EFFECTS OF SELF-EXPLANATION AS CAUSAL MECHANISM ELICITATION METHODOLOGY ON CAUSAL REASONING TASK PERFORMANCE

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri – Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

IOAN GELU IONAS

Dr. David Jonassen, Dissertation Supervisor

MAY 2010

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

EFFECTS OF SELF-EXPLANATION AS CAUSAL MECHANISM ELICITATION METHODOLOGY ON CAUSAL REASONING TASK PERFORMANCE

Presented by Ioan Gelu Ionas,

A candidate for the degree of Doctor of Philosophy,

And hereby certify that, in their opinion, it is worthy of acceptance.

_____

Professor David Jonassen

_____

Professor Charles Franz

_____

Professor Sanda Erdelez

_____

Professor James Laffey

_____

Professor John Wedman

To my daughter Ilinca, my wife

Roxana and my family back home for

their love, patience and support.

ACKNOWLEDGMENTS

The end of a road and the beginning of another is a moment of reflection. I was fortunate to have had the opportunity to be part of a community of brilliant, innovative and dedicated people who shaped me as a scholar, an educator and as a person. Now, when I look back to my years as a doctoral student I see glimpses of familiar faces, faces of people that joined my path and helped me reach my destination.

My greatest debt of gratitude goes to my mentor and advisor, Dr. David Jonassen, who guided me throughout most of my doctoral program by revealing opportunities and posing challenges to my ideas and my research. Without his guidance and support this dissertation wouldn't have been possible.

I would like to express special thanks to the faculty who served on my dissertation committee: Dr. Sanda Erdelez, who jumpstarted my doctoral program as my advisor and continued to help me throughout by challenging my analysis; Dr. Jim Laffey, for his continuous in depth scrutiny of my research methods; Dr. John Wedman, for his help in refining my work. Last but not least I would like to express my gratitude to Dr. Charles Franz, my mentor and friend long before I began my graduate studies at MU, for his continued and selfless support.

I am grateful to the faculty who opened their world of knowledge to me and tolerated my inquisitive questions throughout my coursework. Special thanks to my fellow graduate students Dan Cernusca, Matthew Schmidt, Matthew Easter, and Paul Turner whose insights and ideas were instrumental to the conclusion of this research. I will miss our conversations...

Thank you everybody!

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1: Introduction

People have always tried to find explanations for how and why things happen to and around them. In ancient times they used divine intervention as an explanation for phenomena they did not understand. Today we have laws of physics, chemistry or mechanics to explain the very same phenomena[1], and while in the past many such causal explanations were simple, advances in science and technology now necessitate a deeper understanding of a multitude of causal relations in order to explain the same basic natural phenomena. Nevertheless, today as in ancient times, people need to understand causally linked events to "make sense" of the world around them (e.g., Hastie, 1984) or to come to terms with it.

The fundamental cognitive process which allows people to understand and use causality is called causal reasoning (Rehder, 2003a). According to existing research people engage in causal reasoning to "make sense" of the world around them (Hastie, 1984; Shank & Abelson, 1977) and to understand the physical world they live in (e.g., Brewer, Chinn, & Samarapungavan, 2000; Carey, 1995; Corrigan & Denton, 1996; Schlottmann, 2001; Thagard, 2000; Wellman & Gelman, 1998). Relying on causal propositions, causal reasoning is the foundation of more complex forms of thinking and reasoning across domains (Jonassen & Ionas, 2006). It is required for making predictions,

[1] For example, why and how the sun rises every day, or why, how, and when rain falls.

drawing implications and inferences, and explaining phenomena. As Dickinson (2001) explains:

> Causal learning and representation is a fundamental form of cognition, if not the fundamental form. Without the capacity to learn about and represent causal relationships between our actions and their consequences, the mind would be radically disconnected from the world. (p. 3).

Causal reasoning is also important for scientific thinking (Carey, 1995; Keil, 1989), as relationships between any conceptual entities that interact directly or indirectly with each other are in most cases, causal. Therefore, to understand a domain people need to construct conceptual models using these causal relations.

While understanding causality is fundamental for both scientific and everyday cognition, as well as for learning, research shows troublesome findings about the many problems people have when dealing with causality and causally linked events (Chapman & Chapman, 1969; Pennington & Hasties, 1988; Tversky & Cahneman, 1989). For example, people display a tendency to favor obvious, localized, simple, linear, and sequential causal relations (Bullock, Gelman, & Baillargeon, 1982; Chi, 2000b; Driver, Guesne, & Tiberghien, 1985; Grotzer & Bell, 1999; Perkins & Grotzer, 2000; Spelke, Philips, & Woodward, 1996; Wilensky & Resnick, 1999). That is, people tend to simplify otherwise more complex causal structures – a process which results in a distorted understanding. Also, when solving well-structured problems people tend to use strategies that convert the problem's elements into calculations while missing the conceptual

underpinnings of the problem or domain – a strategy that leads to poor problem solving performance (Hegarty, Mayer, & Monk, 1995; Lucangeli, Tressoldi, & Cendron, 1998).

Scholars also found that instead of updating or building new schemas, people would rather use their old schemas, even when their old schemas are no longer appropriate (Waldmann & Hagmayer, 2006), as long as these old schemas have sufficient situational predictive power. That is, people tend to show resistance to change, even when an update to their mental model is needed as long as the model they currently hold serves their purposes.

The notion of causality has been analyzed and debated for centuries in the scientific and philosophic community, but scholars have never reached common ground (e.g., Mooney Marini & Singer, 1988). Therefore, today there is no universally accepted definition of causality, although multiple attempts have been made over the years. In this controversy, David Hume (1739/1987) brought the most important contribution to understanding causality by showing that the idea of causation arises from the empirical relations of contiguity, temporal succession, and constant conjunction. More commonly understood, causality represents the relationship between two or more entities where the behavior of one or more of them determines the behavior of the other(s). Also, while fundamental, causality proves difficult to capture. That is, many scholars believe it is not possible to observe causality itself, but rather infer it from observable evidence (effects), such as the results of a process (e.g., Buehner & McGregor, 2006; Hume, 1739/1987; Young & Rogers, 2005).

To observe, recognize and understand causality, people need to process at least two significantly different kinds of information – covariational and mechanistic.

Covariational information, seen by some scholars as the foundation of human causal induction, is emphasized by the human sensitivity to statistical data and indicates the probability of a cause producing an effect (Cheng & Novick, 1990, 1992). It is considered to be the quantitative representation of causal reasoning (Thagard, 1998). The mechanism information is focused on knowledge about the causal processes involved in understanding causality (Ahn & Kalish, 2000; T. R. Schultz, 1982). It qualitatively links the cause to its effect by explaining the chain of causal events that lead from the originating cause to the final effect (Thagard, 1998). Mechanism information provides the "how" and "why" of the causal process.

Prior research conducted on the implementation of a cognitive flexibility hypertext[2] development environment as a means for building cause maps[3] in a marketing strategies class revealed that if one asks students to explain how each causal relation happens they build better and more detailed maps[4] (Strobel, et al., 2008). This observation, supported by the students' opinions expressed during a focus group, suggests

---

[2] Cognitive Flexibility Hypertext (CFH) is a computer-based implementation of Cognitive Flexibility Theory (CFT). CFT is a model for designing nonlinear knowledge bases aimed at minimizing the biases that result from using oversimplified, prepackaged instructional materials. The hypertext provides nonlinear access to multiple representations of a domain's content (Strobel, Jonassen, & Ionas, 2008).

[3] Cause maps, also known as Influence Diagrams, are graph representations that can concisely represent a complex structure of interrelated causal relations. They are the result of a family of causal mapping techniques developed for systematical elicitation and representation of individuals' causal beliefs related to a particular issue or event (Axelrod, 1976; Huff & Jenkins, 2002).

[4] Given an originating context or situation, the students were asked to develop an influence diagram to explain the impact of this originating cause on the performance of a company of their choice. That is, they were asked to show the most important direct and indirect influences the originating cause has on the behavior or status of the target company. This task usually involved developing a directed graph using paper and pencil. In this iteration, the students were asked to use the CFH application to perform the same task using a computer. The instructor evaluated their representations.

that asking people for explanations of causal mechanisms when describing a causal relationship can be beneficial. That is, asking students to explain the mechanism(s) by which the cause produces the effect could provide the means to alleviate some of the problems people have when dealing with causality and causal reasoning.

Further study showed that the explanation of causal phenomena plays an important role in the understanding of causal relationships (Lombrozo, 2007; Slusher & Anderson, 1996; Van Fraassen, 1985). For example Ahn (1995) showed that seeking causal explanations is an innate trait people have, while Slusher and Anderson (1996) found that explanation availability is a mediator for the change in beliefs when causal arguments are involved. Researchers also found that explanations seem to be used by people as a foundation for further estimates of, for example, the probability that an event will occur (Lombrozo, 2007), or that to understand physics, people need an explanation to account for the dynamics of the phenomenon (Besson, 2004). Nevertheless, in many situations an explanation from an external source is not readily available. What is always available, though, is self-generated explanation (Chi, Bassok, Lewis, Reinmann, & Glaser, 1989), or self-explanation, which refers to the explanation a learner generates on his or her own, as opposed to explanations provided by an external source (i.e., instructor, book, etc.).

Self-explanations are usually more effective than explanations provided by others because a) they require students to actively elaborate their prior knowledge, thus triggering more constructive learning processes, and b) are usually better targeted toward the learner's specific problem than to the explanations provided by others. Research (e.g., Chi, et al., 1989; Chi, De Leeuw, Chiu, & Lavancher, 1994; Kastens & Liben, 2007;

Lombrozo, 2007) shows that self-explanation provides significant gains in performance for people using it compared to people that do not (e.g., Kastens & Liben, 2007).

Studied in different contexts and for varied age groups, self-explanation proves to be a domain-independent strategy (Nathan, Mertz, & Ryan, 1994) that most high-performance students use in linking prior knowledge to current content (Bielaczyc, Pirolli, & Brown, 1995). Nathan's study shows that self-explanation works better for conceptual reasoning while providing only marginal advantage for procedural contexts. It has also been shown that either guiding (Bielaczyc, et al., 1995) or prompting (Chi, et al., 1994) people to self-explain improves performance. Studies also suggest that no matter how self-explanation happens, for example spoken aloud, in one's head (Didierjean & Cauzinille-Marmèche, 1997) or written by hand or typed (Aleven & Koedinger, 2002), gains in performance are observed. This observation is also true for situations when feedback on the correctness of the explanation is provided (Aleven & Koedinger, 2002) or when it is not (Chi, et al., 1989; Chi, et al., 1994).

Purpose of Study

The purpose of this study is to investigate the possibility of using self-explanation to help learners reason better in contexts where causal reasoning is involved while at the same time advancing existing research in this area. Specifically, this study proposes to use self-explanation to elicit casual mechanism explanations when reasoning about causally linked events. The use of self-explanation or other cognitive strategies to elicit mechanism explanation can in the end become foundations for developing support methodologies for causal reasoning.

Localized in the field of medicine, a domain that relies on the understanding and use of extensive and complex causal processes, this study attempts to answer the following research question: In the medical field, when learners are reasoning causally, does using self-explanation to elicit an explanation of the causal mechanism(s) improve, on average, learners' performance on tasks involving such reasoning processes?

Design of the Study

To answer the research question this study employs both quantitative and qualitative methods for analyzing the data generated by a completely randomized two-group, control and treatment, between-subjects experiment. The design of the experiment is based on the existing literature and on the lessons learned from two pilot studies aimed at developing this research methodology. For this study the treatment is adapted from a methodology used by Atkinson, Renkl, & Merrill (2003) in which they employed prompts to train learners to identify the underlying principle while solving practice problems.

The treatment consists of three multiple choice questions, similar to those medical students and professionals answer during their coursework and during their certification examinations, and for which the participants assigned to the treatment group are asked to provide both an answer a brief account of what mechanisms might be at work in the presented question or context. Focused on short-term transfer, this intervention is intended to provide participants with opportunities to practice self-explanation of causal mechanisms when considering alternative answers to a causal problem, with the expectation that they will perform better, on average, than the control group on a subsequent similar problem.

The performance measurement task consisted of one multiple-choice question with only one correct answer, for which the participants first chose the answer and then explained the mechanisms that supported their choice. Following this, in an attempt to better understand the participants' reasoning processes, they were offered the opportunity to change their answer if they considered this necessary. At the same time, they were asked to explain why the new choice was better than the previous one.

Since prior knowledge is reported to be important for both causal reasoning (e.g., Ahn & Baileson, 1996; Ahn, et al., 1995) and self-explanation (Chi, et al., 1989), this study includes a brief assessment of the participants' prior knowledge in the three fields relevant to the instrument topic: physiology, pathology, and immunology. The assessment of prior knowledge includes two components. First, the participants were asked to provide a self-assessment of their knowledge in each of the three fields. Second, the participants answered a set of six multiple-choice questions, two for each field. The number of questions in this assessment of prior knowledge was severely limited by the time the students had available for this study, as suggested by the participating institution.

About 350 first- and second-year medical students from a midwestern medical school were invited to participate in this experiment. All efforts were made to ensure the participation of a minimum number of 88 subjects[5].

Research Questions

Q1: Does the practice of self-explanation as causal mechanism elicitation technique affects, on average, learners' performance on causal reasoning tasks?

[5] For a .06 medium effect size and .8 power for $\alpha = .05$, a minimum of 44 subjects in each experimental group is recommended (Keppel, 1991, p. 72).

Q2: How does prior knowledge of relevant domains relate to the learners' performance on causal reasoning tasks?

Q3: How does the use of self-explanation as causal mechanism elicitation technique influence participants' accounts of how and why they chose a certain answer or reach a certain solution?

Assumptions and Clarifications

- Considering a typical course schedule at the medical school where this research took take place, first-year medical students at the end of their second-semester had already completed sufficient coursework to allow them to answer the questions included in this experiment.
- Three domains in medicine: physiology, pathology and immunology are the most relevant areas for answering the questions included in this experiment's research instrument.
- The medical profession relies extensively on causality and causal reasoning; therefore, students in the target group are assumed to be familiar with causally linked events and explanations.
- Medical texts and literature use the term “mechanism” about processes that take place in the human body. The use of the term “mechanism” in this study and in materials presented to the participants is in accordance with its accepted meaning and use in the medical field.

Summary

Causal reasoning, a fundamental cognitive process, allows people to understand and use causality and to understand causally linked events (e.g., Rehder, 2003b) as they

try to “make sense” of the world (e.g., Hastie, 1984). Foundational to more complex forms of thinking and reasoning (Jonassen & Ionas, 2006), causal reasoning is required for drawing implications and inferences, making predictions and explaining phenomena in both everyday cognition and scientific thinking (e.g., Carey, 1995).

While fundamental, understanding and using causality has been found to be difficult for people (e.g., Chapman & Chapman, 1969; Tversky & Cahneman, 1989). Research found that people have a tendency to simplify otherwise more complex causal structures (e.g., Bullock, et al., 1982; Perkins & Grotzer, 2000) as well as to resist changing existing schemas if they hold enough predictive power, even though their old schemas are no longer appropriate (Waldmann & Hagmayer, 2006). In addition, while solving well-structured problems people tend to focus more on computations while missing the conceptual underpinnings.

Understanding causality relies on people observing and processing at least two different kinds of information: covariational and mechanistic. Covariational information provides the probability of the cause producing the effect (e.g. Cheng & Novick, 1990), while mechanism information provides the “how” and the “why” of the causal process.

With research showing that mechanism explanation is important for understanding causal relations (e.g. Lombrozo, 2007) and that people have problems understanding and using it, this study focuses on using self-explanation as a methodology to help people reason better in contexts where causal reasoning is needed. Localized in the medical field which relies heavily on causal reasoning, this study tries to find whether using self-explanation to elicit causal mechanism explanation improves, on average, people’s performance on tasks involving such reasoning processes.

Both quantitative and qualitative methods were used to analyze the data provided by a completely randomized two-group between subjects experiment. To collect data on prior knowledge, performance and demographics, about 350 first- and second-year medical students from a Midwestern medical school were invited to participate. The expected outcome was for participants in atreatment group, who had a chance to practice self-explanation strategies by providing explanations for their choices on three multiple choice questions, to outperform participants in acontrol group, who did not have to provide explanations to the practice questions.

CHAPTER 2: Literature Review

Introduction

Research in cognitive psychology shows evidence that people engage in causal reasoning to "make sense" of the world around them (Hastie, 1984; Shank & Abelson, 1977). Since in the 18th century David Hume provided one of the most important contributions to the understanding of causality (Hume, 1739/1987) by showing that the idea of causation arises from the empirical relations of contiguity, temporal succession, and constant conjunction, a significant amount of research has aimed at understanding what causality is, how people reason causally, how they learn from causal events and process, or how they use their understanding of causality to diagnose past or predict future events.

According to the majority of the scientific community, causality plays a fundamental role in the physical world and that causal reasoning, which is the human understanding of physical causality, is a fundamental part of human thinking (e.g., Thagard, 2000), an essential cognitive skill of higher order cognition (Carey, 1995; Schlottmann, 2001; Thagard, 2000), without which people would not be able to understand the world around them. Since David Hume focused mostly on the process of scientific discovery and less on everyday cognition, early work on causality, causal

reasoning, and causal understanding was mostly from the perspective of either the scientific research or the philosophy of science.

For centuries, scholars have debated the meaning of causality but never reached common ground. Therefore, today there is no universally accepted definition of causality, although multiple attempts have been made over the years (for more details see Mooney Marini & Singer, 1988). Among them, David Hume (1987/1739) provided one of the most important contributions to the understanding of causality by showing that the idea of causation arises from the empirical relations of contiguity, temporal succession, and constant conjunction.

Another important attribute of causality is its conjunctive plurality, which, in essence, represents the fact that an effect is rarely the result of a single cause (Einhorn & Hogarth, 1986; Waldmann & Hagmayer, 2001). Scholars consider that two or more causes are usually involved in producing an effect and that all causes need to be concurrently present to produce the effect. On the other hand, the same scholars recognize the disjunctive plurality in causality, also known as "genuine multiple causation" when an effect is produced independently by any of the causes and the joint occurrence of two or more of the causes does not alter the effect.

Despite its apparent simplicity, the notions of causality and causal reasoning are complex and have been approached by researchers and philosophers across the entire spectrum of human activities. While fundamental, causality and causal reasoning are also elusive concepts which are difficult to capture in research. For example, many scholars believe it is impossible to observe causality itself, but infer it from observable evidence,

such as the results (or effects) of the process (Buehner & McGregor, 2006; Hume, 1739/1987; Young & Rogers, 2005).

Cues to Causality and Causal Attribution

Over the years scientists and philosophers observed that certain elements, such as contiguity and temporal order, have a tendency to indicate causal relationships (e.g. Einhorn & Hogarth, 1986). However, there is no guarantee that the presence of these elements will make people perceive a relationship as causal. In principle, the existing literature agrees upon the following major types of cues to causality.

Covariation between two variables is consistent with the traditional notion of cause. As the foundation for the covariational approach to causality, presented in more detail later in this chapter, covariation occurs when two objects are in constant conjunction with each other (Ahn & Kalish, 2000). It is also contended that covariation does not need to be perfect for people to recognize a causal relation (Einhorn & Hogarth, 1986; Kushnir & Gopnik, 2007). Waldmann (1996) shows that covariation data is more informative when used in combination with other assumptions about the system in which the causal relation takes place.

Temporal order of succession has been found to be essential in determining which of the two variables that co-vary is the cause and which is the effect. Nevertheless, while the temporal order has a fundamental importance for causal judgments, it has no role in a formal probability theory (Beauchamp, 1974; Hume, 1739/1987; Mooney Marini & Singer, 1988). Temporal succession, related to the criterion of causal priority (asymmetry), requires a cause to be present for the effect to occur. Therefore, the common understanding of causality assumes that causes occur before their effects.

Lagnado et al. (2007), studying the interaction between temporal order, intervention, and covariational cues found that people use both temporal-order and interventional cues to infer causal structure. They also found that these cues dominate the available statistical information. Their research suggests that people might be using cues, such as temporal order, to generate initial models in order to test them later against covariational data.

Conversely, the use of temporal succession as an indication of causality can be misleading, especially in the social sciences, since the temporal order in which the behavior occurs depends on the individual's abilities to anticipate and plan the future. That is, temporal succession can be viewed as teleologically determined. As a result, causal priority is constructed in the mind in a way that is not reflected by the actual behavioral sequence (Mooney Marini & Singer, 1988).

Contiguity in time and space plays an important role in directing people's attention to contingencies between variables (Michotte, 1946). Relying on these types of cues, causal inference becomes more difficult when temporal and/or spatial contiguity is low (erratic). Contiguity is particularly prone to conflict with other cues, which can in turn create the need for compromises.

Research focused on the influence of time delay on how people judge causal strength shows that as the delay between two events increases, people judge the causal strength of the relationship to be weaker (Shanks, Pearson, & Dickinson, 1989). This holds unless people have reason to expect a delay between the cause and the effect (Buehner & May, 2002; Buehner & May, 2003; Bullock, et al., 1982; Hagmayer & Waldmann, 2002; Tangen & Allan, 2004). Replicating Shank's experiment Reed (1992,

1999) showed that people fail to identify causal relations if there is a more than two second delay between cause and effect.

Considering that these findings are contrary to the richness of everyday cognition, where people seem to be able to recognize without effort causal relations that involve considerable delays, Buehner and May (2003) found that the impact of a temporal delay between cause and effect is largely dependent on the individual's expectations about the timeframe of the causal relation. Their study suggests that free-operand procedures used in previous studies, mainly based on animal behavior, are not well suited to study the direct influences of temporal delays on causal induction because they confound the delay with weaker evidence for the relationship being studied.

Studying preschool children's causal assumptions about spatial contiguity and its use in learning new causal relations, Kushnir and Gopnik (2007) found that correct causal inferences were more likely to happen when causes were spatially close to each other, especially when the subject was facing ambiguous evidence.

Similarity of cause and effect is, according to Nisbett and Ross (1980), a philosophically challenging concept which comes into play when people use, for example, the congruity of the length and strength of two variables, the cause and effect. That is, when the target effect is large in duration and magnitude, for example, people expect the cause(s) to be of comparable duration or magnitude.

Using their own research as well as other prior experimental accounts, Einhorn and Hogarth (1986) summarize the characteristics of causal cues. First, the relation between each cue and the causal phenomenon is probabilistic. That is, while the cue might exist, the causal relation might not, and vice versa. Second, people use multiple

cues in their causal inferences as means to limit the extent of potential errors linked to inferring causality from single cues. Third, the use of multiple cues is often facilitated by redundancy, which reduces the negative effects of omitting cues as well as aids in focusing attention on others. Fourth, while multiple cues can reduce uncertainty, conflicting cues can increase it.

Theories and Models of Causal Reasoning

Initially the domain of philosophers (e.g., Aristotle, Hume, and Kant), theories on causal reasoning have evolved to fit one of two major groups, depending on how they combine the information about covariation and mechanisms (Griffith, 2005): (1) covariation based approaches and (2) mechanism-based approaches. Covariation-based approaches consider that human causal induction is due to a domain independent statistical sensitivity to covariation between cause and effect (Cheng & Novick, 1990, 1992), while the mechanism-based approaches focus more on the role the knowledge about causal mechanisms plays in causal reasoning processes (e.g. Ahn & Kalish, 2000; L. E. Schultz & Griffith, 2007). The models in both approaches explain people's causal judgments, the difference between them being in the emphasis on information about covariation or mechanism.

Covariation-Based Approaches

The covariation-based approaches are successors of the Humean philosophy. They are based on the assumption that an event that exhibits a regularity of association with an effect (i.e., covaries with that effect) is more likely to be identified as a cause to that effect than an event that does not display such regularity of association. For example, based on this principle, the proponents of the *contingency model of causal induction*

(Jenkins & Ward, 1965; Salmon, 1989) argue that causal knowledge is derived from observable events and thus acquired from sensory input (Cheng, 1997). This approach remains the foundation for scientific inquiry ever since David Hume brought it to the attention of the scientific community in the late 1700s.

The fundamental question for all these models is "How do people take advantage of regularities in their environment to induce causal relations and construct mental models based on them?" (Holland, Holyoak, Nisbett, & Thagard, 1986; Nisbett & Ross, 1980). The research has focused on using covariation as a major source of information to support causal relations. The general goal was to find a set of rules people use when they make causal inferences using a set of covarying events.

Starting from the principle of covariation, Kelley (1967) thought of people as "intuitive scientists" who use a mechanism of causal induction similar to ANOVA (Analysis of Variance). He developed his *ANOVA model* in the context of studying causal attributions in interpersonal changes. Kelley's model proposes three dimensions, person (P), stimuli (S), and time/modalities (T) as independent variables. To measure consensus along theses three dimensions, Kelley proposes three information variables: consensus, distinctiveness and consistency. Causal attributions result from comparing the target event with other events along the three dimensions. According to this model, the configuration of these three information variables will determine if the effect will be attributed to a person, a stimuli or to a situation. Kelley's experiments (Kelley, 1967, 1973), as well as later replications strongly support this model (see Ahn, et al., 1995for details).

Subsequent research on causal attribution found that Kelley's ANOVA model presents a number of issues such as (a) a bias against using consensus information, (b) people favoring personal attributions over stimulus and occasion attributions or (c) a tendency towards unpredicted, especially conjunctive, attributions (e.g. Joseph M. F. Jaspars, 1988; Nisbett & Ross, 1980). In response to these issues, Cheng and Novick (1992) proposed their *Probabilistic Contrast Model*. Fundamentally, the model proposes that everyday causal inference is based on contrasts between various combinations of values and it is a probabilistic interpretation of the covariation principle. The model assumes that, for any situation, people have preexisting conceptions about potential causal factors, for which they compute covariations for situations in which the factors are present or absent. The final causal attribution is made if the proportion of events for which the effect occurs in the presence of the factors is noticeably different from the proportion of events for which the effect occurs in the absence of the factors (Ahn, et al., 1995). In essence, the model is still based on covariation, but the covariation is expressed in probability terms that can be computed for any dimension, as opposed to purely through information variables.

