CLARIFYING METACOGNITION THROUGH COMPUTATIONAL MODELLING

by

Brendan Conway-Smith

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

in

Cognitive Science

Carleton University Ottawa, Canada

©2019, Brendan Conway-Smith

Abstract

This thesis presents a novel method of modelling metacognition computationally. Metacognition is commonly described as cognition acting on itself, and is correlated with enhanced performance in memory, reasoning, emotional regulation, and motor skills. How it attains these effects remains unclear. Understanding the mechanisms of metacognition requires surmounting two barrriers: the subject's high-level abstraction and disputed terminology. To overcome these obstacle, and to clarify the workings of metacognition, and how they come to act on themselves. Well-defined computational units are built upon to form increasing complex metacognitive processes. These computational forms of metacognition are then connected to the research literature. Finally, each form of metacognition is built into working models within the cognitive architecture ACT-R. These working models serve as an existence proof of the models' viability and functionality. The intention of this thesis is to help clarify the nature of metacognition, its underlying mechanisms, and its implications for advancing a unified theory of metacognition.

Acknowledgements

I owe a debt of gratitude to my advisor, Robert West for his depth of knowledge, discussion, and patience. I am thankful to have the opportunity to work with Myrto Mylopoulos, whose support and insight I greatly benefit from. I am also grateful to Andrew Brook for his wonderful philosophical guidance. Finally, my colleagues and lab mates have been a great source of knowledge and conversation. I am deeply thankful to have found such excellent support at Carleton University and the Cognitive Modelling Lab.

TABLE OF CONTENTS

Abstract
Acknowledgements ii
Table of Contents iii
List of Tables iv
List of Figures v
List of Appendices vi
1.0 INTRODUCTION
1.1 Background
1.2 Importance to autonomous systems
1.3 Methodology 10
2.0 LITERATURE REVIEW
2.0 Introduction
2.1 Historical background
2.2 Cognitive psychology research
2.2.1 Major components of metacognition
2.2.2 Cognitive knowledge
2.2.3 Cognitive regulation
2.3 Philosophical research
2.3.1 Definitions and epistemology
2.3.2 Attributivism versus evaluativism
2.3.3 Cognitive referent
2.4 Neuroscience research in metacognition
3.0 METACOGNITIVE INSTRUCTIONS

CLARIFYING METACOGNITION

4.0	MODEI	LLING METACOGNITION
	4.0 I	ntroduction
	4.1 (Computational Theory of Mind
	4.2 0	Cognitive modeling as a methodological tool for science
	4.3 (Cognitive modelling
	4.4 U	Unification
	4.5 A	ACT-R
	4.6 A	A production system as computational foundation
	4.7 A	Axiomatic system
5.0	FORMS	OF METACOGNITION
	5.0 I	Forms 1-8
	5.1 (Cognitive referent and intentionality
	5.2 I	Discussion of metacognitive instructions
	5.3 (Connection to cognitive psychology terminology
	5.4 (Connection to philosophy
	5.5 (Connection neuroscience
6.0	METAC	COGNITIVE FEELINGS
	6.1. I	Feelings of knowing
	6.2 I	Representing feelings
	6.3 I	Feelings as Metadata 56
7.0	DISCUS	SSION & CONCLUSION
	7.0 0	General discussion
	7.1 F	Findings contextualized within the research literature
	7.2 P	Potential value of finding
	7.3 (Conclusion
8.0	ADDEN	DUM : ACT-R code for metacognitive models
9.0	REFERI	ENCES

LIST OF TABLES

Table 1.	Connection of terminology	. 51
Table 2.	Noetic feelings of knowing	. 54

LIST OF FIGURES

Figure 1.	Overview
Figure 2.	Chart of metacognitive terminology16
Figure 3.	Chart of metacognitive instructions
Figure 4.	Overview of ACT-R
Figure 5.	Overview of the Common Model
Figure 6.	Euclidean axioms
Figure 7.	Metacognitive mountain
Figure 8.	A hierarchical computational model of cognition and metacognition
Figure 9.	Venn diagram of architecture
Figure 10	Referents interacting with productions
Figure 11	Computational terminology connections
Figure 12	. Cognitive terminology connections
Figure 13	. Metadata matched with symbolic computation
Figure 14	. Metadata interacting with productions
Figure 15.	Summary of thesis results

LIST OF APPENDICES

Appendix A. ACT-R	code in GitHub: For	ns of Metacognition	 65
repending rice rice in		no or metaeogintion	 05

1.0 Introduction

This thesis presents a novel way of representing metacognition computationally and discusses its implications for advancing a unified theory of metacognition. Metacognition is most often described as cognition acting on itself (Flavell, 1976; Dunlosky & Metcalfe, 2009; Proust & Fortier, 2018). After four decades of research, metacognition has been correlated with greater cognitive performance in memory, reasoning, emotional regulation, and motor skills (Dunlosky & Bjork, 2008; Fletcher & Carruthers, 2012; Dobson, 2013; MacIntyre, et al. 2014). Yet *how* it attains these unique effects is unclear, as the mechanisms giving rise to metacognition remain undetermined (Valenzuela, 2019). The problem is that metacognition is nearly prohibitively abstract, and has so far remained a fuzzy concept (Scott, & Levy, 2013). Further, the metacognitive terminology used by researchers is disordered and often contested (Steinbach, 2008). Hence, these are the two main barrier to advancing research in metacognition— the subject's abstractness and its disputed terminology.

