reading Time (statement)	Block Order
							Deviation from reading time RT	Reading Time (statement)
								Deviation from reading time RT

Table 2. Confusion Matrices Without Classified Response Times and Confusion Matrices Predicting Belief

Confusion Matrix: 13 Parameter	Classified as			
Response	FACT	NOT FACT	IDK	
FACT	***7038***	0	335	
NOT FACT	1	***6706***	377	
IDK	841	809	***1365***	57.2 Seconds
Confusion Matrix: 12 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***7043***	0	330	
NOT FACT	1	***6715***	368	
IDK	843	803	***1369***	28.48 Seconds
Confusion Matrix: 11 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4694***	2234	525	
NOT FACT	2296	***4165***	523	
IDK	977	924	***1164***	57.56 Seconds
Confusion Matrix: 10 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4594***	2254	525	
NOT FACT	2296	***4265***	523	
IDK	977	874	***1164***	64.36 Seconds
Confusion Matrix: 9 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4481***	2334	558	
NOT FACT	2303	***4239***	542	
IDK	904	905	***1206***	62.19 Seconds
Confusion Matrix: 8 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4549***	2295	529	
NOT FACT	2252	***4273***	559	
IDK	928	891	***1196***	57.59 Seconds
Confusion Matrix: 7 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4547***	2242	584	
NOT FACT	2255	***4249***	580	
IDK	924	865	***1226***	56.41 Seconds
Confusion Matrix: 6 Parameter	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4481***	2334	558	
NOT FACT	2303	***4239***	542	
IDK	904	905	***1206***	54.02 Seconds
Confusion Matrix: Null Model	*Classified as*			
Response	FACT	NOT FACT	IDK	
FACT	***4705***	2133	534	
NOT FACT	2160	***4301***	623	
IDK	892	894	***1229***	47.51 Seconds

Table 3

CART Models with Classified Response Times and Confusion matrices predicting Fast Slow

Classification and Regression Tree	**Fast Slow**	**Fast Average Slow**
	10 Parameters	10 Parameters
Correctly Classified	13876 (79.4%)	12458 (71.3%)
Incorrectly Classified	3596 (20.6%)	5014 (28.7%)
Kapp	0.5438	0.5091
Root Mean Squared E	0.3811	0.3652
Relative Absolute Erro	61.43%	64%
Number of Leaf Nodes	11	81
Size of Tree	21	161

Confusion Matrix 10 Parameters *Fast Slow*			
	Classified as		
Response	Fast	Slow	
Fast	9669	1773	
Slow	1823	4207	42.08 Seconds

Confusion Matrix: 10 Parameters *Fast, Avg, Slow*				
	Classified as			
Response	Average	Fast	Slow	
Average	236	1077	1519	
Fast	191	7226	1193	
Slow	220	814	4996	163.1 Seconds

Table 4.

CART Models With Classified Response Times Predicting Belief and Confusion Matricies

Classification and Regression Tree	**Fast Slow**	**Fast Avg Slow**
	12 Parameters	12 Parameters
Correctly Classified	15091 (86.4%)	14955 (85.6%)
Incorrectly Classified	2381 (13.6%)	2516 (14.4%)
Kapp	0.7792	0.7655
Root Mean Squared Er	0.2532	0.2566
Relative Absolute Erro	29.23%	40%
Number of Leaf Nodes	137	60
Size of Tree	273	119

Confusion Matrix: 10 Parameters *Fast Slow*				
	Classified as			
Response	Fact	Not Fact	IDK	
Fact	6962	0	411	
Not Fact	1	6689	804	
IDK	771	394	1140	121.25 Seconds

Confusion Matrix: 10 Parameters *Fast, Avg, Slow*				
	Classified as			
Response	Fact	Not Fact	IDK	
Fact	6862	21	411	
Not Fact	1	6679	814	
IDK	771	404	1120	127.31 Seconds