In mathematical terms, the probability of an effect occurring when the cause is present is represented by $P_{(e|i)}$ while the probability of the effect occurring when the cause is absent is represented by $P_{(e|\sim i)}$. Using this notation, causal roles are defined empirically using the following contingency rule:

$$\Delta P_i = P_{(e|i)} - P_{(e|\sim i)}$$

Considering the value obtained for ΔP_i the following attributions can be made: (a) if ΔP_i is positive, the candidate cause i is a facilitator in producing the effect, (b) if ΔP_i is

negative, i is an inhibitor, while (c) when ΔP_i is zero, i is a non-causal factor. In essence, the proponents of contingency type models claim that people are sensitive to the contingency ΔP_i between the cause and the effect and that they use this information as a measure of causal strength or likelihood (Cheng, 1997; Cheng & Novick, 1990, 1992).

As a response to the same issues presented above, Jaspars et al. (1983) propose a deterministic interpretation of Kelley's ANOVA model that they called the *Inductive Logic Model* of causal attribution. According to this model, each piece of information given in an attribution problem is coded in terms of whether it involves (a) the individual in question or other people, (b) the stimulus in question or other stimuli and (c) the occasion in question or other occasions, while the presence or absence of the effect or behavior is noted. According to this model, people observe each potential causal factor or group of factors and note if it or they are present when the effect is present and absent when the effect is absent. In the end, they identify a factor or group of factors that is both necessary and sufficient for the effect to occur.

Using the earlier *Probabilistic Contrast Model* (Cheng & Novick, 1990, 1992) and using elements from the *Causal Power Theory* (Harre & Madden, 1975; P. A. White, 1989), Cheng and Novick (1992) propose the *Causal Strength Model*[6]. While the causal strength model borrowed heavily from the earlier probabilistic contrast model, it was designed to emphasize the probabilistic nature of human causal reasoning.

Later on, Cheng (1997) rejected ΔP as a measure of causal strength, considering it only a measure of covariation, but not causation (Cheng, 1997; Cheng & Novick, 2005), because human judgments suggest assumptions about causality that are not purely

[6] In some works, Cheng's (1992) Causal Strength Model is alternatively called Causal Powers Model, which is different from Harre's (1985) Causal Power Theory.

covariational measures (e.g. ΔP) and conventional statistics. As a consequence, Cheng (1997) proposes what is today one of the most widely agreed upon theoretical frameworks, the *Power PC Theory*[7], as an attempt to make these assumptions explicit. For that she proposes that human causal judgments correspond to "causal power", the probability that the cause C produces the effect E in the absence of all other causes. Therefore, for a generative cause, the causal power can be estimated from contingency data as:

$$power = \Delta P/(1 - P_{(e+|c-)})$$

In this case, the causal power takes ΔP as component and predicts that ΔP's effect will be greater when $P_{(e+|c-)}$ is large. Oppositely, for preventive causes, the causal power can be calculated using the formula below:

$$power = - \Delta P/P_{(e+|c-)}$$

According to this formula, for preventive causes, the smaller $P_{(e+|c-)}$ is, the greater ΔP's effect on the causal power.

Another group of theories, in the same category with the covariation models, use direct acyclic graphs[8] (DAGs) called *Bayes networks* to represent networks of causal relationships (Pearl, 1995, 2000). In this graphical representation, nodes represent variables of interest, such as the features of an object, and links represent causal inferences between these variables. The strength of a causal relation between two of the variables in the graph is represented by the conditional probability that is attached to each causal relation. In a sense, a Bayes network is closer to a model of the environment than

[7] Stands for "Causal Power Theory of the Probabilistic Contrast Model"

[8] Also named "influence diagrams" or "cause maps" by some authors, though the latter two have a wider application than the Bayes nets.

to other knowledge representation schemes (e.g., neural networks). These Bayes networks attempt to simulate how causal mechanisms operate in the environment.

The inner workings of Bayes networks are based on a variety of inference algorithms. The first such mechanism for updating the probabilities in a Bayes network used a message-passing architecture (Pearl, 1998) and worked only on causal trees of singly connected networks. Since Pearl's first attempt, many other techniques have been developed to extend the tree propagation method to more general, multi-connected networks (Shachter, 1988). With roots in statistics, Bayes networks have recently started to be used in psychology research (Glymour, 2001; C. Glymour & P. Cheng, 1998; C. Glymour & P. W. Cheng, 1998; Gopnik, et al., 2004; Griffith & Tenenbaum, 2005; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003).

By using graphical means to represent probabilistic information and knowledge, Bayes networks are able to overcome many of the conceptual and computational difficulties of earlier knowledge based systems. Graphical models provide the means to maintain consistency and completeness in the context of working with probabilistic information and knowledge and to define modular procedures for knowledge acquisition that reduce the amount of work required. In addition, interdependencies can be dealt with explicitly (Pearl, 1998). Despite their utility, Bayes networks have an important limitation: they cannot include cycles, and therefore they allow for only an incomplete representation of certain events and processes (Mazlack, 2007).

While popular, the covariational theories of causal reasoning have also raised criticisms among scholars, exposing a series of fundamental problems. For example, people can confidently infer a strong causal link between two events after only one

instance or co-occurrence of the cause and its effect (Beasly, 1968; Boyle, 1960; Read, 1983). Also, well known in statistics, the fact that covariation does not imply causation is also an important criticism, together with the fact that these covariation-based models cannot explain the phenomenon. Other observed problems or issues are the discounting of causes when there is a difference in strength between alternate options (Goedert & Spellman, 2005) and the fact that the covariation-based models are not sensitive to the temporal aspects of causality.

Mechanism-Based Approaches

Historically speaking, the mechanism view to causal reasoning was probably brought by the research on category formation and causal learning (Buehner & Cheng, 2005). While related, research in this area has not caught up with the research on covariation-based causal reasoning. Nevertheless, psychologists came to the consensus that the notion of causal mechanism is fundamental to people's concept of cause and causation (Ahn, et al., 1995; Bullock, et al., 1982; Cheng & Nisbett, 1993; L. E. Schultz & Griffith, 2007). The accepted interpretation is that the mechanism is the element that makes the difference between true causal relations and mere correlations or coincidences (Ahn & Kalish, 2000; Schlottmann, 2001). It is not the causal link in general that determines a correlation, but the mechanism underlying that causal link that produces it (Cheng, 1997; Maldonado, Catena, Perales, & Candido, 2007). It has been argued that people's ability to infer a causal relation is contingent upon knowledge of or insight in the causal mechanism that produces the effect given a candidate cause (Ahn, et al., 1995). In a series of studies about the way people combine causal mechanism[9] information with

[9] For example, a network of intervening causal relations.

covariation information Rapus (2004) found that the way the information about the strength of the covariation is used depends upon the level of detail of the mechanism information and the scope over which covariation information is derived.

While it seems that there is no generally accepted definition of what a causal mechanism is, many scholars agree that the mechanism is the abstract, unobservable glue between observed events. That is, while physical systems could be analyzed at progressively finer levels, the mechanism, in its absolute sense, always eludes capturing (Schlottmann, 2001). In the same vein, Mazlack (2007) observes that while causal knowledge provides support for a deep understanding of a system and that potential control over a system comes from being able to predict its outcomes, knowledge of at least some of the relationships is inherently imprecise. He also observes that usually, commonsense causal reasoning is more successful in reasoning about a few large-grained events than about many finer-grained ones. That is, simplicity (or simplification) helps causal reasoning, but larger-grained causal objects are by definition more imprecise.

A mechanism explains how a phenomenon is produced (Machamer, Darden, & Craver, 2000), how tasks are carried out (Bechtel & Richardson, 1993), or how the mechanism as a whole behaves (Glennan, 1996). Hobbs (2001) looks at mechanisms as "causal complexes[10]", a collection of events which produces an effect by their occurrence or non-occurrence. Similarly, Glennan (2005) considers the mechanism a complex system that produces a behavior through the interaction of its component parts, where the interactions between parts can be characterized by "direct, invariant, change related generalizations".

[10] A causal complex is the complete set of events and conditions necessary for the causal effect to occur (Hobbs, 2001).

Mechanisms are also characterized as "entities and activities organized such that they are productive of regular changes from start or setup to finish, or termination conditions" (Machamer, et al., 2000). In one of these views, mechanisms are seen as regular because they usually work or behave in the same way under the same conditions. This regularity is present in the way the mechanism (process) runs from the beginning to the end, but what makes it regular is the productive continuity between stages (Darden, 2002). In another view, activities are producers or enablers of change, while entities are the "things" that engage in these activities. That is, in this view, entities and activities are interdependent. From another perspective, Thagard (2000) showed that inferring a causal relation is not contingent on knowledge about a causal mechanism, but is enhanced by it.

The mechanism approach differs from the traditional covariational approaches in at least two important aspects:

- It involves different types of information seeking behavior, by asking "how?" rather than "what?" questions;
- It requires the individual to use prior knowledge to reach a conclusion.

The mechanism-based approaches to causal reasoning seem to be more cognitively complex than their covariation-based counterparts (Hung, 2004). A claim scholars make about the mechanism-based approaches relies on the fact that the mechanism resides on a different cognitive level than the cause and the effect, a level that is removed from the concrete conditions of the underlying evidential phenomena. Removing the mechanism from its concrete context creates the premises for mechanism reuse in other contexts (Kitcher, 1989; Salmon, 1984). As Ahn and his colleagues (1995) suggest, when seeking the cause of an event, individuals primarily try to discover the

processes or mechanisms underlying the relationship between the cause and the effect. In expressing a causal relationship, for example, explaining an accident, the mechanism approach would argue that people would seek explanations that go above and beyond the relationships that are at the level or the target effect. That is, causal explanations would seek to involve a new set of theoretical entities – theories or processes not present in the event description.

It is known that highly complex systems, such as a living cell, organism, or an organization have many concurrent mechanisms working at any given time. That is, they undergo several intertwined processes, at the same time, both on the same level and on different levels. The coexistence of parallel mechanisms is especially noticeable in bio- and social systems. As a result, because more than one mechanism may be at work at any given time in the same system, distinguishing between essential and non essential mechanisms may be necessary. Therefore, scholars proposed various frameworks to deal with such complexity (Glennan, 2002; Machamer, et al., 2000; Railton, 1978; Salmon, 1984, 1989). Elaborating on Railton's (1978) work on causal-mechanical explanation, Salmon (1984, 1989) proposes a concept of what he calls "causal nexus", which he considers to be a vast network of interacting causal processes. Alternatively, borrowing from general systems theory, other scholars propose a complex-system approach to mechanisms (Glennan, 2002), rooted in the work of Wismatt and Herbert Simon. As Glennan proposes:

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the

> interaction between parts can be characterized by direct, invariant, change-related generalizations (Glennan, 2002, p. S344).

Similarly, Glennan (2005) discusses the relationship between mechanisms and people's models of these mechanisms. His research shows that there is no Baconian crucial experiment to decide between competing models.

From a mainly cognitive perspective, Machamer et al. (2002) consider mechanisms a way to explain how phenomena are produced. Taking an action oriented perspective, Bechtel and Richardson (1993) think of mechanisms as the way tasks are carried out. Thus mechanisms "are entities and activities organized in such a way that they are productive of regular changes from start or setup to finish or termination conditions." (Darden, 2002, p. S356)

In understanding and explaining human behavior, Budesheim (1998) suggests that, when exemplars of an individual's behavior are readily available, they will be used before using abstract knowledge if the participants were motivated to engage in effortful processing. This finding suggests that when no effort is required, such as in the case of everyday cognition, people display a tendency to use exemplars they already processed rather than fundamental knowledge. That is, motivation seems to play an important role in what kind of information and reasoning processes people use.

Problems and Issues with Causal Reasoning and Understanding in Humans

From very early on scholars observed that people have problems dealing with causality and causal relationships. Empirical work spanning many years shows that people have problems when reasoning about causes (Chapman & Chapman, 1969; Shanks, Medin, & Holyoak, 1996; Sperber, Premack, & Premack, 1995). For example,

the work of Chapman and Chapman (1969) suggests that people's misunderstanding of causality can result in creating spurious correlations. Pennington and Hasties (1988) suggest that people who do not understand correctly the causality of a phenomenon weigh the same evidence differently, while Tversky and Cahneman (1989) suggest that people reach incorrect conclusions about conditional probabilities.

In their research Schustack and Sternberg (1981) also show a number of such issues. They observed that people tend to favor confirming rather than disconfirming information. That is, people tend to associate more importance to information that confirms their personal theories versus information that disproves them. They tend to "overreact" to negatively presented evidence versus positively presented evidence. That is, there seems to be a strong emotional component to causal reasoning that in certain situations hinders people's ability to reason causally. The same research also shows that people tend to focus too much on the cause to be evaluated to the detriment of other sources of information available to them. In other words, people seem to have a difficulty observing a causal network. They also tend to emphasize the information source at hand versus the environment of which the information source is part.

Another growing body of research (Chi, 2000a; Driver, et al., 1985; Perkins & Grotzer, 2000; Wilensky & Resnick, 1999) supports the findings that people have problems when dealing with causality when engaged in learning activities. It suggests that students have limited notions about the nature of the cause and the effect. A number of studies in ecology (Bell-Basca, Grotzer, Donis, & Shaw, 2000; Green, 1997) show that students' causal reasoning does not correspond to that of scientists. This finding has also been documented in other areas, such as electricity (e.g. Grotzer, 2000; Shipstone, 1985),

ecosystems (e.g. Green, 1997), and force and motion in physics (Halloun & Hestenes, 1985; B. Y. White & Frederiksen, 1995).

For example, in studying pressure (Bell-Basca & Grotzer, 2001), when determining the causes of pressure-related events, students tend to reason using obvious variables rather than considering non-obvious ones. They tend to overlook the systemic nature of a pressure related event and think linearly. In addition, for many students, pressure is unidirectional rather than omnidirectional. These findings are supported by earlier studies (Grotzer & Bell, 1999) which show that students expect obvious causes and effects while often missing effects that involve systems in equilibrium or those that involve passive agents. In other words, students tend to expect obvious causal relationships. Other research (Spelke, et al., 1996) suggests that while students are good at recognizing local or localized relationships, they have difficulties recognizing actions at a distance, either spatial or temporal. This suggests the difficulties in observing causal structures of phenomena.

A study by Anat and Pinchas (1991) in biology shows that student have two main difficulties when attempting to understand causal relationships: 1) inability to organize events according to their correct temporal sequence and 2) inability to identify an event which had been caused by another (given) event.

Research suggests that students tend to assume that the causal patterns they come across are simple, linear and sequential, and in which temporal priority is predominantly visible, linking the causes with their effects (Bullock, et al., 1982). For example, in physics, Barbas and Psillos (1997) show that in studying electricity students tend to think linearly, indicating difficulties in understanding of omni-propagation of disturbances in

electrical circuits. Their study also shows a lack of association between the qualitative description (for variables at the macroscopic level) and underlying microscopic models. In physics education, Besson (2004) shows that students often confuse conditions with actual causes (efficient with contingent causes), that they have difficulties connecting local causes to global effects and that they show a tendency to skip intermediate objects[11]. Grotzer (2000) shows that students' explanations tend to be simple, linear models whereas scientific explanation involves analyzing cause as part of a relation or an interaction. The same study suggests that the type of causality underlying students' models tends to be simple, conductive for simplified interpretations of the information in complex models[12].

In a research paper reporting on the "Understanding the Consequences (UC) Project," Grotzer (2003) shows that exposing students to the underlying causality of phenomena through specifically designed activities and engaging them in discussions about this causality helps them outperform students who learn the same topics, but without these features. These findings held across a range of ages and topics, such as electricity, ecosystems, density and air pressure.

In a qualitative, phenomenographic study, Fyrenius et al. (2007) show that students display frequent misconceptions about underlying principles in medical physiology. He suggests that while causal reasoning is often rewarded in examinations, very long and complex causal relations can be rote memorized instead of understood. These findings are also supported by Dahlgren (1997), who shows that college education

---

[11] To "displace" causes, either skipping intermediate objects, or merging them into a larger one.

[12] Simplified models can be useful in many aspects of everyday life, but they can also distort the scientific interpretation to the extent when it becomes unusable.

will affect students' acquisition of terminology and of problem solving algorithms, but it is rarely able to affect the understanding of central phenomena. Support is also provided by Patel et al. (1989) who show that students often inconsistently apply causal rules based on science information. Their study showed a prevalence of superficial identification of concepts in novices, when more experienced students were able to use general laws. Surveying teachers of physiology about sources of students' difficulties in learning the discipline, Michael (2007) found three main problems. First is that the nature of the discipline, which requires causal reasoning, uses graphs and mathematics, and is highly interactive. Second is the way the discipline is taught. The third is related to what students bring to the task of learning physiology: their belief that learning and memorization is the same thing, their inability or lack of willingness to integrate, and their tendency to compartmentalize.

Indirectly linked to people's difficulties with understanding causal relationships is Waldmann and Hagmayer's (2006) finding that people show a strong tendency to use old conceptual schemas rather than switch to new ones even if and when old categories are no longer appropriate for predicting the new effect. That is, people tend to show resistance to change when updating their causal mental models is needed. Therefore, research suggests that without the necessary expert structural knowledge and a reflective sense of where it applies, students risk imposing limiting structures on new information, which ends up in distorting their own understanding to fit a usually less complex causal structure (Grotzer, 2000; Slotta & Chi, 1999; Wilensky & Resnick, 1999).

Explanation and Self-Explanation

While overlooked and underused (Strobel, et al., 2008), the explanation of the causal phenomenon plays an important part in the understanding of causal relations (e.g. Lombrozo, 2007; Slusher & Anderson, 1996; Van Fraassen, 1985). Also, as Ahn's (1995) work suggests, seeking causal mechanisms (or explanations) seems to be an innate trait people have, as this behavior looks to be "natural". According to the Internet Encyclopedia of Philosophy (Mayes, 2005), historically, the concept of explanation has been associated with causation. That is, to explain an event or phenomenon is to identify its cause in one of two ways: realist or epistemic. The realist explanation is a literal description of reality while the epistemic explanation holds that the role of explanation is to facilitate construction of a consistent empirical model and not to replicate reality. By emphasizing the reason for which things happen, explanation goes beyond mere description, and from a learning perspective, it holds a special place as one of the six core critical thinking skills[13] (Irani, 2006).

Looking at explanations Slusher & Anderson (1996) found that explanation availability is a mediator for the change in beliefs when causal arguments are involved. In other words, asking for an explanation improves the probability for an individual to understand the problem s/he is facing. Studying the role of simplicity and probability in evaluating causal explanations, Lombrozo (2007) found that explanations seem to be used by people as foundation for further estimates (e.g., probability of occurrence). Besson's (2004) study in physics education suggests that to really understand a physical phenomenon people tend to need an explanation to account for the dynamics of that

[13] As defined by the Delphi study on critical thinking.

particular phenomenon. From the scientists' point of view Bechtel and Abrahamsen (2005) suggest that an explanation in a causal context usually employs multiple forms of representation by including, for example, both graphical and textual presentations.

While important, the general domain of explanation is too broad for the purpose of this study. Because the experiment is built using tasks the participants will be performing alone, by themselves, the concept of self-explanation (e.g., Chi, et al., 1989) seems to be better suited. The term self-explanation or self-generated explanation (Chi, et al., 1989) refers to the explanations a learner generates on his or her own as opposed to the explanations provided by an external source (i.e., instructor, book, etc.). Reported gains in science education attributed to the use of self-explanation are overwhelming, with self-explainers often performing twice as well as the non-self-explainers (Kastens & Liben, 2007). Self-explanations are usually more effective than the explanations provided by others because (a) they require students to actively elaborate their prior knowledge, thus triggering more constructive learning processes, (b) they are usually very well targeted toward the student's specific problem, and (c) they are available at the time and place of the learner's need.

Acknowledging that many studies show only that learning and self-explanation co-occur, VanLehn & Jones (1993) propose three possible explanations for why self-explanation works. First, self-explanation seems to motivate learners to detect and fill gaps in their domain knowledge. Second, it seems to help learners abstract solutions and procedures from the current context to a more general description of the problem. And third, self-explanation seems to induce what is called "an analogical enhancement," that is, a richer elaboration of the example or case which facilitates later analogical problem

solving. In their study, the authors found that gap-filling seems to produce most enhancements in learning. Lombrozo (2006) shows that engaging in self-explanation can affect the probability assigned to causal claims, the way probabilities are generalized and learning in general. The reason behind his findings seems to be the fact that explanations accommodate novel information in the context of prior knowledge while doing this in a way that fosters generalization. He finds that even for complex domains, self-explanation is more effective at improving learning and generalization than reading text or receiving feedback.

In a study of fifth-graders learning mathematical equivalence, Rittle-Johnson (2006) found that self-explanation facilitated learning the correct procedures, allowed for these procedures to be adapted to solving novel problems, and boosted procedure retention over time. The study found that children who did not explain had a tendency to revert to old, incorrect procedures. Three mechanisms by which self-explanation helps learning were suggested: a) helped with invention of new problem-solving approaches, b) broadened the range of problems to which children applied correct procedures accurately and c) supported procedure adaptation to solving novel problems in situations when rote application of a known procedure was not possible.

So far, researchers report mostly positive effects when learners use strategies based on self-explanation for a wide variety of topics and learning situations. For example, Chi and her colleagues (1994) show that self-explanation can improve learning when used for learning declarative knowledge from a text. Using a text on the human body circulatory system Chi showed that high explainers (learners who generate a large number of self-explanation statements) had greater understanding of the subject than low

explainers. In her study, all high explainers constructed the correct mental model, while many of the students in the control group as well as the low explainers did not. Exploring the impact of different prompts to students' learning from a text on evolutionary biology, Coleman, Brown and Rivkin (1997) showed that the students who were instructed to explain performed better than those who were instructed to summarize or listen to another student's explanation or summary. The performance increase was strong for both near and far transfer tasks.

For learning from text, a study by Kintsch (1994) suggests that difficulties in reading text might be more conducive to learning when prior knowledge of the domain is adequate. The study suggests that for many the reading process can be too easy, without engaging learners in deep reasoning processes. In these situations strategies for active reading, of which self-explanation seems to be particularly successful, can be supportive for activities that are more constructive and thus improve learning outcomes. Further studies (Ainsworth & Burcham, 2007) show that in reading, self-explanation serves different purposes, depending on the quality of the text. For less coherent texts, self-explanation seems to only help compensate for weaknesses and gaps in the text, while for highly coherent texts it seems to lead learners to detect and repair flaws in their mental models. This research leads to the conclusion that self-explanation is effective when learners use well-structured and explicit texts. A recent study (Yeh, Chen, Hung, & Hwang, 2009) found that the effect of self-explanation prompts on learning depends on the learner's prior knowledge of the domain. For learners with high prior knowledge, prediction-based prompts seem to work best whereas for learners with low prior knowledge, prompts related to reasoning seem to work better.

In the same vein, Bielaczyc et al. (1995) show that self-explanation is a strategy most high-performance students use, for example when linking current concepts with prior knowledge. In his study, the effectiveness of the strategy depends on the learners' domain-general and domain-specific knowledge, the comprehensiveness of the problem being studied as well as the state of students' evolving understanding.

A study by Kirschner, Sweller & Clark (2006) on the advantages and disadvantages of using guidance during instruction shows that the advantage of external guidance diminishes only when learners have sufficient prior knowledge to be able to provide guidance for themselves. That is, for learners with low levels of prior knowledge of a domain, constructivist approaches are not as effective. Both assisting and open prompts have been found to foster both principle-based and rationale-based self-explanation (Berthold, Eysink, & Renkl, 2009). Assisting self-explanation prompts seem to have additional value when compared to open prompts, especially for integrating multiple perspectives, but they come at a cost: they are time consuming.

Looking from the perspective of individual differences, Renkl (1997) shows that successful learners tend to employ more principle-based explanations, explicate the operator-goal combination and use anticipative reasoning. In other words, they assign meaning by identifying underlying domain principles, by identifying goals and sub-goals and exhibit a tendency to anticipate the next step in the solution instead of looking it up. The same study shows that less successful learners use more metacognitive monitoring, that is, they explain a greater number of comprehension problems, indicating a metacognitive awareness of their learning own difficulties.

While performance increases were strong in most of the situations, a study by Pirolli & Recker (1994) on the impact of self-explanation on computer programming skills suggests that while effective, more self-explanation may not necessarily be better. That is, they believe in the law of diminishing returns when using self-explanation based strategies. Pirolly & Recker consider that there are only a limited number of possibilities one can elaborate on a given text or problem until it becomes repetitive. Their research seems to point to the fact that while further pursuit of self-explanations might increase comprehension, it might not be worth the cost in terms of time and effort spent on task.

Further studies on problem solving in physics (Chi & Bassok, 1989), computer programming (Pirolli & Recker, 1994), learning chess (De Bruin, Rikers, & Schmidt, 2007), or even emotion understanding (Tenenbaum, Alfieri, Brooks, & Dunne, 2008) show the effectiveness of self-explanation based strategies. The existing evidence seems to suggest that self-explanation is a domain independent strategy which, once learned, can be successfully used outside the domain in which it was first acquired. Other studies which support this conclusion investigate spreadsheet use in accountancy (Reinmann & Neubert, 2000), experimental design (Lin & Lehmann, 1999) and reading maps (Kastens & Liben, 2007). Looking at self-explanation as a domain independent strategy, Nathan et al. (1994) found that it works better for conceptual reasoning while providing only marginal advantage for procedural contexts. He also noted a decrease in performance when cognitive load is significant, a situation that suggests the existence of a competition for cognitive resources.

Studies show that self-explanation has significant positive impact on performance either when it is presented to others by speaking aloud, or when it happens in one's head

(e.g. Didierjean & Cauzinille-Marmèche, 1997). Both writing or typing self-explanations supports performance improvements, as a study by Aleven & Koedinger (2002) suggests. Researchers also documented the fact that self-explanation based strategies help both when feedback on the correctness of the explanation is provided (Aleven & Koedinger, 2002) and when it is not (Chi, et al., 1989; Chi, et al., 1994). Research also shows that either guiding (e.g. Bielaczyc, et al., 1995) or prompting (Chi, et al., 1994) people to self-explain improves performance.