In this paper I argue that modeling metacognition computationally can help to clarify the literature as well as metacognition's underlying mechanisms. This method accords with John Anderson's (2014) principle that abstract problems become systematic when analyzed within a cognitive architecture. This thesis uses a computational cognitive architecture to define the base units of cognition and how they become "meta" — i.e.: how they come to act on themselves. This paper's major contribution is a novel system of using computational axioms to build increasingly complex cognitive processes. These "forms" of cognition possess the virtue of being well-defined and operationalized. Hence, using this system we overcome the twin barriers of metacognition's abstractness and imprecise terminology.

This work is composed of three main parts. *Part one* consists of a literature review of the major metacognition research in psychology, philosophy, and neuroscience. It maps out the major terminology, issues, and controversies in the respective fields. The review discusses the obstacles to advancing metacognition research. *Part two* presents the main contribution of this thesis — a novel computational system for modelling metacognition. First, the system's philosophical and scientific foundation is established. Second, the foundation is built upon by assembling increasingly complex "forms" of cognition. The metacognitive forms will then be connected back to the research literature. *Part three* contains running computational models of the metacognitive forms detailed in the previous section. Each form of metacognition is built as working computer code, written in the language of the cognitive architecture ACT-R. These working models serve as existence proofs of the system's viability and functionality. The ACT-R code is attached as an addendum connected to GitHub.

The intention of this thesis is to provide a clarifying framework for discussing the research literature on metacognition, to help better understand the underlying mechanisms, terminology, and empirically observed phenomena. Finally, the implications for an improved understanding of metacognition are discussed in terms of its usefulness for artificial systems, education, therapy, and directing future research.

1.1 Background

Aristotle famously called "thinking about thinking" the activity of the Gods (Metaphysics, 7 & 9). It was Aristotle's own meta-thinking that set down the principles of logic and empiricisms that would grow into the scientific method (Whitehead, 1967; Celikovski et al. 1987). This method of rational, deliberate thinking proved the key to gaining our present knowledge of the

universe (Matthews, 2009). Eventually, the ability of cognition to direct itself would attract the attention of scientific inquiry. Early researchers studying psychological phenomena unexpectedly noticed cognition directing its own processes (James, 1897; Piaget, 1950). Psychologist William James (1892) once remarked that the most powerful weapon humans had was "our ability to choose one thought over another." This capacity would later be coined "metacognition" (Flavell, 1976). Since then, decades of research have established metacognition's ability to improve memory, reasoning, emotional regulation, and was even found to be the number one predictor of learning success (Wang et al. 1990; Schraw 1998; Veenman, Prins & Elshout, 2002). Regarding education, at one time, Dewey (1933) noted that people learn more from reflecting on their experience than from the experiences themselves. Later research strongly correlates metacognition with students' overall academic success (Adey & Shayer, 1993; Butler & Winne, 1995; Kuhn & Pearsall, 1998). Individuals with greater metacognitive abilities are generally more successful in their cognitive tasks (Livingston, 1996). Those with poor metacognitive abilities tend to underperform (Dunning et al., 2003). Studies have illuminated metacognition's role in expert performance — the display of exceptional knowledge and skills (MacIntyre, 2014).

While research has converged on metacognition's role in improving cognitive performance, a major question remains. How does cognition "act on itself" to enhance its own operations? Answering this question will require a model of metacognition's underlying mechanisms. In pursuit of this, this thesis aims to help advance Schraw's (2000) call for a "unified theory of metacognition."

1.2 Importance to autonomous systems

The nature of metacognition is not only of growing interest to those studying human cognition. It is increasingly needed for autonomous computational systems that are becoming so

complex that no human can direct their internal operations. Robust artificially intelligence systems are requiring more reliable processes of internal self-control, i.e.: they are requiring a computational model of metacognition (Schmill et. al., 2008). This issue connects to the problem of explainable A.I., which is also said to necessitate a machine form of metacognition (Rosenfeld & Richardson, 2019). As autonomous systems interact more with humans they are increasingly required to anticipate human thinking in systems such as self-driving cars, artificial personal assistants, and robot attendants. To engage smoothly with humans, autonomous robots are said to require a "machine theory of mind" (Rabinowitz et al., 2018). A machine theory of mind must enable a machine to "introspect" — to represent others' as well as its own internal processes. This goal is considered essential for enabling human-machine teaming such as in medical procedures, transportation, disaster relief, and care for the sick or elderly. These complex humanmachine interactions require robots to represent humans' mental processes such as beliefs, goals, and desires (Scassellati, 2002). To resolve this problem, a crucial question must be answered how should we represent the mechanisms of mind for machines to compute, for those of humans as well as their own? The ability of machines to represent the operations of humans intelligence and of themselves is an important step toward building robust, explainable A.I. capable of flexible machine-human interactions. What is required is a model of cognition that includes the functions of metacognition.