A significant body of research (e.g. Crippen & Earl, 2007; Mitchell, Mertz, & Ryan, 1994) is interested in using worked examples associated with explanations and self-explanations to support learning. This research shows that good problem solvers tend to engage in more self-directed elaboration than poor problem solvers when studying worked-out examples in physics (Mitchell, et al., 1994), and that instructional explanations are mainly requested by learners with a low level of prior knowledge (Renkl, 2002). Another study (Atkinson, et al., 2003) found that fading of worked-out examples improves performance on near-transfer tasks, but does not have a reliable effect on far-transfer tasks. The same study shows that including prompts to help self-explanation (such as asking for identification of the principle or mechanism) supports medium to large effects in both near- and far-transfer tasks, with no additional time on task. A more recent study (Crippen & Earl, 2007) shows that a combination of worked-out examples and self-explanation prompts produces improvements in performance, problem solving skill and self-efficacy. If the same worked-out examples are being used without prompting, the intuitive sense learners might have about how to use the examples does not always work.

Self-explanation was studied in the context of gap-detection and gap-filling (Chi & Bassok, 1989; Lin & Lehmann, 1999; VanLehn, Jones, & Chi, 1992), mental model revision (DeLeeuw & Chi, 2003), conflict detection and resolution during knowledge integration (Chi, et al., 1994), thought organization (Lin & Lehmann, 1999) and error detection and self-correction (Kastens & Liben, 2007). By construct, scholars have looked at schema formation and case-based reasoning (Didierjean & Cauzinille-Marmèche, 1997), analogical enhancement (Reinmann & Neubert, 2000), visual/verbal integration (Aleven & Koedinger, 2002), construction of new knowledge (Chi, et al., 1989; DeLeeuw & Chi, 2003; VanLehn, et al., 1992; Wong & Lawson, 2002), connection of principles to action (Lin & Lehmann, 1999), and situational model building (Kintsch, 1994). Self-explanation seems to be a versatile strategy that can be implemented in a variety of learning contexts, both in face-to-face and online environments.

Summary

As one of the fundamental cognitive processes, causal reasoning and the associated causality have both received constant attention over the years. Initiated by David Hume in the 1700s, causality and causal reasoning have been studied by scholars from a variety of backgrounds and perspectives (Bunge, 2004; Cheng, 1997; Fugeslang & Thompson, 2000; Fyrenius, et al., 2007; Grotzer, 2003).

In the past, a covariational, statistics-based view on causality has emerged and received most of the attention of the scientific community (e.g., Ahn & Kalish, 2000; Anderson, 1990; Cheng, 1997; Cheng & Novick, 1990). While historically present, in recent years the mechanistic view on causality has been revitalized (Darden, 2002; Glennan, 1996, 2002; Machamer, et al., 2000). Although the two views on causality were

for some time considered somehow parallel approaches to the same domain, in recent years scholars have begun to come to a common understanding that the two are more or less complementary, with the covariational approach providing the hard data to the mechanistic approach, which in turn, provides the power of causal models (e.g., Axelrod, 1976; C. Glymour & P. W. Cheng, 1998; Huff & Jenkins, 2002). This more recent research suggests that the causes and their effects are facts and that the causal mechanisms or processes that link them, when explained, reside on a higher cognitive level (e.g., Ahn & Baileson, 1996; Kitcher, 1989; Salmon, 1984).

While some scholars have studied the nature of causality and the associated causal reasoning processes, others have taken a different route and have looked at how people use causality in learning, work and everyday life. Spanning many years, this research (e.g., Chapman & Chapman, 1969; Pennington & Hasties, 1988; Shanks, et al., 1996; Sperber, et al., 1995; Tversky & Cahneman, 1989) shows some troublesome findings in that people have problems dealing with causality and causally linked events. The same research tends to suggest that this trend is more important for learners as students have a tendency to favor obvious, localized, simple, linear, and sequential causal relations (e.g. Bullock, et al., 1982; Chi, 2000a; Driver, et al., 1985; Grotzer & Bell, 1999; Perkins & Grotzer, 2000; Spelke, et al., 1996; Wilensky & Resnick, 1999). These findings suggest that students have a tendency to simplify the otherwise more complex causal structure – a process that results in a distorted understanding.

Finally, the construct of self-explanation (e.g., Chi, et al., 1989; Chi, et al., 1994; Conati & VanLehn, 2000; DeLeeuw & Chi, 2003) has the potential to help people improve their causal reasoning skills, as research reports significant gains in performance

on task for people using self-explanation strategies compared to people that do not (e.g. Kastens & Liben, 2007). Studied in different contexts and for varied age groups, self-explanation seems to be a domain-independent strategy (Nathan, et al., 1994) that most high-performance students use (Bielaczyc, et al., 1995) in linking prior knowledge to current content. It has also been shown that either guiding (e.g. Bielaczyc, et al., 1995) or prompting (Chi, et al., 1994) people to self-explain improves performance.

What brings all this research together is its potential to inform a new approach to using causality in learning. It is possible that people's problems with causal reasoning that research unveiled could be diminished by asking learners to overtly explicate the causal mechanism(s) when describing or working on a causal relationship. Stated differently, by requiring learners to explain (self-explain) causal mechanisms, they will be forced to recall, use, evaluate and eventually reevaluate their mental models to fit the current situation. By requesting learners to overtly explain the causal mechanism(s) that link the causes and their effects, learners will be actively influenced to exert more cognitive effort, which has the potential to significantly improve their performance in present and future causal reasoning tasks.

CHAPTER 3: Methods

The main purpose of this study is to determine if the practice of self-explanation as causal mechanism elicitation technique affects learners' performance on causal reasoning tasks. Therefore, the methodology used in this study was developed to make possible the observation of the effects of using self-explanation based cognitive strategies while attempting to maintain all other factors constant. To achieve this goal, this study was conducted as an experiment with two groups: control and treatment.

Research Questions

The research questions this study attempts to answer are:

Q1: Does the practice of self-explanation as causal mechanism elicitation technique affects, on average, learners' performance on causal reasoning tasks?

Q2: How does prior knowledge of relevant domains relate to the learners' performance on causal reasoning tasks?

Q3: How does the use of self-explanation as causal mechanism elicitation technique influence participants' accounts of how and why they chose a certain answer or reach a certain solution?

Experiment Design

To address the requirements of this study a two-group, between subjects, completely randomized, experimental design was chosen. It is a between subjects design

because a subject is member of only one of the two experiment groups: control or treatment (Keppel, 1991, p. 19). It is also a completely randomized design since the participants were randomly assigned to one and only one of the two experimental groups (Keppel, 1991, p. 18). Both quantitative and qualitative data were collected. The quantitative data were collected to study the impact of the treatment, the use of self-explanation, on the participants' performance on a transfer causal reasoning task. The qualitative data were collected to better understand how participants reason causally in the given context.

The design of this experiment is based on existing literature and two pilot studies which were conducted with the intent to help the development of this research methodology. In the first pilot study, a pairing task for reconstructing an influence diagram from a scenario proved to be too complex and cognitively demanding[14] for the participants, as evidenced by many participants exiting the experiment early[15]. The second pilot study simplified the task by only asking the participants to explain the mechanisms that support the functioning of everyday things and phenomena[16]. While in this second case the task was simple enough for most of the participants to complete, the results were mixed. A qualitative analysis of the participants' answers suggests two conclusions: (1) the subject or topic was of no importance for the participants and (2) for everyday things, mechanisms or processes, people build very strong conceptual models

[14] The participants in the treatment group had to work through 4,060 combinations of causes, mechanisms, and effects, while the participants in the control group had to consider only 153 combinations of causes and effects.

[15] Of the 62 participants that answered the invitation, only 22 took the experiment to its conclusion while providing usable data points.

[16] The first scenario was about using a thermostat (house heating system) while the second scenario was about water evaporation (drying clothes).

they will not question or revise if the they do not have powerful enough incentives to do so, which the study did not provide. Also, it appeared that the participants did not even use these conceptual models, but relied only on learned outcomes, effects, or behaviors for everyday use purposes.

Building on the findings of these two pilot studies, the decision was made to ground this study in a domain in which causal reasoning is important for performance and in which incentives are high in order to motivate participants to reassess their mental models. This requirement lead to the selection of medicine as the general field, as physicians as well as medical students rely heavily on understanding causality and therefore on using causal reasoning. An added benefit of selecting this field is that causal mechanisms are present throughout the curriculum, making it possible to circumvent the need for training the participants in how to use and explain the mechanisms.

Focusing on the medical field, the next step was to select a topic around which to develop the elements of the research instrument. Because causal relationships in medicine can be very complex and involve many body systems, the specific topic chosen for this study had to be of average complexity, fairly well bounded, with clear, recognizable causal relationships. Consultations with a subject matter expert (SME) lead to choosing the more general area of immune responses, with focus a on asthma for the specific performance assessment. The decision to use the chosen topic was also determined by who the potential participants were. The targeted group consisted of second year medical students at a midwestern medical school.

The possibility of a small rate of participation among the 175 second-year medical students required the inclusion of first-year medical students as well, since students who,

at the time of the experiment (the end of the first year) should have had at least partial knowledge to allow them to approach the problem. Because differences in knowledge are expected between the first and second year medical students, any further quantitative analysis controlled for the influence of the year of study by including it as a covariate or control variable when appropriate.

Designing the specific items of an instrument for performance assessment as a transfer task required a decision about how the self-explanation treatment was to be delivered. Knowing that people can be trained in self-explanation (e.g. Wong & Lawson, 2002), a protocol adapted from Atkinson et. al. (2003), who used prompts to encourage learners to identify the underlying principle while solving practice problems, was developed for this research. To design the practice activities, a task familiar to the participants was chosen: multiple-choice questions. Associated with the multiple-choice questions received by all participants, the practice designed for the treatment group consisted of prompting them to answer an open-ended question about the mechanisms that are at work in the brief scenario presented to them in the problem definition. Because multiple-choice questions are familiar to medical students, as many of their examinations use multiple-choice questions based on causal scenarios, the task was expected to be simple enough that cognitive overload would not occur.

The treatment spans three multiple-choice questions similar to those which medical students encounter in their tests or board examinations. The participants in the control group were asked to answer the questions by selecting one of the choices presented to them as answers, as they would do during a test. The treatment designed for this experiment was to prompt the participants to practice self-explanation over the same

three multiple-choice questions. Therefore, the participants in the treatment group were asked to first explain what mechanisms or processes were at play before selecting an answer. One of the practice questions is presented in Table 1 as it is presented to the control and treatment groups

Table 1

Practice problems for control and treatment groups

Control group	Treatment group
A 10-year-old child presents with a persistent sinus infection. Over the course of the last 5 years, the patient has had several dozen sinus and upper respiratory infections. Blood tests reveal abnormally low levels of IgA immunoglobulin, but normal levels of other isotypes. The patient's recurring infections are most likely due to a defect in which of the following? a) Fixation of complement b) Mast cell activation c) Mucosal immunity d) Neutrophil activation e) Opsonization of bacteria	A 10-year-old child presents with a persistent sinus infection. Over the course of the last 5 years, the patient has had several dozen sinus and upper respiratory infections. Blood tests reveal abnormally low levels of IgA immunoglobulin, but normal levels of other isotypes. *What mechanisms are most probable to be at play in this case?* _____ _____ _____ The patient's recurring infections are most likely due to a defect in which of the following? a) Fixation of complement b) Mast cell activation c) Mucosal immunity d) Neutrophil activation e) Opsonization of bacteria

The final step in designing this experiment, besides developing the research instrument, presented below, was to decide upon what other elements the research should include. Since prior knowledge is considered to be important for both causal reasoning and self-explanation, this study includes a section designed to assess participants' prior

knowledge in the fields related to the performance assessment question. In addition, it was considered important to collect the participant intended medical specialty among other demographic data as this could be relevant to the subject's performance since s/he might have invested more effort toward that specialty.

Research Instrument

Within the overall experiment design, the specific instrument designed for this study and the rationale for the inclusion of each of the components is outlined in Table 2. A workflow of the instrument can be found in Appendix 2 while the full-length instrument is presented in Appendix 3.

Table 2

Structure of the research instrument

Assessment of prior knowledge	
Located at the beginning of the instrument so that it is not biased by any of the other instrument items. It is composed of two different measures, intended to provide an assessment in three areas: physiology, pathology, and immunology, all playing a role in answering the performance question in the transfer task.	
Self-rating	Self-rating of prior knowledge in three domains. Positioned at the very beginning so that it is not biased by how the participants perceive their performance on subsequent tasks.
Multiple-choice assessment	Set of six multiple choice questions designed to provide a brief assessment of prior knowledge of the domain. There are two questions per area, half having multiple answers (check all that apply). The questions were adapted from similar questions found in standardized tests medical students take for their board examinations. Examples of assessment questions are presented below this table.

Practice (treatment)	
A set of three multiple-choice questions. Both groups, control and treatment, were asked to answer these questions. The difference between the control and treatment group was that, together with answering each question, the participants in the treatment group were prompted to explain the mechanisms that they think are at work in the presented scenario. The questions were adapted from similar questions found in standardized tests. An example is presented in Table 1.	
Performance assessment (transfer task)	
Multiple-choice question	One multiple-choice question, developed together with the subject matter expert based on an influence diagram representing the main physiological and immunological causal relations in an asthma attack, also developed together with the subject matter expert. The question has only one correct answer among the presented choices.
Mechanism explanation	The participant was asked to explain the mechanisms that support his or her choice of the answer.
Self-confidence question	The participant was asked to rate his or her own confidence in the correctness of the answer chosen. This question is intended as reflection on the previous answer and prompt into the next step.
Change answer	Allowed the participants to change their answer to the performance assessment question. If the answer was NO, participants would go directly to the demographics section of this instrument.
Review answer to multiple-choice question	If the participant wished to make changes, he or she was shown the multiple choice question again, with the previously chosen answer highlighted and accompanied by the associated mechanism explanation presented earlier. The participant was also asked to explain why the new choice was better than the previous one.
Demographics	
The following demographics variables are collected: 1. Age group 2. Gender 3. Income group 4. Undergraduate major 5. Intended medical specialty	

Two prior-knowledge assessment questions for the same domain are presented below. For the all that apply question, the choices that should be included in the correct answer (selected) are marked with "(T)".

All that apply question: In the presence of inflammation the following are raised (check all that apply):

1. Platelets (T)
2. Ferritin (T)
3. Caeruloplasmin (T)
4. Fibrinogen (T)
5. Complement proteins (T)

Single choice question: An 18 month old infant male presents with sudden onset of cough. Suspecting a foreign body aspiration in which area of the lung does this happen more frequently?

1. Left upper pulmonary lobe
2. Right upper pulmonary lobe
3. Right lower pulmonary lobe (Correct answer)
4. Left lower pulmonary lobe

The performance assessment task was composed of three mandatory steps and a fourth one left at the participant's choice. The mandatory portion presented a brief scenario the participant and asked him or her to answer a multiple-choice question as shown below.

Scenario: "A 3-year-old male, known with seasonal allergies, presents at the clinic with difficulty breathing - increased respiratory rate, wheezing, and cough.

Symptoms started two days ago. No fever. Parents noted this morning a fine rash appearing on his abdomen. The rash seems to be getting worse."

Question: Which of the following is the most probable diagnostic?

1. Anaphylaxis
2. Asthma exacerbation (Correct answer)
3. Pneumonia
4. Foreign body aspiration
5. Roseola

After providing an answer to the multiple-choice question, the participants were asked to provide an account about the mechanisms which suggest the chosen diagnostic in an open-ended format. The question asked was: "What are the mechanisms that suggest your diagnosis?"

Once the participant answered this question, s/he was asked to provide an estimation of their own confidence in the correctness of their selected diagnostic and the provided explanation. The question asked was: "On a scale from 1 to 10, where 1 is 'Not confident at all' and 10 means 'I am sure', how confident are you that your answer is correct?"

The last step in the performance assessment sequence was to offer the participants the option to change their answer if they considered it necessary. If they chose to change their answer, they were presented with their previous answer to change and asked to provide the reason why the new answer is better than the previous one.

Sample, Sample Size, and Participant Recruitment

The invitation to participate was extended by e-mail to about 350 first- and second-year medical students pursuing a medical degree at a midwestern medical school. Considering a medium (.06) effect size and a power of .8 for an $\alpha = .05$, the minimum number of subject in each of the two experiment groups is 44 (Keppel, 1991, p. 72). Therefore, for a two-group experimental design, a minimum of 88 participants was needed.

While the experiment task was designed with second-year students in mind, first-year medical students were invited to participate as well. This decision was made because in addition to the expected low participation of the second-year medical students, the experiment was scheduled at the end of the academic year, when many extra activities compete for the students' time and attention. It was also determined by the fact that at the end of their first year, given the official schedule, first year students should have garnered enough knowledge of the appropriate domains to be able to approach the problem appropriately. To further improve participation, a lottery with four prizes of $200 each was run.

Data Collection

The data was collected online, using a web-based application developed for this study. The web-based instrument was designed to replicate the research instrument presented above and to store the participants' answers in a database. The URL of this web-application was http://rmed.gionas.webfactional.com. Selected screenshots of the experiment website are presented in Appendix 8.

The workflow of the web-application follows closely the workflow for the experiment (Appendix 2). When the participant opened the application URL, s/he was greeted and presented with the informed consent, and then asked to provide his or her name and e-mail address to acknowledge consent to participate. By continuing to the application, the participant was directed to the first step of the prior knowledge assessment section. In between, the application assigned the participant to one of the two experiment groups, alternatively. That is, if a participant was assigned to the control group, the next participant was assigned to the treatment group. The randomization was provided by the order in which the participants opened the web application, which was not under the researcher's control.

From this point on, the web application followed the workflow of the experiment until the last section, the demographics, was completed. At the end of the experiment, the last page presents the researcher's acknowledgment of the participant's effort and contribution and provides an opportunity for the participant to receive feedback to the answers provided to the 10 multiple choice questions that were part of the study[17]. If the participant was interested in receiving feedback, all 10 questions were presented to him or her, highlighting both the participant's answers and the correct choices.

Data Analysis

The collected data were analyzed both quantitatively and qualitatively. The answer to research questions one and two was based on a quantitative analysis, while the answer to research question three was based on a qualitative analysis.

[17] 6 questions in the prior knowledge assessment section, 3 treatment questions, and 1 performance assessment question.

Variables

Group (independent variable, nominal scale), is determined by the group to which the participant is assigned: control or treatment.

Prior knowledge score (independent variable, interval scale) has two variants: a) *measured prior knowledge* and b) *self-reported prior knowledge*. Measured prior knowledge attempts to provide an objective measure of the participants' knowledge of the three main domains included in the scenario. Self-reported prior knowledge attempts to capture participants' opinion of their knowledge of the three domains. This second measure of prior knowledge was introduced as an attempt to compensate for the small number of questions used to measure prior knowledge, due to the constraints of this study. The researchers' opinions on the validity of self-reporting in assessing knowledge are divided. Nevertheless, as part of their academic activities, by the end of the school year, the students participating in this research worked through most of the content appropriate for their year of study, organized in multiple one-month long modules with assessments at the end of each module. Therefore, it was expected for the self-reported prior knowledge to include not only the opinion about self, but also others' assessment based on the various forms of assessment associated with each of the academic modules they completed or are in process of completing. This decision is supported by existing research studies and meta analyses which report, for example, that no consistent over- or underestimation was found in self-assessment (Boud & Falchikov, 1989) or that broad agreement between self-assessed and objective knowledge can be observed (Ackerman, Beier, & Bowen, 2002). In addition, medical students have been found to be relatively accurate in predicting test scores for example, which suggests well-developed self-

assessment skills (Fitzgerald, Gruppen, White, & Davis, 1997) and that self-assessment accuracy is relatively stable for the first two years of medical school (Fitzgerald, White, & Gruppen, 2003).

Measured prior knowledge is calculated as the sum of the answers to the first six multiple-choice questions in the research instrument. Because two types of multiple-choice questions, single choice and all that apply, were used in the measurement of prior knowledge, the scoring method tried to account for their relative difficulty. Therefore, a first rule used in scoring was to assign each question a number of points equal to the number of possible answers. The multiple-choice questions with a single answer choice were scored on the "all or nothing" principle. That is, for a correct answer the score was equal to the number of answers available for that question, which accounts for its relative difficulty measured in terms of number of answer choices, while for an incorrect answer the score was zero.

For the all that apply questions, a correct choice was represented either by a checked answer if that answer was supposed to be checked, or by an unchecked answer if that answer was supposed to be unchecked. Therefore, for these questions scores were calculated as a sum of individual answer scores as follows. If the participant correctly assessed the answer choice by either checking it or leaving it unchecked, as appropriate for each situation, that answer choice received one point. If the participant incorrectly assessed the answer by either checking it or leaving it unchecked when the reverse would have been true, zero points were awarded. The minimum score for each question was zero while the maximum score is equal to the number of answer options in each question.

Self-reported prior knowledge was computed as the sum of three domain-specific self-rating scores, provided on a scale from one to 10, at the very beginning of the experiment, resulting in a final range of values from three to 30.

Performance score (transfer task, dependent variable). The performance score was determined in three ways, resulting in three different performance scores for the same task: a) a categorical performance score (nominal scale), b) a total performance score (interval scale) and c) an adjusted performance score (interval scale). The *categorical performance score* reflects only the answer (choice) the participant provides to the multiple-choice performance question, unbiased by other variables in the experiment. The *total performance score* accounts for the answer to the multiple-choice question and for the quality of the associated explanation. The *adjusted performance score* includes the participant's confidence in his or her own answer.

The *categorical performance score* was designed to assess participants' performance on the multiple choice question by itself, not biased or influenced by any other experiment variable. As the experiment was designed in such a way that the participant would not be able to go back and change the answer to the multiple-choice performance question after s/he moved on, this choice represents the unbiased outcome of the performance assessment task, which follows immediately after the practice session. It reflects the score of zero for an incorrect choice or one for the correct choice.

The *total performance score* was calculated as the sum between the score for the multiple-choice question and a score ranging from zero to one resulted from applying a scoring rubric to assess the explanation associated with the answer to that multiple-choice question. The scoring rubric was developed based on both the causal diagram and an

interview with a subject matter expert (SME). During this interview the SME was asked to evaluate in an open-ended format the participants' answers while a think-aloud protocol was used to capture the expert's thoughts and observations. The researcher intervened only to keep the interview on track, at crossroads and to help clarify certain words, expressions, or situations. The think-aloud protocol for this session is presented in Appendix 5.

Furthermore, an *adjusted performance score* was calculated to account for participants' confidence in the correctness of their answer. For this, the confidence value, represented on a scale from one to 10, was converted to its percent value and applied differentially, depending on the correctness of the answer to the multiple choice question. This scoring variant reflects the assumption that the more confident the participant is in a correct answer the higher the score should be and the more confident the participant is in a wrong answer, the lower the score should be. That is, if the participant answered the multiple choice question correctly, the score would be equal to the percent value of the confidence score, whereas if the participant answered the multiple choice question incorrectly, the score would be equal to the negative value of the percent value of the confidence score.

Year of study (covariate). The inclusion of first-year medical students in this experiment together with the second-year medical students on a task that was designed for the average second-year student suggests that second-year students will probably perform better than first-year students. Therefore, to better observe the effects of the applied treatment, the year of study was included as a covariate to control for its impact on the performance measures.

Analysis

To answer research questions one and two, two statistical tests are used, depending on the type of dependent variable in the analysis. For the quasi-continuous variants of the performance score, the total and the adjusted performance scores, the analysis was conducted using a one-way, between groups ANCOVA statistical test, with the performance score as dependent variable, experiment group as independent variable, and year of study as covariate.

Prior knowledge was introduced as an independent variable while searching for an answer to the second research question to account for its role as a possible moderator. In this case, since prior knowledge is expected to be dependent on the year of study, the year of study was introduced as a covariate in the analysis. As a result, to answer the second research question, an ANCOVA statistical test was used, with year of study as a covariate. For the categorical performance score, the analysis was conducted using a binomial logistic regression statistical test (Pedhazur, 1997) involving the same dependent and independent variables. This analysis is designed to provide information about both variables individually, and about any interaction that might occur between them. This analysis tests if prior knowledge is a moderator (Baron & Kenny, 1986) of the relationship between the treatment and the participant's performance. For this analysis, the prior knowledge variable is dichotomized to "high" and "low" using the sample mean as cutoff point to group the data. Since this is a regression type analysis, all included variables will be considered as part of the equation. Therefore the year of study was be included in this analysis alongside prior knowledge variable.

The answer to research question three was approached through a qualitative analysis. From a qualitative perspective, the analysis looked at the answers provided by 12 participants, three at each of the two ends of the performance measure, for each experiment group. That is, the analysis included the three participants with the lowest performance scores and the three participants with the highest performance scores, for each experiment group. The analysis looked at how well the explanation matched the expert's influence diagram and how well the concepts, relationships and mechanisms were defined.

From a different perspective, the qualitative analysis also looked at the answers provided by the participants that chose to change their answers to the performance question, and at the rationale they provide for this change. The answers provided by subjects in both experiment groups were considered.

Summary

To study the effects of using practice of self-explanation as methodology to help people's causal reasoning processes, a two-group between subjects, completely randomized experiment was designed, informed by the research literature and two pilot studies. This experiment was conducted with medical students, as the medical field relies heavily on causal reasoning for achieving performance.

The research instrument was designed to provide the means to allow the participants in the treatment group to practice self-explanation strategies while controlling for other factors. Delivered online, using a web-based application designed for this purpose, the research instrument allowed for collection of data for both qualitative and quantitative analyses. The quantitative analyses showed whether an effect

of using self-explanation existed, while the qualitative analysis furthered the understanding of how it helped causal reasoning.

CHAPTER 4: Results

The data were collected online using an application designed and programmed by the researcher. The experiment website was open for four weeks toward the end of the academic year, during the month of April, 2009. During this time, three e-mail messages were sent to about 350 first- and second-year medical students. The first message sent was the initial invitation to participate, while the other two were reminders. Answers from 147 participants were collected. After removing duplicate and incomplete answers, there were 117 valid data points, of which 44 were first-year medical students and 73 were second year medical students. Duplicate records were removed by deleting the newest record. When the subject quit the experiment early and attempted to continue later the two records were merged whenever possible[18] by keeping the older of the two sets of values. A record was considered incomplete when one or more of the following data was missing: one or more scores from the self-reported prior knowledge set, one or more of the answers to the prior-knowledge measurement question set, and/or the complete answer to the performance assessment question.

Variables

A summary of the variables included in the analysis is presented in Table 3.

[18] Same experiment group and no missing data between the two records.

Table 3

Variables

Name	Description	Range/Values
	Dependent variable(s) (measure performance on the same task using three alternative scores)	
Categorical performance score	Performance assessment multiple-choice question answer, disregarding other variables or explanations. Represents the unbiased initial answer.	Values are 0 or 1
Total performance score	A score calculated to account for the correctness of the answer choice and the quality of the explanation.	Range from 0 to 2
Adjusted performance score	The categorical performance score adjusted by the self-reported confidence in the correctness of chosen diagnostic. Aimed at representing the strength of the participant's constructed mental model.	Range from 0 to 2
	Independent variables	
Experiment group	The experiment condition to which the participant was assigned.	Either "control" or "treatment"
Self-reported prior knowledge	The sum of self-reported scores for three different knowledge domains. Converted to a categorical variable using the sample mean as cutoff point.	Values are "Low" or "High"
Measured prior knowledge	A value calculated based on the answers the respondents provided to the six prior knowledge questions. Converted to a categorical variable using the sample mean as cutoff point.	Values are "Low" or "High"

Dependent Variable(s)

Categorical Performance Score

The *categorical performance score* was calculated by assigning a value of zero to a wrong answer choice or a value of one to the correct answer choice. Because this is a categorical variable, the analysis was performed using a logistic regression statistical test (Pedhazur, 1997), specifically designed for the study of categorical variables through regression analysis.

Table 4

Frequency table for the categorical performance score

	Correct answer	Incorrect answer	Total
First year	12	32	44
Second year	37	36	73
Total	49	68	117

Total Performance Score

To compute the *total performance score*, the first step was to construct the scoring rubric for participants' explanations. The scoring rubric was developed based on the influence diagram developed for the domain (Appendix 1) and an interview with a subject matter expert (SME). During this interview the SME was asked to use a think aloud protocol while analyzing participants' answers. The interview was audio recorded and transcribed. The transcription was reviewed and analyzed to understand and extract the elements the SEM was looking for when analyzing the participants' explanations.

The analysis suggests that the subject matter expert looked at three main themes or elements: 1) presence of the correct symptoms and/or of related keywords indicative of these symptoms, 2) the level of detail of the answer and 3) the presence of mechanisms expressed as one or more correct sequences of events. For each criterion, using the influence diagram and the interview data, a set of scores was developed. Once the scoring rubric was developed, it was presented for evaluation to the same subject matter expert who considered it appropriate and proposed no changes. Each of these criteria and the associated scores is explained in Appendix 5 along with specific examples. Using the rubric, two raters, the researcher and a physician (other than the SME), scored the participants' answers. The two raters discussed the rubric and decided to proceed by reading and analyzing each answer separately, scoring it separately, and negotiating a final score. In the end, the total performance score was computed by summing the score for the answer choice (zero or one) with the score for the explanation converted to a percentage[19]. The summary statistics for this variable are presented in Table 2.

Table 5

Summary statistics for the total performance score

Statistic	Value
Min	-2.00
Max	7.00
Median	1.00
Mean	1.584
SD	2.094

[19] The conversion to percentage was performed to represent the score for the explanation as a value between zero and one to allow for equal weight for the answer and the explanation in the final score.

Adjusted Performance Score

The third variant of the performance score, the *adjusted performance score*, was determined based on the assumption that the confidence the participant had in his or her own answer has different meaning for correct and wrong answer choices. That is, one can infer that the more confident a subject is in a correct answer, the closer the constructed mental model of the scenario is to the expert's model. Conversely, the more confident a subject is in an incorrect answer, the farther away the constructed model can be considered to be from the expert's model. Therefore, for a correct answer choice, the score is one plus the value of the self-confidence while for the wrong answer choices, the score is the one minus the value of the reported confidence.

$$Adjusted\ \ performance\ \ score = \begin{cases} 1 + confidence\ \ (\%) \\ 1 - confidence\ \ (\%) \end{cases}$$

Table 6

Summary statistics for the adjusted performance score

Statistic	Value
Min	0.22
Max	1.78
Median	1.00
Mean	0.97
SD	0.3839

Independent Variables

Experiment Group

Table 7 shows the distribution of the participants by experiment group and year of study.

Table 7

Participants distribution by experiment group and year of study

	Control	Treatment	Total
First year	25	19	44
Second year	36	37	73
Total	61	56	117

Prior Knowledge

Prior knowledge was measured in two ways. The multi-domain *self-reported prior knowledge* variable was computed as the sum of three self-reported single domain measures of prior knowledge: immunology, physiology and pathology. The scale variable was transformed into a two-level categorical variable (low/high) by dividing the sample in two groups against its mean (14.1966). After transformation, there were 51 participants in the high prior knowledge category and 66 participants in the low prior knowledge category. Table 8 shows the participants distribution by level of self-reported prior knowledge and year of study.

Table 8

Participants distribution by level of self-reported prior knowledge and year of study

	Low prior knowledge	High prior knowledge	Total
First year	39	5	44
Second year	27	46	73
Total	66	51	117

The *measured prior knowledge* variable was computed as the sum of the scores of six multiple-choice questions. The process for calculating the measured prior knowledge score is presented in detail in the methodology section. For this analysis, the obtained scale variable was converted to a two-level categorical variable (low prior knowledge and high prior knowledge) using its mean (13.6581) as cutoff point. After conversion, there were 65 subjects in the low measured prior knowledge category and 52 subjects in the high measured prior knowledge category. Table 9 shows participant distribution by level of measured prior knowledge and year of study.

Table 9

Participants distribution by level of measured prior knowledge (PK) and year of study

	Low PK	High PK	Total
First year	26	18	44
Second year	39	34	73
Total	65	52	117

Analysis of normality

Kurtosis and skewedness calculations are used to test the assumption of normality of the dependent variable as required by the analysis of covariance statistic test. For the continuous variables in this study, the values of the two parameters and the diagnostic decision are presented in Table 10.

The z-scores for measured prior knowledge suggest the existence of one outlier. After further analysis of the corresponding data point, the decision was made to keep it in the analysis. The decision is also supported by the values of skewedness and kurtosis that are within the recommended limits, suggesting that the assumption of normality is verified.

Table 10

Summary of normality and outlier analysis

Variable	Skewedness	Kurtosis	z-scores	Diagnostic decision
Total performance score	0.576	-0.649	Min: -1.71 Max: 2.59	Normally distributed
Adjusted performance score	0.115	-0.624	Min: -1.95 Max: 2.11	Normally distributed
Self-reported prior knowledge	0.134	-0.124	Min: -2.16 Max: 2.31	Normally distributed
Measured prior knowledge	0.435	0.781	Min: -2.80 Max: 3.75	Normally distributed

Note: *Boundary values for skewedness, kurtosis, and z-scores are ±3.*

Research Results

Research Question 1

Does the practice of self-explanation as causal mechanism elicitation technique affects, on average, learners' performance on causal reasoning tasks?

Table 11 summarizes the types of statistical analyses performed to answer this question. Given that both first and second year medical students were invited to participate in the experiment, the year was introduced in the analysis as a covariate variable to control for its impact.

Table 11

Research Question 1: Types of statistical analyses

Independent variable	Dependent variable		
	Categorical Performance Score	Total Performance Score	Adjusted Performance Score
Experiment group	Logistic Regression	One-way ANCOVA	One-way ANCOVA

For the *categorical performance score*, a logistic regression statistic test was used. The results show no significant interaction effect and no significant main effect of the experiment group. However, the results do show a significant main effect of the year of study. The results are presented in Table 13.

Table 13

Logistic regression results for experiment group x year of study

Predictor	β	SE β	Wald χ^2	df	p	e^{β}
Constant	0.328	0.305	1.154	1	0.283	1.388
Experiment group (1)	-0.608	0.389	2.443	1	0.118	0.544
Year of study (1)	-0.948	0.416	5.599	1	0.018*	0.347
Total			χ^2	df	p	
Overall model evaluation						
Likelihood ratio			8.814	2	0.012*	
Goodness of fit						
Hosmer & Lemeshov test			1.540	2	0.463	

Note: The analysis was conducted using the Binary Logistic Regression in SPSS, using the enter method. The significance level was set at 0.05. Cox and Snell $R^2 = 0.073$, Nagelkerke $R^2 = 0.098$.

The results suggest that the experiment group does not influence the odds for choosing the correct answer versus choosing an incorrect answer on the performance assessment multiple-choice question. From a different perspective, being a first-year medical student decreases the odds of answering correctly versus answering incorrectly by a factor of 0.347, when compared to second-year medical students.

For the *total performance score*, the results of the ANCVOA analysis are presented in Table 14. The data show that the treatment group had no effect on performance. The covariate (year of study) was significantly related to the total performance score.

Table 14

ANCOVA table - Experiment group × year of study for the total performance score

Source	df	F	η	p
Year of study	1	5.798	0.055	0.018*
Experiment group	1	1.477	0.014	0.227
Error	98	(45.355)		

Note. Values enclosed in parentheses represent mean square errors. * $p < 0.05$

Performing the same statistical test for the *adjusted performance score* suggests similar outcomes. That is, the ANCOVA analysis shows no significant main effect of the experiment condition ($F(1,114) = 1.771$; $p = 0.186$) on the total performance score. The covariate (year of study) was significantly related to the total performance score ($F(1,114) = 4.37$; $p = 0.039$ (< 0.05)).

Research Question 2

How does prior knowledge of relevant domains relate to the learners' performance on causal reasoning tasks?

For each type of performance score included in this analysis, Table 15 shows the statistical test run for each DV (Dependent Variable) – IV (Independent Variable) combination. The treatment condition and the covariate were always included in the analysis.

Table 15

Research Question 2: Types of statistical analyses

		Dependent variable		
		Categorical performance score	Total performance score	Adjusted performance score
Independent variable	Self-reported prior knowledge	Logistic regression	Factorial ANCOVA	Factorial ANCOVA
	Measured prior knowledge	Logistic regression	Factorial ANCOVA	Factorial ANCOVA

Categorical Performance Score

The categorical performance score associates the value of one (true) or zero (false) with a correct or incorrect answer to the performance assessment question respectively. Binary logistic regression is the statistical test which was run to analyze the categorical performance score. The same analysis was performed for each of the two measures of prior knowledge, measured and self-reported by the participants, as the independent variable.

Measured prior knowledge. The results of the logistic regression test (Table 16) show a significant interaction between measured prior knowledge and the experiment condition. That is, the odds of answering the multiple-choice performance assessment question correctly are different for different levels of prior knowledge for the participants in different treatment conditions. Further analysis (Appendix 6) shows that for the high prior knowledge group, the treatment condition does not influence these odds. For the participants in the low measured prior knowledge group, the odds of answering the

question correctly versus answering it incorrectly are *decreased* by a factor of 0.387 by being in the control group as opposed to being in the treatment group. The coding of the dependent variables is presented in

Table 17.

Table 16

Logistic regression results for the analysis with the measured prior knowledge

Predictor	β	SE β	Wald χ^2	df	p	e^{β}
Constant	-1.718	0.734	5.482	1	0.019*	0.179
Experiment group (1) by Measured prior knowledge (1)	-0.950	0.443	4.590	1	0.032*	0.387
Year of study	1.014	0.420	5.829	1	0.016*	0.368
Total			χ^2	df	P	
Overall model evaluation						
Likelihood ratio [b]			11.217	2	0.004**	
Goodness of fit						
Hosmer & Lemeshow test [c]			2.362	2	0.307	

Note: [a] The analysis was conducted using Binary Logistic Regression in SPSS using the backward elimination process. The elimination process continued until nothing could be eliminated from the model anymore. The significance level was set at 0.05. Cox and Snell $R^2 = 0.091$, Nagelkerke $R^2 = 0.123$. The elimination process was in three steps. The model significance was 0.0.023 in the first step, 0.010 for the second step, and 0.004 for the third step. [b] $p < 0.05$ means there is adequate fit of the data to the model, meaning that at least one predictor is significantly related to the response variable. [c] A finding of non-significance indicates that the model adequately fits the data.

Table 17

Categorical variables parameter coding for logistic regression with measured prior knowledge

		Frequency	Parameter coding (1)
Measured prior knowledge	Low	65	1.000
	High	52	0.000
Group	Control	61	1.000
	Treatment	56	0.000

Self-reported prior knowledge. The results of the binary logistic regression analysis are presented in

Table 20. The coding of the categorical factors is shown in Table 17. The logistic regression test shows a significant interaction effect between the treatment condition and self-reported prior knowledge.

Table 17

Categorical variables parameter coding for logistic regression with self-reported prior knowledge

		Frequency	Parameter coding (1)
Self-reported prior knowledge	Low	66	.000
	High	51	1.000
Group	Control	61	.000
	Treatment	56	1.000

Table 20

Logistic regression results for the analysis with self-reported prior knowledge

Predictor	β	SE β	Wald χ^2	df	p	e^{β}
Constant	0.682	0.441	2.387	1	0.122	1.978
Experiment group (1) by Self-reported PK[a] (1)	1.583	0.805	3.864	1	0.049*	4.868
Year of study (1)	-1.140	0.487	5.477	1	0.019	0.320
Total			χ^2	df	P	
Overall model evaluation						
Likelihood ratio [c]			13.101	3	0.004*	
Goodness of fit						
Hosmer & Lemeshow test [d]			1.530	4	0.821	

Note: [a] PK represents "prior knowledge". [b] The analysis was conducted using Binary Logistic Regression in SPSS using the backward elimination process. The elimination process continued until nothing could be eliminated from the model anymore. The significance level was set at 0.05. Cox and Snell $R^2 = 0.106$, Nagelkerke $R^2 = 0.143$. The elimination process was in two steps. The model significance was 0.011 in the first step and 0.004 for the second step. [c] $p < 0.05$ means there is adequate fit of the data to the model, meaning that at least one predictor is significantly related to the response variable. [d] A finding of non-significance indicates that the model adequately fits the data.

Further analysis (Appendix 7) shows that, for the participants with low self-reported prior knowledge, the treatment condition does not influence de odds of answering correctly versus answering incorrectly. For the participants who report high levels of prior knowledge, the odds of answering the question correctly *are increased* by a factor of 4.385 for the participants in the treatment group compared to the participants in the control group.

Total Performance Score

The total performance score was calculated by combining the participant's score for the performance question and the score for the provided explanation. For both measures of prior knowledge, the analysis was performed using a factorial ANCOVA statistical test. The analysis began with measured prior knowledge as the independent variable, which was replaced for further analysis by self-reported prior knowledge.

Measured prior knowledge. The two-way between–groups ANCOVA with two between-groups factors included the year of study as a covariate to control for its effects. This analysis shows no significance for either the interaction between experiment group or the measured prior knowledge ($F(1,96) = 0.213$; $p = 0.645$) and the simple effects of experiment group ($F(1,96) = 0.985$; $p = 0.324$) and measured prior knowledge ($F(1,96) = 1.419$; $p = 0.236$). The covariate, year of study, was significantly related to the total performance score ($F(1,96) = 5.426$; $p = 0.022$ (< 0.05)).

Self-reported prior knowledge. The analysis shows a *significant interaction* between the experiment group and the self-reported level of prior knowledge ($F(1,96) = 4.799$; $p = 0.031$ (<0.05)) after controlling for the year of study. The nature of this interaction is shown in *Figure 1*. Further analysis shows that both simple effects, experiment group ($F(1,96) = 1.851$; $p = 0.177$) and self-reported prior knowledge ($F(1,96) = 0.997$; $p = 0.32$) are not significant. The covariate, year of study, was significantly related to the total performance score ($F(1,96) = 7.801$; $p = 0.006$ (< 0.05)).

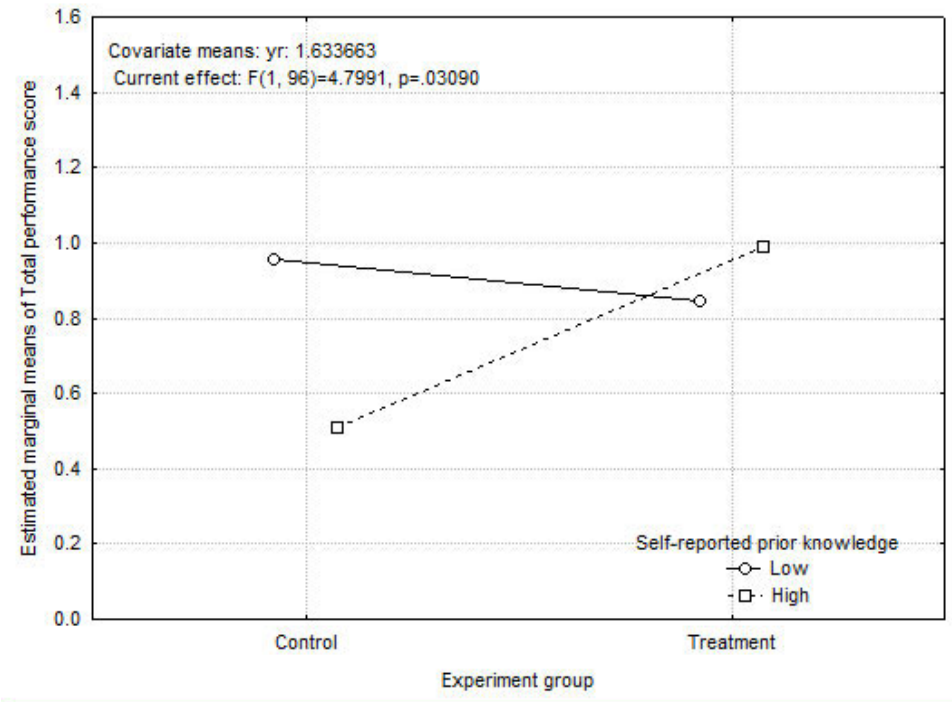


Figure 1. Estimated marginal means for total performance score

Post-hoc analysis[20] suggests that the difference is more pronounced between the subjects reporting high levels of prior knowledge in the control and treatment groups, followed by the difference between the participants in the treatment group who reported low versus high prior knowledge levels. The representation in *Figure 1* suggests that practicing self-explanation considerably affects the performance of the subjects in the high self-reported prior knowledge category but has little effect on the subjects reporting low levels of prior knowledge. That is, the effect of the applied treatment seems to be dependent on the level of prior knowledge reported by the participants.

[20] The analysis was conducted using Fisher LSD and Tukey HSD tests.

Adjusted Performance Score

M*easured prior knowledge*. A two-way ANCOVA analysis with two between-subject factors was performed. The results (Table 19) show no significant interaction between the experiment group and measured prior knowledge. Further analysis shows no significant main effect of the experiment group, but indicates a significant main effect of the measured prior knowledge factor. That is, the subjects with higher levels of measured prior knowledge show a tendency to perform better than the subjects with lower levels of prior knowledge. The covariate (year of study) shows borderline correlation with the adjusted performance score.

Table 19

ANCOVA table - Experiment group x Measured prior knowledge, with year-of-study

Source	*df*	*F*	η	*p*
Year of study	1	3.877	0.537	0.051
Experiment group (G)	1	1.428	0.198	0.235
Measured prior knowledge (MPK)	1	4.261	0.590	0.041*
G x MPK	1	0.475	0.066	0.492
Error	112	(15.512)		

Note. Values enclosed in parentheses represent mean square errors. * $p < 0.05$

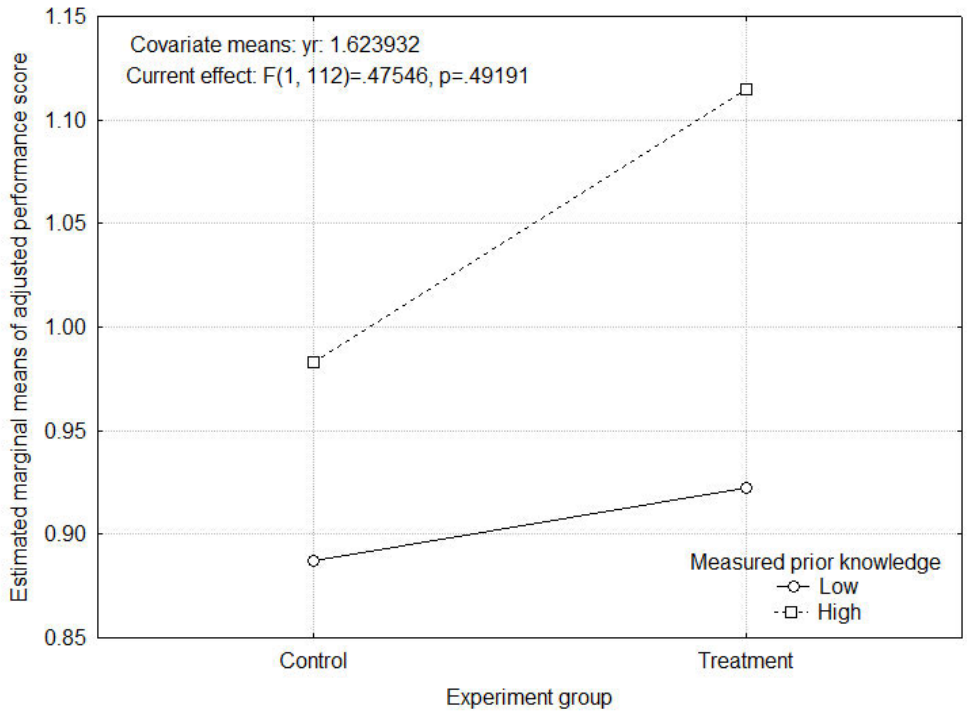


Figure 2. Estimated marginal means of adjusted performance score for measured prior knowledge

Self-reported prior knowledge. The data were analyzed using two-way ANCOVA with two between-groups factors and year of study as covariate. The results are presented in Table 20. This analysis shows a significant interaction between the experiment group (treatment condition) and self-reported prior knowledge, while showing no significant main effects. The nature of the interaction is shown in *Figure 3*. The covariate (year of study) was significantly related to the adjusted performance score.

Table 20

ANCOVA table - Experiment group x self-reported prior knowledge for the adjusted performance score

Source	df	F	η	p
Year of study	1	6.219	0.848	0.014*
Experiment group (G)	1	2.247	0.306	0.137
Self-reported prior knowledge (SRPK)	1	1.451	0.198	0.231
G x SRPK	1	4.781	0.652	0.031*
Error	112	(15.271)		

Note. Values enclosed in parentheses represent mean square errors. * $p < 0.05$

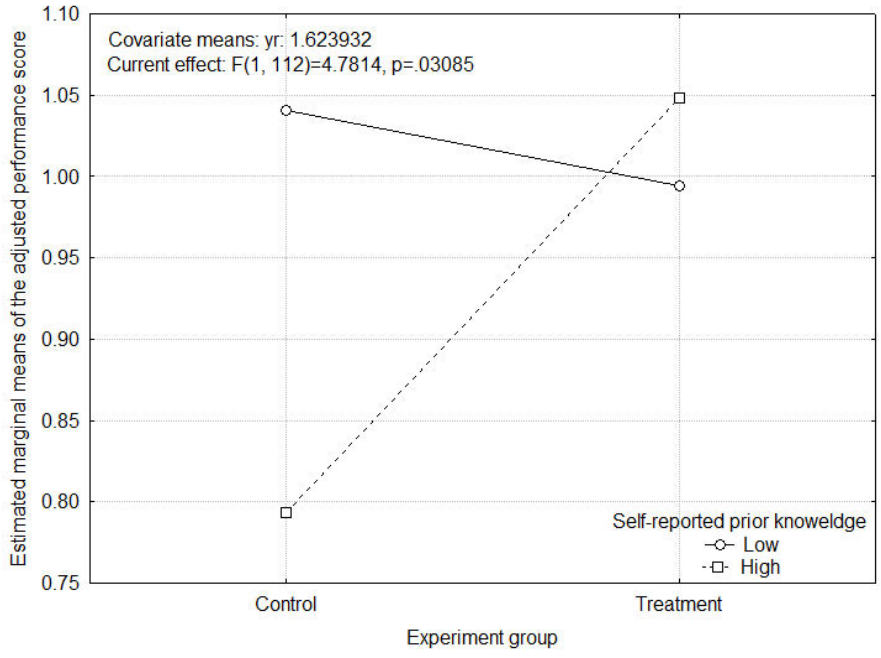


Figure 3. Estimated marginal means of the adjusted performance score, controlling for year of study, for self-reported prior knowledge

Further analysis suggests that the treatment condition has different effects for different levels of prior knowledge. That is, practice of self-explanations seems to have

significant effects on performance for the participants reporting higher levels of prior knowledge and almost no effects on performance for the participants reporting lower levels.

Qualitative Analysis

Research Question 3

How does the use of self-explanation as causal mechanism elicitation technique influence participants' accounts of how and why they chose a certain answer or reach a certain solution?

To answer this last research question an in-depth qualitative analysis was initially planned. Due to this experiment's limitations, the data collected did not present the expected depth and breath. These limitations include (1) the delivery at the end of the school year, which constrained the amount of time and effort the participants were able and willing to give for the experiment, (2) the fact that medical school training targets performance on national examinations, thus favoring quick answers to multiple choice questions over explanations, (3) oversimplification of the communication with those outside the field due to complexity and difficulty of understanding of medical explanations and (4) communication with peers based on keywords and vocabulary that is expected to express, many times in very few words, an entire network of causal relations. Therefore, less than optimal data makes this analysis more difficult and limits its depth. Since the qualitative data collected for this experiment were not extensive, no special tools or software applications were used in the analysis. A subject matter expert was consulted during the qualitative analysis, whenever needed, about the various aspects of the answers the participants provided.

The analysis looked at two aspects of the participants' explanations. First, it looked at the explanations the participants provided to support their answer to the performance assessment question, for both control and treatment groups. The explanations provided by the participants with the lowest performance scores were compared with those provided by the participants with the highest performance scores. Second, the analysis looked at how and why the participants decided to change their answer when provided with the opportunity to do so.

Mechanism Explanations

Two scores, the total performance score and the adjusted performance score were used to select the three best and three worst performers. When more than three participants shared a score at either end of the spectrum, random selection was used to reduce the group to three subjects. In analyzing the explanations, several themes were observed across cases.

A first theme is based on the observation that the answers included in the low score category, for both variants of the performance score, in both experiment groups, show that of the total of 12 cases analyzed, 11 point to "Roseola" as the answer of choice. Statement such as "presence of a rash" or "Roseola is a three-day rash with similar symptoms" use either the answer choice directly or use the "rash" as the symptom which was determinant in their decision. "Roseola" was one of the answer choices presented to the participants and was included because one of the symptoms presented in the scenario, rash, is one of the symptoms that could suggest this diagnosis. Nevertheless, the other symptoms that would make the diagnosis of "Roseola" certain are missing from the scenario, suggesting that rash is a symptom related to a different condition.

On the other hand, there seems to be a tendency of the subjects at the higher end of the performance score to attempt to explain the rash in the context of the correct answer, even if it is not necessarily part of it. For example, subject 137 explains:

> Subject 137:
>
> "The allergens have triggered histamine release which causes bronchiolar constriction and vasodilation of the vessels of the skin, producing a rash."

Or subject 2, who states "… histamine could cause his respiratory [symptoms] as well as the fine rash that had appeared on the abdomen."

Most answers provided by the participants in the lowest performance score subsample are brief compared with the answers of those participants in the highest performance score. Statements composed of only a few words, similar to "The presence of a rash" or "symptoms followed by a rash a few days later" seem to be common.

At the higher end of the performance score, explanations include references to causal mechanisms, mostly as sequences of events, interpreted by the SME as representations of causality. In this context, comparing the answers with the causal map the expert developed was difficult, as the researcher and the SME needed to infer the participants' thought process. For example, looking at the explanation provided by subject 73, "allergy, Th2, humoral response" and comparing it to the influence diagram, one can observe that, for that subject, the entire map was reduced to three components, each including multiple causal mechanisms. In addition, the network of causal relations represented in the map was reduced to a simple linear causal relation.

In effect, in some answers, causal relations are linear, with at most three or four nodes. For example, the explanation provided by subject 73 presented above contains three nodes, as shown in *Figure 4*.

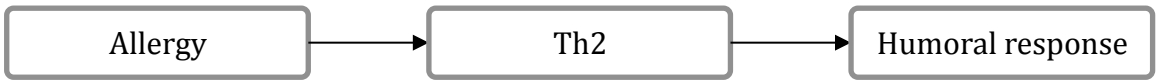


Figure 4. Influence diagram presentation of the explanation provided by subject 73

In other cases, while the answer starts in a linear fashion, the participant's attempt to provide a more detailed explanation creates the need for two paths at the end of the causal chain. For example, the explanation provided by subject 137 can be converted to a causal diagram as follows (*Figure 5*).

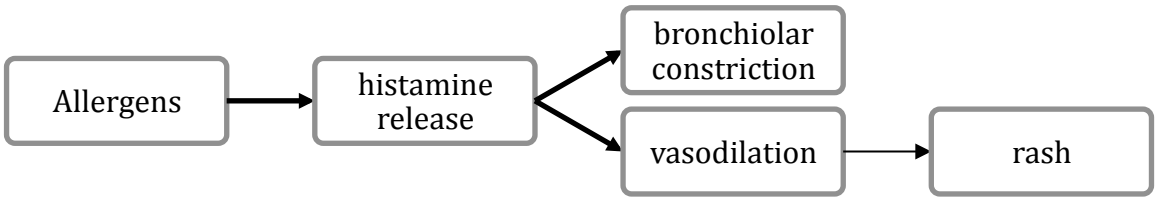


Figure 5. Influence diagram presentation of the explanation provided by subject 137

Subject 25 presents a different explanation.

> Subject 25:
>
> "Histamine receptor activation leading to mast cell proliferation leading to wheezing (bronchial constriction), cough, increased respiratory rate."

This explanation would be represented graphically as follows (*Figure 6*):

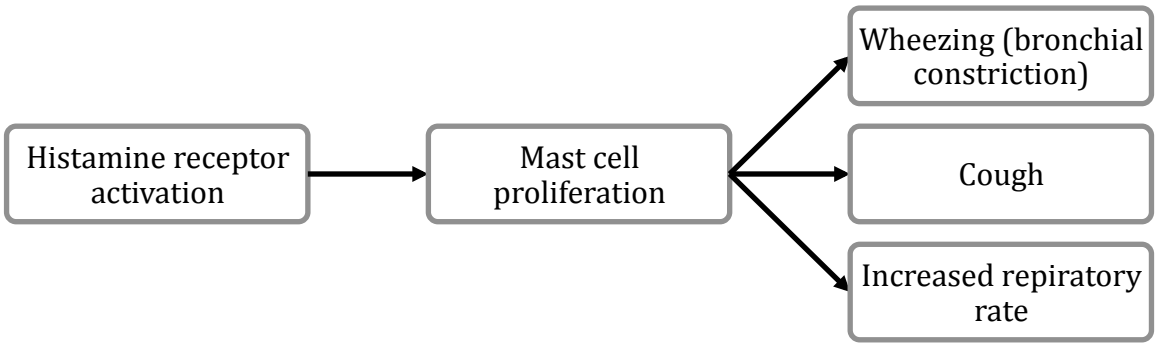


Figure 6. Influence diagram presentation of the explanation provided by subject 25.

Compared to the influence diagram for asthma presented in Appendix 1, the diagrams developed based on the participants' explanations are simple, composed of a small number of nodes (two to six nodes on average), with each node representing a high-level concept or process which, in the expert's diagram, is composed of multiple, interrelated concepts and causal relations. In addition, the diagrams, while correct, are relatively linear and lack long branches or feedback loops.

Another common theme that emerged in multiple instances is the presence of some form of a differential diagnostic between some or all of the five diagnostics offered as possible answers for the performance assessment question. For example, subjects 10 and 155 provide the following explanations:

> Subject 10
>
> "No fever so that means it was likely not pneumonia. Anaphylaxis is more severe and would require attention sooner than 2 days. Wheezing and cough happen to try to disperse the foreign body from the respiratory tract, and the respiratory rate increased to make up for the lack of air passing through."

Subject 155

"Well, I actually tried to rule out the other options. There was no fever so I ruled out roseola. The Presentation did not seem to be acute like anaphylaxis. Beyond that I wasn't really sure. It seemed to me that the remaining options could cause the respiratory symptoms, but I'm not sure that Pneumonia or a foreign body aspiration would cause a rash. Asthma is IgE mediated and could therefore present with a skin irritation similar to anaphylaxis."

In both cases, the participant goes through some or all the diagnostics provided and attempts to disprove each of them using the symptoms presented in the scenario. In the case of subject 155, the participant clearly presents the elimination thought process "… tried to rule out other options" after which the participant proceeds to list the options and associate them with the existing symptoms. While the participants were asked to explain the mechanisms they considered when they answered the question, they explained a much more extensive causal reasoning process which contains the sought explanation as a component. That is, understanding of the causal processes behind their answer is only part of the differential diagnostic process they needed to perform to decide upon it.

Control vs. Treatment Group

The explanations provided by the participants in the low performance score category do not seem to indicate notable differences between the control and the treatment groups. For example, subject 1 in the control group supports the diagnostic for Roseola with "The presence of a rash", similar to the explanation provided by subject 4 in

the treatment group: “presence of rash around his abdomen”. The same similarity can be found between subject 17 in the control group who explains “Roseola is a 3 day rash with similar symptoms” and subject 23 in the treatment group who explains “symptoms followed by rash a few days later” or subject 95 in the same group who explains “Onset of rash a few days after the other symptoms.”

Looking at the high performance score group, the subjects in the treatment group tend to explain the causal chains of events, even if rather brief, better than the subjects in the control group. For example, subject 140 in the control group explains:

> Subject 140
>
> “Patient is known to have an IgE mediated process already in seasonal allergies. With 3 days worth of coughing and wheezing and no fever, an infectious process, though not off the differential is unlikely. It is also likely that his rash is allergic in origin and is worsening with the allergen likely to have caused the exacerbation of his asthma.”

While the above explanation contains a number of causal relations, some of them only restate the information in the scenario: “… known to have an IgE mediated process already in seasonal allergies.” In this sentence, the subject fills in part of the mechanism by which seasonal allergies work, the behavior of the Immunoglobulin E (IgE), but fails to associate it with the current symptoms. S/he also works through an elimination process, associating the absence of fever with the absence of an infectious process. This statement is correct by itself, but it does not explain a mechanism that supports the correct answer either. It is, however, a mechanism that invalidates an alternative answer.

In the end, the subject tries to explain the rash in the context of the correct answer, but generally fails to provide an explanation that supports the chosen diagnostic.

On the other hand, the subjects in the treatment group tend to provide, even if just in a few words, explanations incorporating more elaborate causal relations, such as subjects 73 or 137 whose explanations were analyzed before. Their explanations also tend to provide more details related directly to the causal processes that are behind the correct answer than those of the participants in the control group, who seem to have a tendency to work through a differential diagnosis rather than to provide a direct explanation, such as subject 104's explanation presented above.

Reasons for Changing the Answer

A first look at the participants that changed their answer reveals that of 12, only two participants were in the treatment group, while the remaining 10 were in the control group condition. This observation suggests that the subjects that practiced self-explanation considered their answer to the performance measurement question more carefully than the participants that did not practice self-explanation, thus showing less need to revisit it when given the chance. Looking at the confidence levels expressed by the subjects that decided to change their answer one can observe that on a scale from one to 10, none of the participants rated their confidence with more than five, with most of them in the one and two range.

Further analysis suggests that the main reason why subjects changed their answer is related to the need to reread or reexamine the scenario in order to elaborate the explanation required by the performance assessment question. Subject 80, instead of presenting an explanation in his or her first explanation, admits "Now that I'm reading

the question again I know that it can NOT be Roseola as …" Then, when explaining the rationale for which the answer was changed s/he follows up with an explanation of the correct answer. Alternatively, subject 90, who in his or her initial explanation states: "No fever, history, and I think I am wrong now that I am looking at it" but still misses the answer afterwards.

Once more, these explanations show the impact of the last symptom presented in the scenario text, the rash. It looks like many of the subjects in this category considered this symptom to be determinant for the diagnostic and somehow disregarded all the other symptoms presented to them earlier in the description. After rereading the scenario, most likely while trying to explain the mechanisms, the subjects tended to reassess the importance of the rash symptom and decide upon the correct answer choice.

There is at least one subject who admited that s/he made the wrong choice on accident as noted by "Clicked the wrong button" and then took the chance to correct that error. One other subject admitted that s/he was "… a bit distracted doing multiple things at once while taking this survey" and got it right the second time.

Summary

For this research, data collection was conducted during the month of April 2009 at a midwestern medical school. 147 first and second-year medical students agreed to participate, of which 117 provided usable data points. The dependent variable in the quantitative studies was the performance score, with its three variants, total performance score, categorical performance score and adjusted performance score. The main independent variables in the various analyses were the experiment condition and the measure of prior knowledge, with its two variants, measured prior knowledge and self-

reported prior knowledge. Due to the inclusion of two years of study in this experiment, the analysis controlled for the year of study by including it as a covariate where appropriate.

Table 21 presents the summary of the results for the first research question: "Does the practice of self-explanation as causal mechanism elicitation technique affects, on average, learners' performance on causal reasoning tasks?" Overall, if one does not account for the year of study, the treatment does not seem to have any significant impact on the subjects' performance. The year of study, introduced as covariate, was significantly related to the performance score.

Table 21

Research Question 1 - Summary of results

Performance score (PS)	Type of statistical test	Results
Categorical PS	Logistic regression	No significant effects found Significant effect of the Year of study
Total PS	ANCOVA	No significant effect of experiment group found Significant covariate relationship
Adjusted PS	ANCOVA	No significant effect of experiment group found Significant covariate relationship

Note: DV = Performance score, IV = Experiment group, Covariate = Year of study

For the second research question, "How does prior knowledge of relevant domains relate to the learners' performance on causal reasoning tasks?" the results are shown in Table 22. In summary, the results show that self-reported prior knowledge interacts with the experiment group, when the statistical test controls for the year of study, for all performance scores. Measured prior knowledge does not seem to interact with the treatment condition in the case of the total performance score; it seems to have a

significant simple effect for the adjusted performance score and it appears to interact with the treatment condition for the categorical performance score.

Table 22

Research Question 2 – Summary of results

Performance score (PS)	Prior knowledge (PK)	Type of statistical test	Results
Total PS	Measured PK	ANCOVA	No significant interaction effects No significant simple effects
	Self-reported PK	ANCOVA	Significant interaction Simple effects were discarded
Adjusted PS	Measured PK	ANCOVA	No significant interaction Significant simple effect of PK Not significant effect of treatment
	Self-reported PK	ANCOVA	Significant interaction effect Simple effects were discarded
Categorical PS	Measured PK	Logistic regression	Significant interaction effect Simple effects were discarded
	Self-reported PK	Logistic regression	Significant interaction effect Simple effects were discarded

Note: DV = Performance score (PS), IV1 = Experiment group, IV2 = Prior knowledge (PK), Covariate = Year of study.

Despite the fact that collecting qualitative data was impacted by several significant limitations, a few trends emerged. For low performing participants, the answers tended to be brief with no reference to causal mechanisms or causal chains, whereas for the high performing participants, the answers are sensibly more elaborate, included references to causal mechanisms and eventually causal chains. Nevertheless, the causal chains that were extracted from the participants' answers were short, linear, and sometimes with one or two branches, yet never included feedback loops.

No significant differences could be observed in the answers of the participants at the low end of the performance scale between the control and the treatment groups. In

comparison, at the high end of the performance range, the participants in the treatment group seemed to incorporate more elaborate causal relations in their answers and provide more details than those in the control group.

The inclusion of a symptom unrelated to the correct causal chains, the rash, prompted many participants to either use it as the most important symptom for their diagnostic (choosing Roseola as answer) or tried to explain it in the context of the correct answer (asthma attack). Finally, the decision to change the answer seems to have been determined by the participant's need to reread the scenario statement in order to provide the explanation they were asked for, in order to support their answer choice.

CHAPTER 5: Discussion and Implications

Understanding, explaining and using causal phenomena are integral parts of everyday human activities. In Dickinson's own words "Without the capacity to learn about and represent causal relationships between our actions and their consequences, the mind would be radically disconnected by the world" (2001, p. 3). Nevertheless, while causal reasoning and understanding is important for learning and everyday and scientific reasoning (Carey, 1995; Keil, 1989) scientific research reveals the many problems people have in dealing with causality and causally linked events (Chapman & Chapman, 1969; Chi, 2000b; Perkins & Grotzer, 2000; Tversky & Cahneman, 1989). In this respect this study looked at how cognitive strategies – more specifically self-explanation – support causal reasoning and understanding. For this purpose, an experiment was designed and deployed and data were collected and analyzed.

Results and Analysis

Research Question 1

The first research question asked if the practice of self-explanation as causal mechanism elicitation strategy affects, on average, learners' performance on causal reasoning tasks. Its aim was to observe whether significant effects on performance exist on near-transfer causal tasks for the participants who practiced self-explanation prior to this task as compared to those participants who did not. During the experiment the

participants in the treatment group practiced self-explanation on a set of three multiple-choice questions by responding to a prompt that asked them to provide an explanation of the mechanisms behind their decision. The participants in the control group answered the same three multiple-choice questions, but without being asked for an explanation. No explanation about what self-explanation based strategies are was provided to the participants and no training in how it should be used was delivered. Three performance scores were calculated from the collected data: 1) a categorical performance score which, represents the raw true/false answer to the multiple-choice performance question, 2) a total performance score, which accounts, in addition to the raw score, for the quality of the mechanism explanation and 3) an adjusted performance score, which accounts for the participants' confidence in the correctness of their answer. For analysis, logistic regression was used to analyze the categorical performance score and factorial ANCOVAs were used for the total and adjusted performance scores.

Data analysis shows that no significant differences can be observed between the participants in the control group and those in the treatment group. The only study variable that shows any significant relationship is the year of study, included in the analysis as a covariate. That is, the logistic regression for the categorical performance score shows a significant effect of the year of study. More specifically, being a first-year medical student decreases the odds of answering correctly versus answering incorrectly by a factor of 0.347, when compared to second-year medical students. The ANCOVA analyses for Total Performance Score and Adjusted Performance Score show a significant covariate relationship, with second year students performing better than first year students. If one assumes that the year of study is a proxy for prior knowledge, the

results indirectly support prior findings which shows that performance is linked to the level of prior knowledge for both causal reasoning (e.g., Ahn & Baileson, 1996; Ahn, et al., 1995; Thagard, 2000) and for self-explanation (e.g., Bielaczyc, et al., 1995; Chi, et al., 1989; Kintsch, 1994). That is, for causal reasoning, the higher the level of prior knowledge, the better the performance, whereas for self-explanation, the higher the level of prior knowledge, the stronger the effect of using the self-explanation strategy on performance. Several factors, such as the type and strength of the treatment condition, the duration of the treatment, the timing of the experiment or the level of prior knowledge the participants had, can be held accountable for the observed outcome.

One factor that could account for the observed outcome is the type and strength of the treatment condition. Study participants were introduced to the practice of self-explanation indirectly through three successive practice exercises and without being aware of the strategy they were asked to use. Furthermore, the prompt used was generic and open-ended, with no prior scaffolding, instructions or directions. This argument is supported by prior findings which suggest that mere exposure to self-explanation leads to some improvement, but significantly less than when participants are instructed in how to use these strategies (Bielaczyc, et al., 1995). In similar studies, the success of cognitive strategies based on self-explanation is associated with more direct prompts (e.g., Berthold, et al., 2009; Hausmann, VanLehn, Nokes, & Gershman, 2009; Kirschner, et al., 2006), usually in multiple steps (e.g., Bielaczyc, et al., 1995). That is, overall, the treatment might have been too weak to produce recognizable effects at this level.

The time spent practicing self-explanation and the type of task could also account for the observed result. More specifically, the task used for practicing self-explanation

was one familiar to the participants (answering multiple-choice questions) to which the treatment added an open-ended mechanism explanation component. Medical students are trained to quickly answer multiple-choice questions without too much consideration to the explanation behind the scenario which is consistent with the test-taking behavior considered to be effective for medical board examinations. This means that if the participants used a cognitive strategy based on self-explanation before, they might have had the time to observe what they were asked to do. Since the three practice questions followed six multiple-choice questions designed to assess prior knowledge, some participants might have entered into "test-taking" mode and quickly skimmed over the question and answers, overlooking the explanation they were asked to provide. In this respect, they might have been what Renkl (1997) calls "superficial explainers"[21] and "passive explainers"[22] which raises the question of the effectiveness of the type of prompt used for self-explanation, supported by prior research which suggests that if the participants are not assisted to self-explain, the strategy does not show its full potential and produces diminished returns (Berthold, et al., 2009).

As suggested by existing research (e.g., Bielaczyc, et al., 1995; Renkl, 2002), participants' prior knowledge of the three domains at the time of the experiment could also be considered to have played a role in the observed outcome. With both first- and second-year medical students being invited to participate in the study, a look at the sequence of courses in the curriculum shed some light on the observed results. With the experiment designed for second-year medical students, an analysis of the curriculum

---

[21] Superficial explainers are explainers who assign little time to the task of explaining.

[22] Passive explainers are explainers with a very low level of self-explanation activity.

sequence of courses shows that, on average, first-year medical students should have had the knowledge to answer the question. Nevertheless, during their second year, more advanced courses on the same subjects as well as on associated topics deepen understanding of the subject matter. Considering that the Year of Study could be a rough representation of the level of prior knowledge the participants had, the analysis of the impact that participants' level of prior knowledge had on performance might be able to provide a better explanation. A second research question attempts to further study the effects of prior knowledge on the effectiveness of self-explanation based cognitive strategies for causal reasoning contexts.

Research Question 2

Starting from existing research (Bielaczyc, et al., 1995; Renkl, 2002) the second research question asks if the prior knowledge of the relevant domains (immunology, pathology and physiology) is related in any way to the learners' performance on causal reasoning tasks. To answer this research question, two measures of prior knowledge were included in the analysis. One measure was self-reported, for which the participants self-assessed their level of prior knowledge in each of the three domains related to the performance assessment question. The second one measured prior knowledge using a set of six multiple-choice questions, two for each of the three domains. The same scores analyzed for the first research question were studied using logistic regression and factorial ANCOVAs for both measures of prior knowledge.

For the categorical performance score, the treatment (use of self-explanation as causal mechanism elicitation strategy) seems to improve performance for the participants with low prior knowledge but does not show any effect for the participants with high

prior knowledge. That is, for the participants who were found to have low levels of measured prior knowledge, being in the treatment group increases their odds of answering the question correctly. For self-reported prior knowledge, this strategy works for participants that report high levels of prior knowledge, while it shows no effect for the participants who report low levels of prior knowledge.

The categorical performance score was used to measure the effect of using the self-explanation-based cognitive strategy on participants' raw performance scores. In both analyses, the interaction between prior knowledge and treatment condition was significant, supporting extant research that shows prior knowledge to be important for performance in both causal reasoning (e.g., Ahn & Baileson, 1996; Ahn, et al., 1995; Thagard, 2000) and self-explanation (e.g., Bielaczyc, et al., 1995; Chi, et al., 1989; Kintsch, 1994). The apparent issue here is that while for measured prior knowledge the effect is found to be beneficial for the participants in the low prior knowledge category, for self-reported prior knowledge, the effect seems to be beneficial for the high prior knowledge participants. With the effects of the year of study removed by including it as a covariate in the statistical analysis, the only differentiating factor remains the method by which prior knowledge was measured.

The use of two methods to estimate prior knowledge was necessary because of the experiment access, time and length limitations imposed by the participating institution which limited the number and complexity of questions that could be used to measure this variable. It is likely that six multiple-choice questions (two per domain) are too few to

reliably measure the participants' prior knowledge[23]. For this reason, considering existing research (e.g., Ackerman, et al., 2002; Anaya, 1999; Boud & Falchikov, 1989; Falchikov & Boud, 1989; Fitzgerald, et al., 1997; Fitzgerald, et al., 2003), the self-reported prior knowledge can be used reliably as a substitute measure of prior knowledge for advanced learners, such as graduate students, for the given experiment conditions.

The total performance score was introduced to account for the quality of the mechanism explanation associated with the answer choice. For this performance score the analysis shows no main effects or interaction when looking at measured prior knowledge, but shows a significant interaction when looking at self-reported prior knowledge. For self-reported prior knowledge, a positive impact of the treatment condition was found for the participants who report having high levels of prior knowledge. No significant influence was found for the participants reporting low levels of prior knowledge.

The inclusion of a score for the quality of the mechanism explanation provided a better estimate of the participants' understanding of the phenomenon. As a side effect, the increase in variance also helped the statistical analysis.

For self-reported prior knowledge, analysis of this measure of performance suggests that a certain level of prior knowledge is a prerequisite for a strategy based on self-explanation, applied without direct prompting, to work. That is, a prior knowledge threshold needs to be reached before improvements can be observed as a result of using this strategy. Looking from the perspective of VanLehn's (1993) gap-filling explanation,

[23] For comparison, the board examinations the students have to take to enter a residency program, the USMLE STEP 1 and STEP 2 CK have 336 and 370 multiple-choice questions respectively. Considering the number of topics reported to be assessed in each of the two step examinations there are, on average, 20 to 25 questions per domain.

it is probable that too little prior knowledge limits learners' ability to recognize and fill knowledge gaps by themselves, without any outside intervention, as suggested by Kirschner (2006). This suggests that when using self-explanation as a cognitive strategy, a different approach for low and high prior knowledge learners, similar to those discussed by existing research (e.g., Atkinson, et al., 2003; Hilbert, Schworm, & Renkl, 2004), could be more successful.

The adjusted performance score uses participant reported confidence in his or her own answer to account for the strength of participants' causal mental model and attenuate the effects of the participants "guessing" the answer instead of choosing it based on sound causal reasoning. For the adjusted performance score, for the measured prior knowledge variable, the analysis shows a significant main effect of prior knowledge, with those participants with higher levels of prior knowledge performing better than those with lower levels of prior knowledge, but does not show any interaction between prior knowledge and the treatment condition. For self-reported prior knowledge the analysis shows a significant interaction between the level of prior knowledge and the treatment condition, suggesting significant positive effects of the treatment for those reporting high levels of prior knowledge but with little or no effect for those reporting low levels of prior knowledge. The results suggest an outcome similar to that observed for the total performance score and, as before, the addition of a quasi-continuous score increases variance and thus provides better support for the statistical analysis.

Figure 7 presents an overview of the trends observed so far for the second research question. The uneven outcomes obtained when using an objective measure of prior knowledge are most likely due to the measurement shortcomings discussed before. Using a self-reported measure of prior knowledge provides more consistent behavior, with the analysis for all three scores showing a significant interaction between the treatment condition and the level of prior knowledge. Further analysis suggests that a cognitive strategy using practice of self-explanation, similar to the one used in this experiment, seems to help the participants in the high self-reported prior knowledge group, while having little or no influence for the participants reporting lower levels of prior

knowledge. Therefore the prior knowledge of the domain(s) seems to have a threshold value or range above which a cognitive strategy based on self-explanation could have positive effects, similar to the one used in this study.

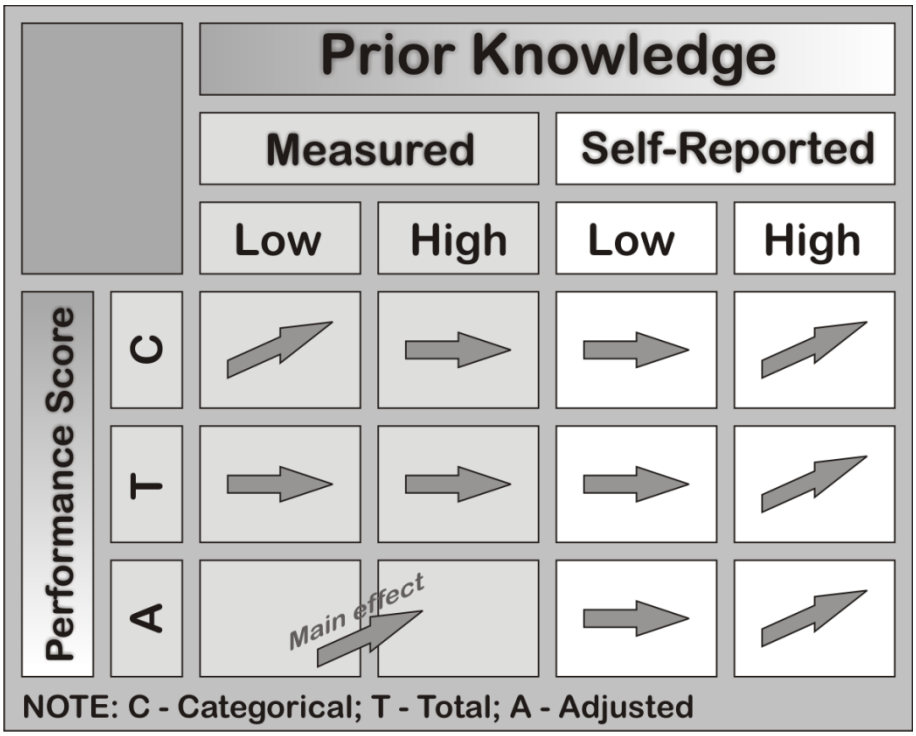


Figure 7: Research Question 2 - Overview of findings

Research Question 3

The lack of appropriate incentives for the participants to provide detailed answers to the performance assessment explanation prompt question significantly impacted the quality of the answers, thus limiting the ability of the researcher to perform an in-depth analysis. Nevertheless, even with suboptimal data, some of the problems researchers report for causal reasoning can be recognized in the participants' answers. For example, the focus on the rash symptom, unrelated to the correct causal chains, as a determinant factor for the diagnosis for some of the participants, resonates with the issue of differentiated weight assigned to evidence based on understanding of the causal

phenomenon (Pennington & Hasties, 1988). From a different perspective, once the participant decided upon a diagnostic of Roseola the observed behavior is consistent with the tendency to favor confirming versus disconfirming evidence (Schustack & Sternberg, 1981). The same observed behavior could also be viewed as evidence for too much focus on a possible cause to the detriment of other available information (Schustack & Sternberg, 1981).

The tendency to express simple, linear, and sequential causal models (Bullock, et al., 1982; Grotzer, 2000) can also be observed in the participants' answers. At the low end of the performance spectrum the causal models are nonexistent, whereas at the higher end, the models extracted from the participants' explanations are simple and show the tendency to skip intermediate objects (e.g., Besson, 2004). Future research should either use a different approach to collecting qualitative data and/or include appropriate incentives and prompts to improve for depth and breadth of this data.

Conclusions and Implications for Research and Practice

Using of self-explanation based strategies, even in their weaker forms, to manage the issues people have with causal reasoning and causal understanding shows promise. Using a weak and nonintrusive intervention based on a brief practice of self-explanation before the task this study shows that learners reporting higher levels of prior knowledge of the relevant domains seem to benefit from the use of this strategy, while the learners reporting lower levels do not. From a theoretical research perspective, this finding suggests the existence of a threshold value or range of prior knowledge in the relevant domain(s) that needs to be reached before cognitive strategies based on self-explanation similar to the one used here could become effective.

A direct practical implication is that if the level of prior knowledge exists, practice-based self-explanation cognitive strategies could be easily implemented in both face-to-face and online situations, with almost no time cost for the learner. It is probable that the threshold is different for different domains and contexts. Therefore, success or failure to use this strategy could serve as an indication that more involved methods, such as multi-step prompting and training in the use of self-explanation strategies, are needed to support students' causal reasoning until they accumulate sufficient domain knowledge to reach the threshold. For more difficult subjects learned over longer periods of time, more involved self-explanation based cognitive strategies could be used at the beginning of the period, faded afterwards to less intrusive strategies, such as the one used in this experiment.

As this experiment was conducted entirely online, using a web-based application, there are implications for the design of online instruction and of online learning environments. With proper preparation, support for self-explanation based cognitive strategies is easy to implement in online environments, and has the potential to provide significant gains in learner performance with minimal intrusion and cost. Expanding on this idea, use of appropriate scaffolding and prompting could impact a variety of computer-based tools which could benefit from implementing support for cognitive strategies to aid causal reasoning and understanding through causal mechanism elicitation. For example, when needed, causal mapping or concept mapping tools could include components to elicit and use causal mechanisms when developing causal or concept maps in cases where increased depth and reliability are needed.

Limitations

Various events and conditions weighed on the design, deployment, and data collection in this study. Most of the limiting factors were introduced throughout this document to account for design decisions, data quality issues or unexpected outcomes. Below is a summary of the most relevant of these limitations and of their effects.

The participating institution imposed the most important limitations to this research. The areas affected were the type of access to their students and the length and complexity of the tasks these students could be asked to perform. First, the number of instrument elements that could be dedicated to measuring prior knowledge by objective measures was severely limited, leading to less useful data. Without access to other objective measures of prior knowledge, such as examination or test grades, the analysis was limited to self-reported measures of prior knowledge. Improved and more fine-grained measurement of prior knowledge would have helped in better defining the prior knowledge threshold from which such strategies become effective. Second, the suggested length (time) and difficulty of the instrument components limited the depth and breadth of the performance measurement tasks, thus constraining the types of statistical tests that could be used to analyze the data and thus the power to observe and confirm trends and causal relations.

The online delivery of this experiment impacted the ability of the researcher to control for participants' behavior during the experiment. This limitation mainly affected the quality and detail of participants' explanations, limiting the depth and breadth of the qualitative analysis designed for the experiment. In addition, no control over the environment in which the experiment took place left open the possibility of external

interfering factors affecting the participant's performance. Therefore, the results provided by the qualitative analysis must be interpreted with care and only within the boundaries (including the limitations) of the present experiment.

Another factor that could have affected the outcomes of this experiment both positively and negatively is its timing at the end of the semester and school year when students are facing more concurrent activities than usual. Among the positive effects are the ability to recruit first year medical students to participate in the research and the increased reliability in self-reported measures of prior knowledge of relevant domains. The negative effects are related to the time and effort the participants were willing to spend on completing the experiment's tasks, with negative effects on the depth, breadth and quality of the data collected.

Implications for Future Research

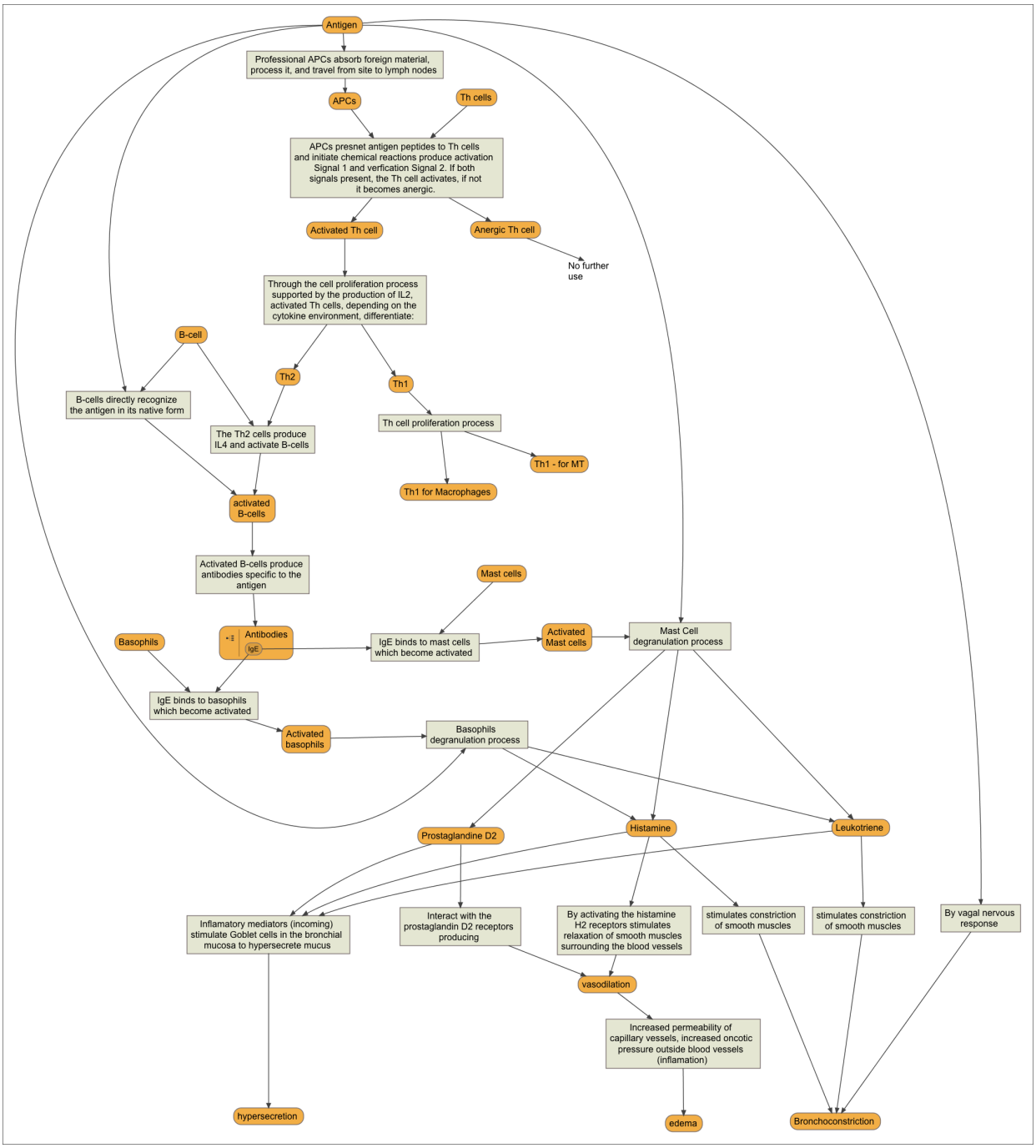
This research makes only a small step toward understanding how cognitive strategies can support causal reasoning and understanding in people. As a natural continuation, future research should be developed to test the observed behaviors using a better objective measure of prior knowledge as well as an expanded performance measurement task. Further studies should also look at how the type and timing of prompts used to elicit self-explanation affect performance. The potential outcome would be the development of a best practices roadmap for the use of self-explanation based strategies to aid learning and the understanding of causality. With this in mind, the research can be further extended to other cognitive strategies aimed at eliciting mechanism explanations during causal reasoning.

The nature and level of the prior knowledge threshold above which practice of self-explanation based cognitive strategies begins to become effective is worthy of further investigation. From this perspective, it would be interesting for example, to find what type(s) of knowledge (e.g. factual, process, etc.) is more important and why. As practical applications of cognitive strategies such as the one used in this study depend on the ability to estimate the various variables in the equation, future studies should also look for ways of estimating this threshold level.

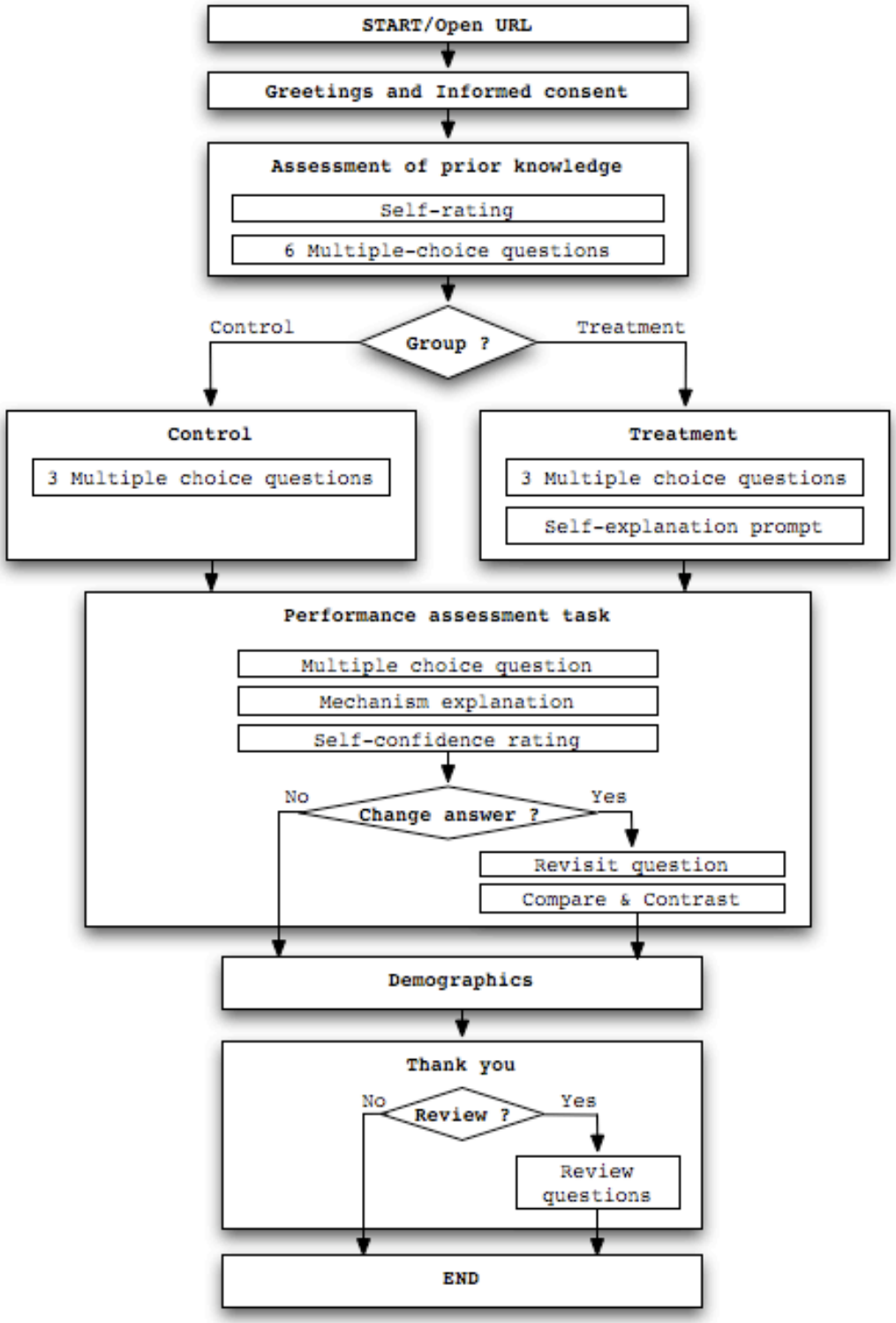
With little research, at best, on methodologies and tools capable of improving people's causal reasoning and understanding, the research field is open to look at a variety of cognitive strategies. Looking only at self-explanation, future research could study its effects on causal reasoning and understanding in various domains and disciplines, using different delivery methods (e.g., mixed), or for different types of groups (e.g., small vs. large).

With causal mechanism elicitation cognitive strategies able to improve people's performance on causal reasoning tasks, future studies could look at various methods of implementing such strategies in a variety of practical applications, ranging from concept mapping to decision making.

APPENDIX 1 - Influence Diagram for Asthma



APPENDIX 2 - Instrument Workflow



APPENDIX 3 - Research Instrument

The instrument presented below is mirrored online on the website where the experiment is delivered. The application URL is [http://rmed.gionas.webfactional.com].

Informed Consent

You are invited to participate in this study titled "Impact of self-explanation on causal reasoning". This study is not intended to assess your knowledge, but to study the way you use causal reasoning and explanations.

This consent requests your permission to allow the researcher to collect and analyze your answers and work done while using this website, as well as use the results of such analyzes in presentations at conferences and for printed publications.

Your participation in this study is voluntary and there are no anticipated risks or discomforts to you. You have the right to withdraw from the study at any time with no questions asked and no repercussions to you. If you are a student, the decision to decline participation or withdraw will not affect your grades or academic standing in any way. All answers are confidential and will be available to the researcher only. Your identity or any data that could identify you will never be communicated to anyone. For analysis identification data will be replaced with codes.

If you have any questions about this research and your participation, now or at any time, please feel free to contact:

Ioan Gelu IONAS

School of Information Science and Learning Technologies

University of Missouri - Columbia

111 London Hall

igieeb@missouri.edu

For additional information regarding human participation in research please contact either the KCOM IRB Office at (660) 626-2316 or the UMC Campus IRB Office at (573) 882-9585.

By providing your name and e-mail address as requested below and continuing to the website, you acknowledge that you understand the conditions of this invitation, voluntarily agree to participate in this study, and allow the results to be used for the research purposes stated above.

[First Name entry field]

[Last Name entry field]

[e-mail address entry field]

<Continue>

Assessment of Prior Knowledge

Researcher note: Both experiment groups will be asked to perform the same tasks in the prior knowledge assessment section.

Self-Assessment of Prior Knowledge

On a scale from 1 to 10, where 1 means "Nothing at all" and 10 means "I am an expert", how would you rate your knowledge of:

- Physiology
- Pathology
- Immunology

Multiple Choice Questions for Prior Knowledge Assessment

Question 1: The effect of sympathetic nervous system include (check all that apply:)

6. Contraction of the bladder detrusor muscle (F)
7. Papillary dilation (T)
8. Reduced gastrointestinal motility (T)
9. Constricts bronchiole smooth muscle (F)
10. Constricts skin and mucus membrane blood vessels (T)

Question 2: A 28 week premature infant receives surfactant through endotracheal tube. Which of the following would you expect?

11. Decrease in alveolar surface tension and increase in lung compliance (Correct)
12. Increased alveolar surface tension and decreased lung compliance
13. Increased alveolar surface tension and no change in lung compliance
14. Decrease in alveolar surface tension and decrease in lung compliance.

Question 3: In the presence of inflammation the following are raised (check all that apply):

15. Platelets (T)
16. Ferritin (T)
17. Caeruloplasmin (T)
18. Fibrinogen (T)

19. Complement proteins (T)

Question 4: An 18 month old infant male presents with sudden onset of cough. Suspecting a foreign body aspiration in which area of the lung does this happen more frequently?

20. Left upper pulmonary lobe
21. Right upper pulmonary lobe
22. Right lower pulmonary lobe (Correct)
23. Left lower pulmonary lobe

Question 5: The following are true about lymphocytes (check all that apply):

1. T cells account for 20% of the circulating lymphocytes (F)
2. In the spleen, B cells are found in the periarteriolar areas of white pulp (F)
3. In the lymph nodes, T cells occupy the paracortical area surrounding the germinal centers (T)
4. B cells but not T cells express surface IgG (T)
5. B cells undergo maturation in the bone marrow (T)

Question 6: Three days ago, a 13-year-old boy developed a pruritic, migratory, urticarial rash on his torso, face and extremities, and mild edema of his hands and feet. He had been well until he developed symptoms of an upper respiratory tract infection. He had a fever initially, but that has now resolved. He has taken one dose each of

acetaminophen and ibuprofen, but no other medications. He has no history of urticaria, drug allergy, or adverse reactions to food.

Which of the following is the most likely cause of his illness?

1. Allergy to food flavoring
2. Nonsteroidal anti-inflammatory drug reaction
3. IgE response to acetaminophen
4. Infectious-triggered urticaria (Correct)
5. Underlying autoimmune disease

Treatment

Researcher note: In this section, which provides the participants with self-explanation practice opportunities, the treatment consists of the overt request to explain mechanisms that are at work within the presented context. This requirement is for the participants assigned to the treatment group only.

Scenario 1

Both groups: A 10-year-old child presents with a persistent sinus infection. Over the course of the last 5 years, the patient has had several sinus and upper respiratory infections. Blood tests reveal abnormally low levels of IgA immunoglobulin, but normal levels of their isotopes.

Treatment group only (open-ended mechanism self-explanation): What mechanisms are most probable at play in this case?

Both groups: The patient's recurrent infections are most likely due to a defect of which of the following?

1. Fixation of complement

2. Mast cell activation
3. Mucosal immunity (Correct)
4. Neutrophil activation
5. Opsonization of bacteria

Scenario 2

Both groups: A 9-year-old boy is brought to the emergency room because of vomiting, abdominal cramps, and difficulty breathing. Symptoms began 30 minutes ago while he was dining in a restaurant. Temperature is 37.2^0C, pulse rate 140/min, respiratory rate 30/min, and blood pressure 90/60 mm Hg. Eczema is seen in the flexural folds and blanching, erythematous rash is present on the trunk and face. Expiratory wheezing is heard.

Treatment group only (open-ended mechanism self-explanation): What kind of reaction is this and which is/are the most probably mechanism(s) by which it is produced?

Both groups: Which of the following is the most likely diagnosis?

1. Anaphylaxis (Correct)
2. Asthma
3. Food poisoning
4. Gastroenteritis
5. Drug ingestion

Scenario 3

Both groups: In a teenager who gets stung by a bee, urticaria develops within 30 minutes of the sting.

Treatment group only (open-ended mechanism self-explanation): What type of reaction is this and which is/are the most probable mechanism(s) by which a bee sting produces urticaria?

Both groups: This reaction is most probably mediated by:

1. Complement component C3
2. Immunoglobulin E (IgE) antibodies (Correct)
3. Neutrophils
4. Natural killer (NK) cells

Transfer Task

Researcher note: The tasks in this section are the same for all participants.

Step 1: Multiple-choice Question

A 3-year-old male, known with seasonal allergies, presents at the clinic with difficulty breathing - increased respiratory rate, wheezing, and cough. Symptoms started two days ago. No fever. Parents noted this morning a fine rash appearing on his abdomen. The rash seems to be getting worse.

Which of the following is the most probable diagnostic?

1. Anaphylaxis
2. Asthma exacerbation (Correct answer)
3. Pneumonia
4. Foreign body aspiration
5. Roseola

Step 2: Open-ended Mechanism Explanation

What are the mechanisms that suggest your diagnosis?

Step 3: Confidence Question

On a scale from 1 to 10, where 1 is "Not confident at all" and 10 means "I am sure", how confident are you that your answer is correct?

Step 4: Choice for Review

Do you want to change your answer?

Researcher note: If the answer to this question is "Yes", then go to Step 5. If the answer is "No", go to next section.

Step 5: Review

Researcher note: Present information in Step 1 again and highlight previous answer. Show answer (choice and explanation) from question in Step 2. Allow participant to choose a different answer.

Ask question (open-ended): Why did you change your answer and why and how is the new answer better than the first one?

Demographics

Question 1: In which age group are you? (choices: <20; 20-25; 25-30; 30-35; 35-40; >4)

Question 2: Are you a Male or a Female

Question 3: In which of the following income groups are you? (choices <$20K, $20K-$40K, $40K-$80K, $80K-$150K, >$150K)

Question 4: Which medical specialty do you intend to pursue? (open-ended)

Wrap-up and Feedback

Researcher note: This last section shows a thank you note for participating in the research and offers the participant the option to review his/her answers to the 10 multiple-

choice questions in this study. If the participant decides s/he wants to review, each question will be reviewed on a separate page.

APPENDIX 4 - Think-aloud Protocol

[Introduction]

"We're here with Dr. _____ to assess, from an expert's point of view, the answers subjects provided to a complex question which combines a multiple choice answer with an explanation for an experiment on causal reasoning and understanding."

[Task description]

"Dr. _____, your task today is to score the answers presented to you, in the order they are presented. For scoring, I would suggest you to use a differential scale, ranging from -3 to +3, where -3 for the worst answers while +3 is for the best answers. If you believe that this scale is not appropriate for scoring these answers, could you suggest a different one?"

"In the table in front of you, for each line, when beginning assessment of a new subject's answer, speak aloud the number in the first column. This number uniquely identifies a subject's answer so it can be used in further analysis. The second column is the answer choice the subject made and the third column is the explanation s/he provided for that choice. The correct choice was 'Asthma exacerbation'. Please use the last column to mark the score and make notes if necessary."

"During the entire process please verbalize your scoring process by verbalizing your thoughts. Be aware that I might intervene occasionally, depending on situation, to maintain this process in line with these instructions. For example, if you stop verbalizing your thoughts, I will intervene and remind you to do so."

[Address expert's questions and concerns]

"This is all I had to say. At this point, do you have any questions or concerns you need to address?"

[If Yes]

Go on and address the expert's questions and concerns.

[Ask for consent]

Do you agree for the information you provide during the session today to be recorded and used, as is or interpreted, for the purposes of this research, in publications, and reports? If you agree, please say 'Yes, I agree'. If not, please say 'No, I do not agree'.

[If consent not obtained]

"Thank you for accepting to consider helping with this research. Since you do not accept to allow the researcher to record and use the information you provide during this session we cannot continue at this time."

"Would you like to change your decision?" [state the consent statement again]

[Begin]

"Let's start with the first subject."

[At the end]

"We have concluded this assessment session and I would like to thank you for your help in conducting this research."

[Address expert's questions and concerns]

"If you have any questions, concerns, or remarks or comments you would like to ask or make, please feel free to do so."

[If Yes]

Listen to what the expert has to say and address any eventual questions and concerns s/he may have.

[Thank you]

"Thank you again for helping with this research."

Contingencies

[If the subject becomes quiet]

Use one of the prompts below or other similar prompt, but do not interfere with the expert's thought process.

I'd like to hear what you're thinking

If you could just say whatever words come to your mind

Please speak up

[At key junctions]

When learning something new that is key to understanding, summarize the event and the thinking that the user explored, very briefly. Ask the participant to provide more detail.

APPENDIX 5 - Scoring Rubric

Symptoms and related keywords. They represent mostly a repetition of the correct and relevant symptoms presented to the participants in the problem: seasonal allergies, difficulty breathing (increased respiratory rate, wheezing, cough, no fever), allergens. The presence of relevant keywords are an indication that the participant has an understanding of the symptoms, recognized the situation, and decided not to repeat the symptoms from the scenario (IgE, mast cell, histamine, cytokine, T-helper lymphocytes). For this item, scores are awarded as follows:

0 if none of the relevant symptoms or keywords are found in the answer

+1 all or some of the relevant symptoms are present

+2 presence of relevant keywords with or without symptoms.

-1 if incorrect or non-relevant symptoms are present in the wrong context (the rash for example, which is not necessarily relevant in this case, but only if it presented by itself or in the wrong context)

Level of detail is represented by how deep the participant reaches with the explanation. The most detailed level of explanation expected by the expert was at the cell level. There are two elements for a complete answer: a) mentioning the Immunoglobulin E (IgE) and b) one mentioning the cell processes (mast cell degranulation processes). Scores are awarded as follows:

0 if no detail is present in the explanation (see the other scores for examples)

+1 if the participant mentions either a) or b)

+2 if the participant mentions both a) and b)

+1 An additional point can be awarded for really detailed explanations for each or both of the two elements. Keywords indicative for a more detailed explanation are: T helper cells, cytokine).

Presence of mechanisms, represented by the presence of correct sequences of relevant events that occur in an asthma attack. Scores are awarded as follows:

0 if no evidence of a relevant mechanism is found in the answer

+1 if the answer mentions disparate correct sequences (aka. Mechanisms). For example, mentioning just one causal link such as a) the presence of a mediated reaction, b) mast cell degranulation, c) histamine or cytokine release, d) mast cell proliferation, e) mediated response to allergens, etc could serve as an indication for the existence of a causal relation.

+2 for the presence of correct chains of such causal relations. For example, a more complex example here would be a) inflammatory reaction goes in the asthma, leading to mast cell proliferation, leading to wheezing, by bronchoconstriction, to cough, increased respiratory rate or b) mentioning the mast cells and the histamine and the fact that histamine produces other relevant reactions at the bronchiolar level: bronchospasm, inflammation, and edema of the bronchiole tree.

-1 for the presence of incorrect causal relations (linking the rash to Roseola - it is correct to link it since rash is part of the roseola symptoms, but not in the context of this scenario since many other viral infections can produce a rash). This should be carefully considered to be in the appropriate context (e.g., some respondents might have attempted to explain by eliminating the other alternative answers).

APPENDIX 6 - Extended calculations for logistic regression with measured prior knowledge

The generic equation for the logistic regression is:

$$Logit(p) = \beta_0 + \beta_1\, sgrp + \beta_2 mpk + \beta_3 sgrpmpk + \beta_4 yr$$

where:

sgrp – represents the experiment group variable

mpk – represents the measured prior knowledge variable

yr – represents the covariate, year of study.

From the logistic regression output (Table 16) the beta coefficients of this regression equation are:

$\beta_0 = -1.718$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = -0.950$ $\beta_4 = 1.014$

To help the analysis the regression equation will be analyzed for each of the two levels of the self-reported prior knowledge individually. This analysis will produce two regression equations, one for each level of the variable (Low and High). The year of study is included as a covariate for control purposes and will not be included in the analysis.

For *high measured prior knowledge* participants (setting the *mpk* variable to 0 according to the categorical variables coding scheme), the regression equation is:

$$Logit(p) = -1.718 + 1.014yr$$

The obtained equation does not contain a term for the experiment group. This leads to the conclusion that the experiment group *does not influence the odds* of answering the problem correctly versus answering it incorrectly for the participants in the *high measured prior knowledge* group.

For the *low measured prior knowledge* participants (setting the *mpk* variable to 1 according to the categorical variables coding scheme), the regression equation is:

$$Logit(p) = -1.718 - 0.950sgrp + 1.014yr$$

According to this regression equation, the odds ratio for the experiment group variable is:

$$odds\ ratio = e^{-0.950} = 0.387$$

This odds ratio shows that for the participants with *low measured prior knowledge* the odds of answering the question correctly versus the odds of answering it incorrectly are *decreased* by a factor of 0.387 by being in the control group (the group for which the prediction is being made, coded with 1) than in the treatment group (the reference group, coded with 0).

APPENDIX 7 – Extended calculations for logistic regression with self-reported prior knowledge

The generic equation for the logistic regression is:

$$Logit(p) = \beta_0 + \beta_1\, sgrp + \beta_2 mpk + \beta_3 sgrpmpk + \beta_4 yr$$

where:

sgrp – represents the experiment group variable

srpk – represents the prior knowledge variable

yr – represents the covariate, year of study.

From the logistic regression output (

Table 20) the beta coefficients of this regression equation are:

$\beta_0 = -2.052$ $\beta_1 = 0$ $\beta_2 = -0.999$ $\beta_3 = 1.472$ $\beta_4 = 1.130$

To help the analysis the regression equation will be analyzed for each of the two levels of the self-reported prior knowledge individually. This analysis will produce two regression equations, one for each level of the variable (Low and High). The year of study is included as a covariate for control purposes and will not be included in the analysis.

For *low self-reported prior knowledge* (setting the *srpk* variable to 0 according to the categorical variables coding scheme), the regression equation is:

$$Logit(p) = -2.052 + 1.130yr$$

The obtained equation does not contain a term for the experiment group. This leads to the conclusion that the experiment group *does not influence the odds* of answering the problem correctly versus answering it incorrectly for the participants in the *low self-reported prior knowledge* group.

For the *high self-reported prior knowledge* participants (setting the *srpk* variable to 1 according to the categorical variables coding scheme), the regression equation is:

$$Logit(p) = -2.052 - 0.999 + 1.472sgrp + 1.140yr = -3.051 + 1.472sgrp + 1.140yr$$

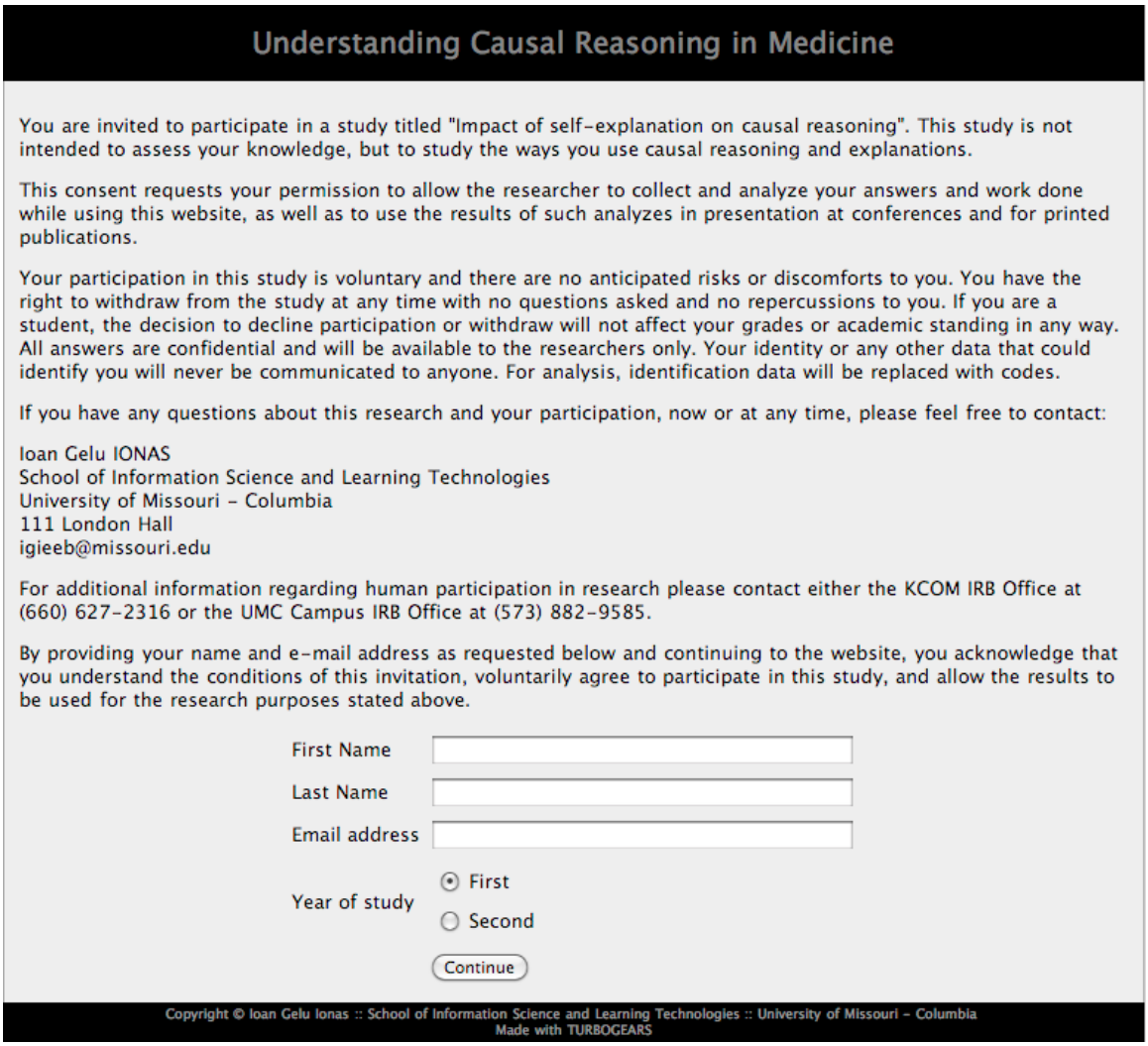
According to this regression equation, the odds ratio for the experiment group variable is:

$$odds\ ratio = e^{1.472} = 4.385$$

This odds ratio shows that for the participants with *high self-reported prior knowledge* the odds of answering the question correctly versus the odds of answering it incorrectly are *increased* by a factor of 4.358 by being in the treatment group (the group for which the prediction is being made, coded with 1) than in the control group (the reference group, coded with 0).

APPENDIX 8 – Screenshots of the experiment website

The following screenshots present the experiment website pages in the order they were presented to the participants. The differences between the control and treatment groups are mentioned in the caption text.



Understanding Causal Reasoning in Medicine

You are invited to participate in a study titled "Impact of self-explanation on causal reasoning". This study is not intended to assess your knowledge, but to study the ways you use causal reasoning and explanations.

This consent requests your permission to allow the researcher to collect and analyze your answers and work done while using this website, as well as to use the results of such analyzes in presentation at conferences and for printed publications.

Your participation in this study is voluntary and there are no anticipated risks or discomforts to you. You have the right to withdraw from the study at any time with no questions asked and no repercussions to you. If you are a student, the decision to decline participation or withdraw will not affect your grades or academic standing in any way. All answers are confidential and will be available to the researchers only. Your identity or any other data that could identify you will never be communicated to anyone. For analysis, identification data will be replaced with codes.

If you have any questions about this research and your participation, now or at any time, please feel free to contact:

Ioan Gelu IONAS
School of Information Science and Learning Technologies
University of Missouri - Columbia
111 London Hall
igieeb@missouri.edu

For additional information regarding human participation in research please contact either the KCOM IRB Office at (660) 627-2316 or the UMC Campus IRB Office at (573) 882-9585.

By providing your name and e-mail address as requested below and continuing to the website, you acknowledge that you understand the conditions of this invitation, voluntarily agree to participate in this study, and allow the results to be used for the research purposes stated above.

First Name
Last Name
Email address
Year of study
First
Second
Continue

Copyright © Ioan Gelu Ionas :: School of Information Science and Learning Technologies :: University of Missouri - Columbia
Made with TURBOGEARS

Introduction page and login

Understanding Causal Reasoning in Medicine

Instructions

To complete this survey quickly and accurately, plese read and follow the directions below:

DO NOT use the brower's back button. Even if you remember that you missed something in the previous question, do not attempt to go back. Once you move on the next page this website will not record any of the changes you make on any of the previous pages.

Please answer all questions to the best of your abilities. Make sure you understand what each question asks you to do as some questions have more than one right answer. These questions are indicated by the comment "Check all that apply" and have check boxes at the left of each option.

Please explain everything to the best of your knowledge. You do not need to write whole sentences as long as what you write conveys the information correctly and completely.

IMPORTANT NOTE: If for some reason you receive an error while using this website, please wait for two or three seconds and refresh the page. If this action does not work, please restart from the beginning and provide the same identification information as the first time. Thank you for understanding!

Continue

Instructions for the participants

Understanding Causal Reasoning in Medicine

Using a scale from 1 to 10, where 1 means "Nothing at all" and 10 means "I am an expert", rate your knowledge in the following areas:

Physiology

	1	2	3	4	5	6	7	8	9	10	
Nothing at all	○	○	○	○	○	○	○	○	○	○	**I am an expert**

Pathology

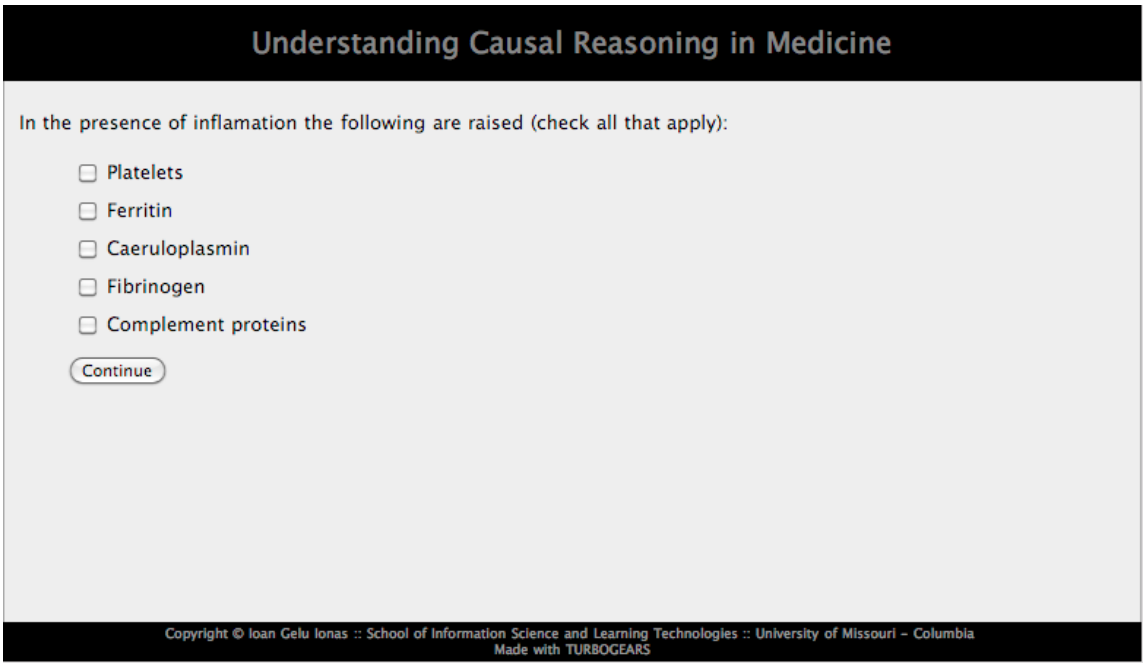
	1	2	3	4	5	6	7	8	9	10	
Nothing at all	○	○	○	○	○	○	○	○	○	○	**I am an expert**

Immunology

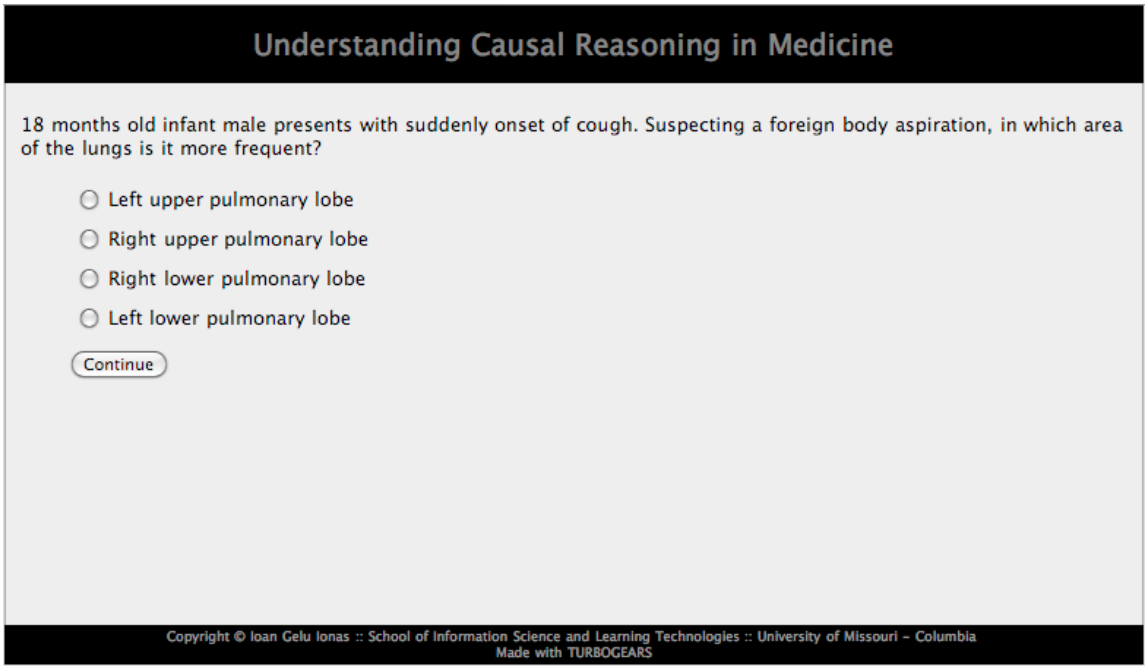
	1	2	3	4	5	6	7	8	9	10	
Nothing at all	○	○	○	○	○	○	○	○	○	○	**I am an expert**

Continue

Self-reported prior knowledge



Prior knowledge measurement – Example of all that apply multiple-choice question



Prior knowledge measurement – Example of all that apply multiple-choice question

Understanding Causal Reasoning in Medicine

A 10-year-old child presents with a persistent sinus infection. Over the course of the last 5 years, the patient has had several dozen sinus and upper respiratory infections. Blood tests reveal abnormally low levels of IgA immunoglobulin, but normal levels of other isotypes.

This patient's recurring infections are most likely due to a defect in which of the following?

- ○ Fixation of complement
- ○ Mast cell activation
- ○ Mucosal immunity
- ○ Neutrophil activation
- ○ Opsonization of bacteria

Continue

Example of practice question – Control group (NO explanation required)

Understanding Causal Reasoning in Medicine

A 10-year-old child presents with a persistent sinus infection. Over the course of the last 5 years, the patient has had several dozen sinus and upper respiratory infections. Blood tests reveal abnormally low levels of IgA immunoglobulin, but normal levels of other isotypes.

What mechanisms are most probable to be at play in this case?

This patient's recurring infections are most likely due to a defect in which of the following?

- ○ Fixation of complement
- ○ Mast cell activation
- ○ Mucosal immunity
- ○ Neutrophil activation
- ○ Opsonization of bacteria

Continue

Example of practice question – Treatment group (explanation required)

Understanding Causal Reasoning in Medicine

A 3-year-old male, known with seasonal allergies, presents at the clinic with difficulty breathing - increased respiratory rate, wheezing, and cough. Symptoms started two days before. No fever. Parents noted this morning a fine rash appearing on his abdomen. The rash seems to be getting worse.

Which of the following is the most probable diagnostic?

- ○ Anaphylaxis
- ○ Asthma exacerbation
- ○ Pneumonia
- ○ Foreign body aspiration
- ○ Roseola

Continue

Performance measurement question

Understanding Causal Reasoning in Medicine

A 3-year-old male, known with seasonal allergies, presents at the clinic with difficulty breathing - increased respiratory rate, wheezing, and cough. Symptoms started two days before. No fever. Parents noted this morning a fine rash appearing on his abdomen. The rash seems to be getting worse.

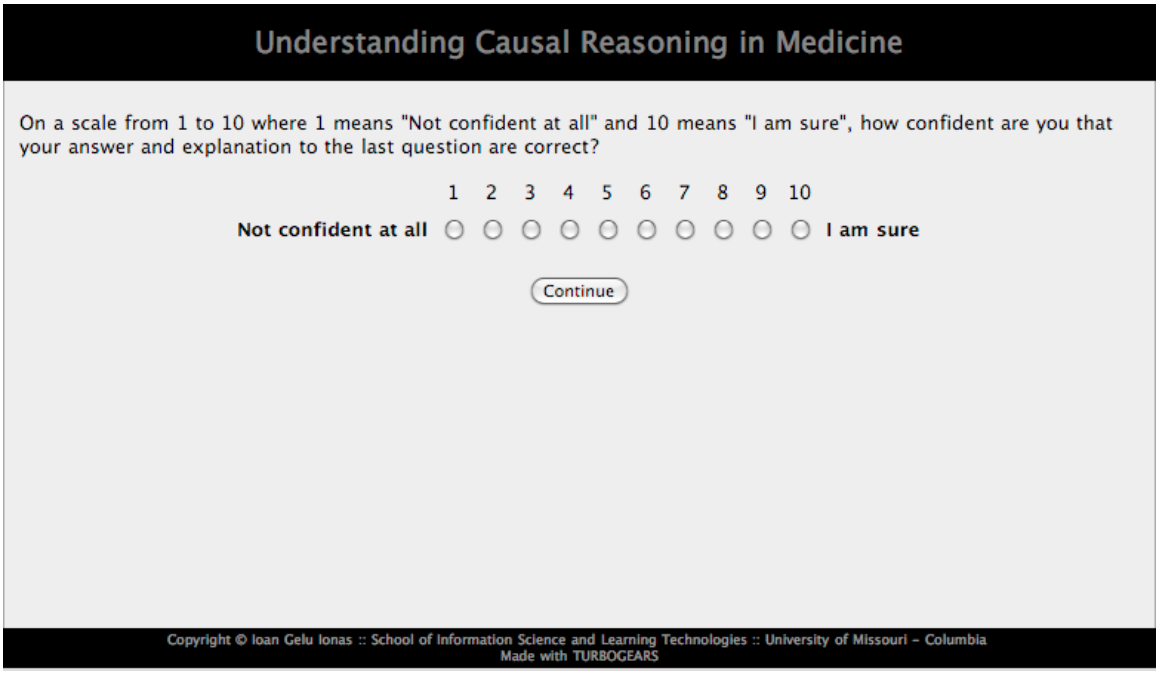
To the question: Which of the following is the most probable diagnostic?

You did not choose an answer to this question!. If, after answering the question below you want to choose an anser, you will have this chance later.

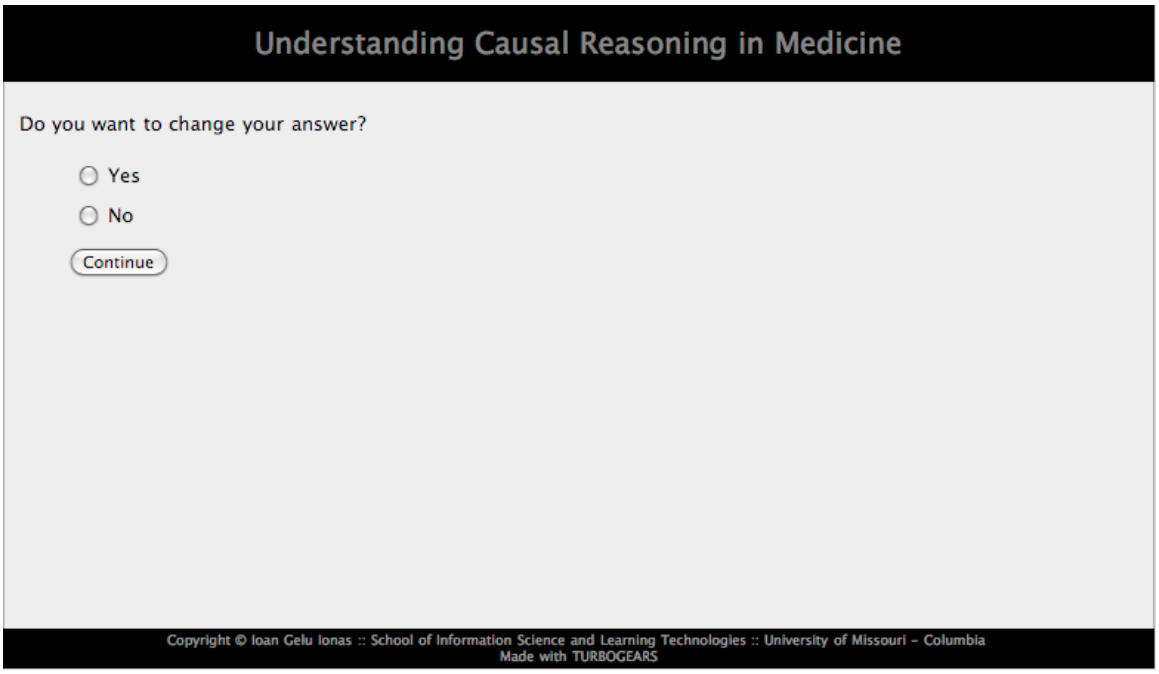
What **mechanisms** suggest your diagnostic? Please explain.

Continue

Mechanism explanation required for the performance measurement question



Confidence in the correctness of own answer



All participants are offered the possibility to change the answer

Understanding Causal Reasoning in Medicine

A 3-year-old male, known with seasonal allergies, presents at the clinic with difficulty breathing - increased respiratory rate, wheezing, and cough. Symptoms started two days before. No fever. Parents noted this morning a fine rash appearing on his abdomen. The rash seems to be getting worse.

Which of the following is the most probable diagnostic?

Your previously chose: **Pneumonia**

Your explanation was: **adfadf**

Your new answer is:

- ○ Anaphylaxis
- ○ Asthma exacerbation
- ○ Pneumonia
- ○ Foreign body aspiration
- ○ Roseola

Why did you change your answer and why and how is the new answer better than the first one? Please use the field below to explain.

Continue

Answer change page – participants are asked to provide an explanation for the change

Understanding Causal Reasoning in Medicine

In which age group are you?

- ○ <21
- ○ 21-25
- ○ 26-30
- ○ 31-35
- ○ 36-40
- ○ >40

Are you a Male or a Female?

- ○ Male
- ○ Female

In which income group would you or your family be?

- ○ < $20K
- ○ $20K-$40K
- ○ $40K-$80K
- ○ $80K-$150K
- ○ >$150K

Which medical specialty do you intend to pursue?

Continue

Demographic data collection

Understanding Causal Reasoning in Medicine

Thank you for participating in this study. If you want to learn more about its results, please contact Gelu Ionas.

If you want to review your answers the 11 questions included in this study, click the "Review" button, otherwise, click "Continue".

Review Continue

Offer to review the answers

Understanding Causal Reasoning in Medicine

Review

The answers are marked with (T) if the answer is correct or with (F) if it is not. The answer(s) you selected are colored in red.

The effect of sympathetic nervous system include (check all that apply):

1. (F) Contraction of the bladder detrusor muscle
2. (T) Papillary dilation
3. (T) Reduced gastrointestinal motility
4. (F) Constricts bronchiole smooth muscle
5. (T) Constricts skin and mucous membrane blood vessels

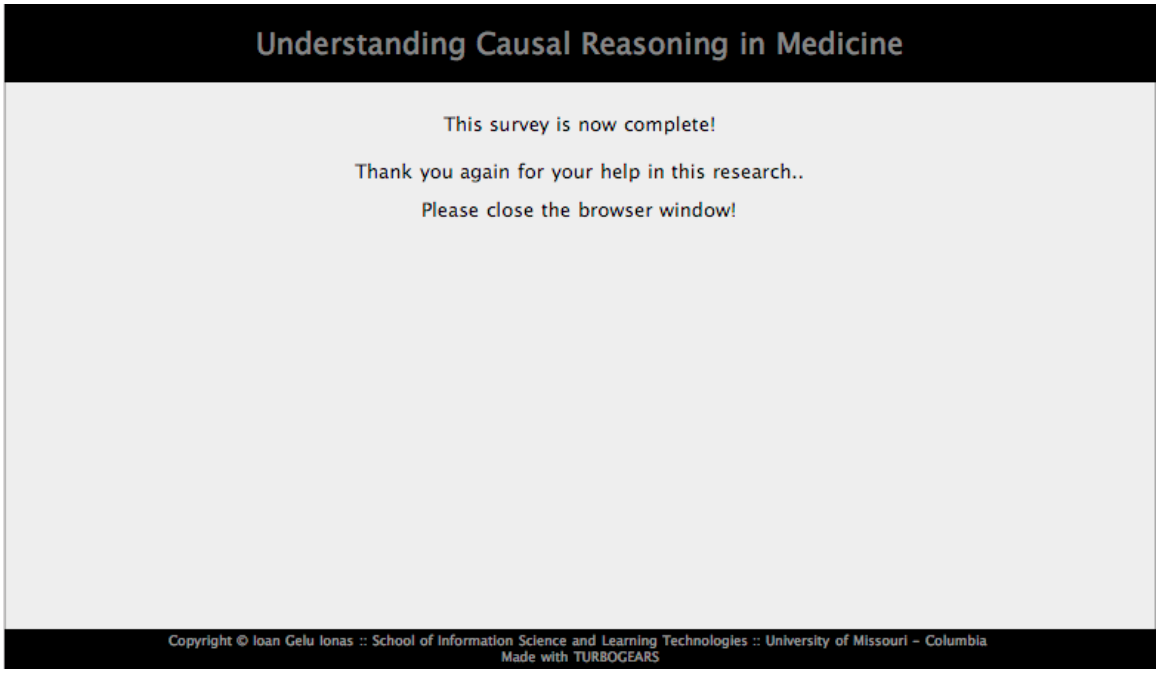
A 38 weeks premature infant receives surfactant via endotracheal tube. Which of the following would you expect?

1. (T) Decrease in alveolar surface tension and increase in lung compliance
2. (F) Increased alveolar surface tension and decreased lung compliance
3. (F) Increased alveolar surface tension and no change in lung compliance
4. (F) Decrease in alveolar surface tension and decrease in lung compliance

In the presence of inflamation the following are raised (check all that apply):

1. (T) Platelets
2. (T) Ferritin
3. (T) Caeruloplasmin
4. (T) Fibrinogen
5. (T) Complement proteins

Answer review page (10 questions = 6 for prior knowledge, 3 for practice, and 1 for performance measurement.



Experiment completion message.

BIBLIOGRAPHY

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*, 587-605.

Ahn, W.-K., & Baileson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31*, 82-123.

Ahn, W.-K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and Cognition* (pp. 199-225). Cambridge, MA: MIT Press.

Ahn, W.-K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*(3), 299-352.

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction, 17*(3), 286-303.

Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*(2), 147-179.

Anat, Z., & Pinchas, T. (1991). Assessing students' difficulties in causal reasoning in biology - a diagnostic instrument. *Journal of Biology Education, 25*(4), 302-308.

Anaya, G. (1999). College impact on student learning: Comparing the use fo self-reported gains, standardized test scores, and college grades. *Research in Higher Education, 40*(5), 499-526.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale NJ: Erlbaum.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out examples. *Journal of Educational Psychology, 95*(4), 774-783.

Axelrod, R. (Ed.). (1976). *The structure of decision: Cognitive maps of political elites.* Princeton, NJ: Princeton University Press.

Barbas, A., & Psillos, D. (1997). Causal Reasoning as a Base for Advancing a Systemic Approach to Simple Electrical Circuits. *Research in Science and Education, 27*(3), 445-459.

Beasly, N. E. (1968). The extent of individual differences perception of causality. *Canadian Journal of Psychology, 22*, 399-407.

Beauchamp, T. L. (1974). *Philosophical Problems of Causation*. Encion, CA: Dickenson.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History & Philosophy of Biological & Biomedical Sciences, 36*(2), 421-441.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research.* Princeton: Princeton University Press.

Bell-Basca, B., & Grotzer, T. A. (2001). *Focusing on the Nature of Causality in a Unit of Pressure: How Does it Affect Student Understanding?* (No. ED456029): National Science Foundation.

Bell-Basca, B., Grotzer, T. A., Donis, K., & Shaw, S. (2000). *Using domino and relational causality to analyze ecosystems: realizing that what goes around comes around.* Paper presented at the Annual Meeting of the National Association of Research in Science Teaching.

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science, 37*(4), 345-363.

Besson, U. (2004). Some features of causal reasoning: common sense and physics teaching. *Research in Science & Technological Education, 22*(1), 113-125.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findigns. *Higher Education, 18*(5), 529-549.

Boyle, D. G. (1960). A contribution to the study of phenomenal causality. *Quarterly Journal of Experimental Psychology, 12*, 171-179.

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (2000). Explanation in scientists and children. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and congition* (pp. 279-289). Cambridge, MA: MIT Press.

Budesheim, T. L. (1998). The Use of Abstract Trait Knowledge and Behavioral Exemplars in Causal Explanation of Behavior. *Personality and Social Psychology Bulletin, 24*(6), 575-587.

Buehner, M. J., & Cheng, P. W. (2005). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 143-168). Cambridge, MA: Cambridge University Press.

Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking and Reasoning, 6*(4), 269-295.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgment of causality: Effects of prior knowledge, experience, and reinforcement procedure. *The Quarterly Journal of Experimental Psychology, 56A*(5), 865-890.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking and Reasoning, 12*(4), 353-378.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209-254). New York, NY: Academic Press.

Bunge, M. (2004). How Does It Work? The Search for Explanatory Mechanisms. *Philosophy of the Social Sciences, 32*(2), 182-210.

Carey, S. (1995). On the origin of causal understanding. In D. Sperber, D. Premack & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 268-302). Oxford: Clarendon Press.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271-280.

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

Cheng, P. W., & Nisbett, R. E. (1993). Pragmatic Constraints on Causal Deduction. In R. E. Nisbett (Ed.), *Rules for Reasoning* (pp. 207-227). Hillsdale: Lawrence Erlbaum Associates.

Cheng, P. W., & Novick, L. R. (1990). A Probabilistic Contrast Model of Causal Induction. *Journal of Personality and Social Psychology, 58*(4), 545-567.

Cheng, P. W., & Novick, L. R. (1992). Covariation in Natural Causal Induction. *Psychological Review, 99*(2), 365-382.

Cheng, P. W., & Novick, L. R. (2005). Constraints and Nonconstraints in Causal Learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review, 112*(3), 694-707.

Chi, M. T. H. (2000a). *Misunderstanding emergent processes as causal.* Paper presented at the Annual Conference of the American Eductional Research Association (AERA), New Orleans, LA.

Chi, M. T. H. (2000b). Self-explaining expository texts: The dual process of generating inferences and repairing mengtal models. . In R. Glaser (Ed.), *Advances in instructional psychology.* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self explanations. . In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 251-282). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Bassok, M., Lewis, M., Reinmann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improve understanding. *Cognitive Science, 18*(3), 439-477.

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientifc texts. *Journal of the Learning Sciences, 6*(4), 347-365.

Conati, C., & VanLehn, K. (2000). Toward Computer-Based of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education, 11*, 389-415.

Corrigan, R., & Denton, P. (1996). Causal understanding as a developmental primitive. *Developmental Review, 16*, 162-202.

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education, 49*(3), 809-821.

Dahlgren, L. O. (1997). Experience of Learning. In F. Marton, D. Hounnsell & N. Entwistle (Eds.), (2 ed., pp. 23-38): Scottish Academic.

Darden, L. (2002). Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. *Philosophy of Science, 69*, S354-S365.

De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology, 32*(2), 188-205.

DeLeeuw, N., & Chi, M. T. H. (2003). Self-explanation: Enriching a situation model or repairing a domain model? . In G. Sinatra & P. Pintrich (Eds.), *International conceptual change.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dickinson, A. (2001). The 28th Bartlett Memorial Lecture. Causal learning: Association versus computation. *Current Directions in Psychological Science, 10*, 127-132.

Didierjean, A., & Cauzinille-Marmèche, E. (1997). Eliciting self explanations improves problem solving: What processes are involved? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16*, 325-351.

Driver, R., Guesne, E., & Tiberghien, A. (1985). *Cognition and Instruction* (Vol. 10).

Einhorn, H. J., & Hogarth, R. M. (1986). Judging Probable Cause. *Psychological Bulletin, 99*(1), 3-19.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395-430.

Fitzgerald, J. T., Gruppen, L. D., White, C. B., & Davis, W. K. (1997). *Medical student self-assessment abilities: Accuracy and calibration.* Paper presented at the Annual Meeting of the American Educational Research Association.

Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education, 37*, 645-649.

Fugeslang, J. A., & Thompson, V. A. (2000). Strategy selection in Causal Reasoning: When Beliefs and Covariation Collide. *Canadian Journal of Experimental Psychology, 54*(1), 14-32.

Fyrenius, A., Silen, C., & Wirell, S. (2007). Students' conceptions of underlying principles in medical physiology: an interview study of medical students' understanding in a PBL curriculum. *Advances In Physiology Education, 31*, 364-369.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49-71.

Glennan, S. (2002). Rethinking Mechanism Explanation. *Philosophy of Science, 69*, S342-S353.

Glennan, S. (2005). Modeling Mechanisms. *Studies in History & Philosophy of Biological & Biomedical Sciences, 36*(2), 443-464.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Glymour, C., & Cheng, P. (1998). Causal mechanisms and probability: a normative approach. In M. Oaksford & N. Charter (Eds.), *Rational models of cogniton*. Oxford: Oxford University Press.

Glymour, C., & Cheng, P. W. (1998). Rational models of cognition *Causal mechanisms and probability: A normative approach.* New York: Oxford University Press.

Goedert, K. M., & Spellman, B. A. (2005). Nonnormative discounting: There is more to cue interaction effects than controlling for alternative causes. *Learning & Behavior, 33*(2), 197-210.

Gopnik, A., Glymour, C., Sobel, D. M., Schultz, L., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review, 111*(1), 3-32.

Green, D. W. (1997). Explaining and envisaging an ecological phenomenon. *British Journal of Psychology, 88*, 199-217.

Griffith, T. L. (2005). *Causes, coincidences, and theories.*

Griffith, T. L., & Tenenbaum, J. B. (2005). Structure and Strength in Causal Induction. *Cognitive Psychology, 51*(4), 334-384.

Grotzer, T. A. (2000). *How Conceptual Leaps in Understanding the Nature of Causality Can Limit Learning: An Example from Electrical Circuits* (No. ED441699). Arlington, VA: National Science Foundation.

Grotzer, T. A. (2003). Transferring structural knowledge about the nature of causality: An empirical test of three levels of transfer., *The National Association of Research in Science Teaching (NARST) Conference*. Philadelphia, PA.

Grotzer, T. A., & Bell, B. (1999). Negotiating the funnel: Guiding students toward understanding elusive generative concepts. In L. Hetland & S. Veenema (Eds.), *The Project Zero Classroom: Views on Understanding.*: Fellows and Trustees of Harvard College.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition, 30*(7), 1128-1137.

Halloun, I., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53*(11), 1056-1065.

Harré, R. (1985). *The philosophies of science.* (2nd ed.). Oxford: Oxford University Press.

Harre, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity.* Oxford: Basil Blackwelll.

Hastie, R. (1984). Causes and Effects of Causal Attribution. *Journal of Personality and Social Psychology, 46*(1), 44-56.

Hausmann, R. G. M., VanLehn, K., Nokes, T. J., & Gershman, S. (2009, July 29 - August 1). *The design of self-explanation prompts: The fit hypothesis.* Paper presented at the 31st Annual Meeting of the Cognitive Sciences Society, Amsterdam, The Netherlands.

Hegarty, R. A., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*, 18-32.

Hilbert, T. S., Schworm, S., & Renkl, A. (2004). Learning from worked-out examples: The transition from instructional explanations to self-explanation prompts. In P. Gerjets, R. Elen, R. Joiner & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning.* (pp. 184-192). Tubingen: Knowledge Media Research Center.

Hobbs, J. (2001). *Causality.* Paper presented at the Common Sense 2001, Fifth Symposium on Logical Formalizations of Commonsense Reasoning., New York University, New York.

Holland, J., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge MA: MIT Press.

Huff, A., & Jenkins, H. M. (2002). *Mapping strategic knowledge*. London: Sage.

Hume, D. (1987). *A treatise of human nature* (2nd ed.). Oxford: Clarendon Press (Original work published 1739).

Hung, W. (2004). *An investigation of the role of causal reasoning methods in facilitating conceptual understanding of college students in physics.*

Irani, T. (2006). Teaching the Critical Thinking Skill of Explanation. *The Agricultural Education Magazine, 78*(6), 21-23.

Jaspars, J. M. F. (1988). Mental models of causal reasoning. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 335-358). Cambridge, MA: Cambridge University Press.

Jaspars, J. M. F., Hewstone, M. R. C., & Fincham, F. D. (1983). Attribution theory and research: The state of the art. In J. M. F. Jaspars, F. D. Fincham & M. R. C. Hewstone (Eds.), *Attribution theory: Essays and experiments.* (pp. 3-36). Orlando, FL: Academic Press.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes *Psychological Monographs: General and Applied* (Vol. 79, pp. 1-17).

Jonassen, D. H., & Ionas, I. G. (2006). Designing effective support for causal reasoning. *Educational Technology Research and Development, Online First*.

Kastens, K. A., & Liben, L. S. (2007). Eliciting Self-Explanations Improves Children's Performance on a Field-Based Map Skills Task. *Cognition and Instruction, 25*(1), 45-74.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192-238). Lincoln: University of Nebraska Press.

Kelley, H. H. (1973). The Process of Causal Attribution. *American Psychologist, 28*, 107-128.

Keppel, G. (1991). *Design and analysis: a researcher's handbook.* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist, 49*, 294-303.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Kitcher, P. P. (1989). Explanatory unification and the causal structure of the world. In P. P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science* (Vol. 13, pp. 410-505). Minneapolis: University of Minnesota Press.

Kushnir, T., & Gopnik, A. (2007). Conditional Probability Versus Spatial Contiguity in Causal Learning: Preschoolers Use New Contingency Evidence to Overcome Prior Spatial Assumptions. *Developmental Psychology, 43*(1), 186-196.

Lagnado, D. A., Waldman, M., Hagmayer, Y., & Sloman, S. (2007). Byond covariation: Cues to causal structure. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

Lin, X., & Lehmann, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*(7), 837-858.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Science, 10*, 464-470.

Lombrozo, T. (2007). Simplicity and Probability in Causal Explanation. *Cognitive Psychology, 55*(3), 232-257.

Lucangeli, D., Tressoldi, P. E., & Cendron, M. (1998). Cognitive and metacognitive abilities involved in the solution of mathematical word problems: Validation of a comprehensive model. *Contemporary Educational Psychology, 23*, 257-275.

Machamer, P. (2002). Activities and causation: the methaphysics and epistemology of mechanisms.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.

Maldonado, A., Catena, A., Perales, J. C., & Candido, A. (2007). Cognitive Biases in Human Causal Learning. *The Spanish Journal of Psychology, 10*(2), 242-250.

Mayes, R. G. (2005). Theories of Explanation. *The Internet Encyclopedia of Philosophy* Retrieved May 16, 2008, from http://www.iep.utm.edu/e/explanat.htm

Mazlack, L. J. (2007). *General imprecise causal reasoning representations.* Paper presented at the Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics.

Michael, J. (2007). What makes physiology hard for students to learn? Results of a faculty survey. *Advances In Physiology Education, 31*, 34-40.

Michotte, A. (1946). *La perception de la causalité*. Paris: Vrin.

Mitchell, N. J., Mertz, K., & Ryan, R. (1994). *Learning through self-explanation of mathematics examples: Effects of cognitive load.* Paper presented at the Annual Meeting of the American Educational Research Association.

Mooney Marini, M., & Singer, B. (1988). Causality in the Social Sciences. *Sociological Methodology, 18*, 347-409.

Nathan, M. J., Mertz, K., & Ryan, R. (1994). *Learnign through Self-Explanation of Mathematics Examples: Effects of Cognitive Load.* Paper presented at the Annual Meeting of the American Educational Research Association.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Cognitive Science in Medicine. In D. A. Evans & V. L. Patel (Eds.), (pp. 53-111). London: Massachusetts Institute of Technology.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669-710.

Pearl, J. (1998). Graphical Models for Probabilistic and Causal Reasoning. In D. M. Gabbay & P. Smets (Eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision* (pp. 367-389). Dordrecht: Kluwer Academic Publishers.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research. Explanation and prediction.* (3rd ed.): Thomson Learning.

Pennington, N., & Hasties, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 521-533.

Perkins, D. N., & Grotzer, T. A. (2000). *Models and Moves: Focusing on Dimensions of Causal Complexity to Achieve Deeper Scientific Understanding* (No. ED441698). Arlington, VA: National Science Foundation.

Pirolli, P. L., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12*, 235-275.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science, 67*, 1-25.

Rapus, T. L. (2004). *Integrating information about mechanism and covariation in causal reasoning.*

Read, S. J. (1983). Once Is Enough: Causal Reasoning From a Single Instance. *Journal of Personality and Social Psychology, 45*(2), 323-334.

Reed, P. (1992). Effect of a signaled delay between an action and outcome on human judgment of causality. *Quarterly Journal of Experimental Psychology, 44B*(2), 81-100.

Reed, P. (1999). Role of a stimulus filling an action-outcome delay in human judgments of causal effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes, 25*(1), 92-102.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27*(5), 709-748.

Rehder, B. (2003b). A Causal-Model Theory of Conceptual Representation and Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1141-1159.

Reinmann, P., & Neubert, C. (2000). The role of self-explanation in learning to use a spreadsheet through examples. *Journal of Computer Assisted Learning, 16*(4), 316-325.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1-29.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529-556.

Rittle-Johnson, B. (2006). Prompting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*(1), 1-15.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton, NJ: Princeton University Press.

Salmon, W. C. (1989). Four decades of scientific explanation. In P. P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science* (Vol. 13, pp. 3-219). Minneapolis: University of Minnesota Press.

Schlottmann, A. (2001). Perception Versus Knowledge of Cause and Effect in Children: When Seeing is Believing. *Current Directions in Psychological Science, 10*(4), 111-115.

Schultz, L. E., & Griffith, T. L. (2007). Can Being Scared Cause Tummy Aches? Naive Theories, Ambiguous Evidence, and Preschoolers' Causal Inferences. *Developmental Psychology, 43*(5), 1124-1139.

Schultz, T. R. (1982). Causal reasoning in the social and nonsocial realm. *Canadian Journal of Behavioral Science, 14*(4), 307-322.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of Evidence in Causal Inference. *Journal of Experimental Psychology: General, 110*(1), 101-120.

Shachter, R. D. (1988). Probabilistic inference and influence diagrams. *Operations Research, 36*(4), 589-604.

Shank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Shanks, D. R., Medin, D. L., & Holyoak, K. J. (1996). *The psychology of learning and motivation* (Vol. 34 Causal learning). San Diego, CA: Academic Press.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *Quarterly Journal of Experimental Psychology, Section B, 41*, 139-159.

Shipstone, D. (1985). Electricity in simple circuits. In R. Driver, E. Guesne & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 33-51). Philadelphia, PA: Open University Press.

Slotta, J., & Chi, M. (1999). *Overcoming robust misconceptions through ontology training.* Paper presented at the The American Educational Research Association Annual Meeting.

Slusher, M. P., & Anderson, C. A. (1996). Using Causal Persuasie Arguments to Change Beliefs and Teach New Information: The Mediating Role of Explanation Availability and Evaluation Bias in the Acceptance of Knowledge. *Journal of Educational Psychology, 88*(1), 110-122.

Spelke, E., Philips, A., & Woodward, A. (1996). Infant's knowledge of object motion and human action. In D. Sperber, D. Premack & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate.* Oxford: Clarendon Press.

Sperber, D., Premack, D., & Premack, A. J. (Eds.). (1995). *Causal cognition: A multidisciplinary debate*. Oxford, UK: Oxford University Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal netwoks from observations and interventions. *Cognitive Science, 27*, 453-489.

Strobel, J., Jonassen, D. H., & Ionas, I. G. (2008). The evolution of a collaborative authoring system for a non-linear hypertext: A design-based research study. *Computers & Education, 51*(1), 67-85.

Tangen, J. M., & Allan, L. G. (2004). Cue interaction and judgments of causality: Contributions of causal and associative processes. *Memory & Cognition, 32*(1), 107-124.

Tenenbaum, H. R., Alfieri, L., Brooks, P. J., & Dunne, G. (2008). The effects of explanatory conversations on children's emotion understanding. *British Journal of Developmental Psychology, 26*(2), 249-263.

Thagard, P. (1998). Explaining disease: Correlations, causes, and mechanisms. *Minds and Machines, 8*, 61-78.

Thagard, P. (2000). Probabilistic networks and explanatory coherence. *Cognitive Science Quarterly, 1*(1), 91-114.

Tversky, A., & Cahneman, D. (1989). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, NJ: Erlbaum.

Van Fraassen, B. C. (1985). Salmon on Explanation. *The Journal of Philosophy, 82*(11), 639-651.

VanLehn, K., & Jones, R. M. (1993). *What mediates the self-explanation effect? Knowledge gaps, schemas or analogies?* Paper presented at the Fifteenth Annual Conference of the Cognitive Science Society.

VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of self-explanation effect. *Journal of the Learning Sciences, 2*(1), 1-59.

Waldmann, M. R. (1996). Knowledge-Based Causal Induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *Causal Learning, The Psychology of learning and motivation* (Vol. 34, pp. 47-88). Sand Diego: Academic Press.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*(1), 27-58.

Waldmann, M. R., & Hagmayer, Y. (2006). Categories and Causality: The Neglected Direction. *Cognitive Psychology, 53*(1), 27-58.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon, D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception and language.* (pp. 523-573). New York, NY: Wiley.

White, B. Y., & Frederiksen, J. R. (1995). *An overview of the thinker tools inquiry project* (No. Causal Models Research Group Report No. CM-95-04). Berkeley, CA: School of Education, University of California at Berkeley.

White, P. A. (1989). The theory of causal processing. *British Journal of Psychology, 80*, 431-454.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to make sense of the world. *Journal of Science Education and Technology, 8*(1), 3-19.

Wong, R. M. F., & Lawson, M. J. (2002). The efect of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction, 12*(2), 233-262.

Yeh, Y.-F., Chen, M.-C., Hung, P.-H., & Hwang, G.-J. (2009). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers & Education, In press*.

Young, M. E., & Rogers, E. T. a. B., Joshua S. (2005). Causal impressions: Predicting when, not just whether. *Memory & Cognition, 33*(2), 320-331.

VITA

Ioan Gelu Ionas is originally from Romania where he received his bachelor's degree in Mechanical Engineering at the Technical University of Cluj-Napoca in 1990. He holds a PhD degree in Management (2000) from the "Lucian Blaga" University of Sibiu, Romania, and a Masters in Business Administration degree (2003) from the University of Missouri – Columbia. In 1990 Gelu began working as a faculty member for the School of Engineering at the "Lucian Blaga" University of Sibiu where he taught courses, conducted research, and published in mechanical engineering, management, and consumer behavior until 2001.

In August 2001 Gelu returned to the University of Missouri – Columbia to pursue and MBA degree and later a PhD in Information Science and Learning Technologies from which he graduated in May 2010. During his studies he participated in various research and development projects and contributed to research papers published or presented at national or international conferences. His areas of interest are causality and causal reasoning and design and development of learning and research web-based environments.