1.3 Methodology

To arrive at a model of metacognition we employ the computational method of cognitive modeling. Cognitive modeling is a falsifiable methodology for studying cognition (Prezenski et al., 2017). As a scientific method, it hypothesizes and implements executable cognitive models

that are then compared to empirical data. Much has been written on the value of cognitive modeling as a tool for science (Simon & Wallach, 1999). Particularly, cognitive modeling can be used to describe, predict, and prescribe human behavior (Marewski & Link, 2014). This process is elaborated on in section 2.0. The paper asserts that computational terminology is ideal for developing models of human metacognition. The functions of computation forces definitional clarity. Hence, it can help operationalize what has so far been endlessly abstract. This is why a computational method is ideal for investigating metacognition to determine its inner mechanisms. Moreover, applying a computational lens to cognition is suitable when we consider that the field of cognitive science was born from an information-processing view of the mind.

To ground this method in human cognition we employ the Standard Model of Cognition (Laird, Lebiere, & Rosenbloom, 2017), now renamed the Common Model. The Common Model defines the mechanisms common to all computational architectures modelling human cognition. This paper employs the Common Model architecture as instantiated within ACT-R. Specifically, we use productions to represent procedural memory, chunks to represent declarative knowledge, and buffers to represent working memory. We first define the base units of cognition and then built upward toward processes that are distinctly metacognitive. Following this, we compare these models to the leading metacognitive theories within psychology, philosophy, and neuroscience. We intend for our results to produce a novel view of metacognition based on computational mechanisms rather than solely philosophical definitions. We expect this model to assist in clarifying metacognitive processes and integrating its terminology. This new philosophical method constructs metacognition hierarchically, founded upon a computational cognitive architecture and the present research literature.

LITERATURE REVIEW

2.0 Introduction

The following presents a review of some of the major literature in metacognition research, specifically within the fields of psychology, philosophy, and neuroscience. This review will be structured thematically, laying out the major terminology, issues, trends, and controversies within the respective fields. Finally, it will identify gaps within the literature and lay out a proposal for addressing them.

2.1 Historical Background

Throughout Western intellectual history, the notion of "thinking about thinking" traces its roots at least as far back as Plato and Aristotle (Georghiades, 2004; Colman, 2015). Plato had once claimed that "when the mind is thinking it is talking to itself" (Jowett, 2018). Aristotle believed that a divine intellect "understands understanding" and that "it is indeed the best thing" (Metaphysics, 9). Ancient philosophy referred to metacognition using largely implicit language, as it lacked modern, explicit terminology. Centuries would pass before formal terms would be developed by cognitive psychology — terms philosophy later adopted and advanced. By this circuitous route, philosophy obtained much of its present metacognitive vocabulary that it has since adapted in a unique way.

Modern scientific inquiry into metacognition was commenced by cognitive psychologist Piaget (1950) who first referred to "knowing about knowing," and "thinking about thinking." It was not until 1976 that the term "metacognition" was originated by psychologist John Flavell in his work on metamemory. The term "hypercognition" was coined by Demetriou in 1993, yet common usage would retain Flavell's prefix "meta," from the greek word meaning "beyond" (Metcalfe, 1994). It was Flavell's work that directed scientific focus to the vital role metacognition plays in directing cognitive processes.

2.2 Cognitive psychology research

The modern literature on metacognition was originated by studies within the field of cognitive psychology (Hart, 1965; Piaget, 1950; Peters, 2007; Steinbach, 2008). These terms were also developed by research in early social development psychology (Vygotsky, 1962; Garofalo & Lester, 1985; Martinez 2006). The most common definitions of metacognition entail possessing information about one's own cognitive structure and the capacity to intentionally direct those structures (Flavell, 1979; Jacobs & Paris, 1987; Schraw, 1994; Livingston, 1997; Dunlosky & Hertzog, 2000; Georghiades, 2004).

Psychologist John Flavell (1976) preformed much of the pioneering work and advanced an early definition of metacognition as "knowledge concerning one's own cognitive processes and products or anything related to them" (p.232). Flavell later refined this as "cognition about cognitive phenomena" (1979). Future definitions would not stray far from this central notion. Developmental psychology would later refer to metacognition as "The knowledge and control children have over their own thinking and learning activities" (Cross & Paris, 1988, p.131). While psychologists Kuhn & Dean (2004) referred to it as "awareness and management of one's own thought." Martinez (2006) considered it "The monitoring and control of thought."

While these definitions appear to converge, an analysis of the literature on metacognition reveals that the definitions that researchers and theorists employ differ widely. In a review of metacognitive terminology, Dinsmore, Alexander, and Loughlin (2008) concluded that, in the 255 studies reviewed, researchers provided explicit definition of metacognition only 32% of the

time. Complicating things further, Steinbach (2008) reported that even more terms have emerged in recent years and they are used interchangeably with metacognition. Researchers now use terms synonymous with metacognition including "self-management" (O'Neil & Speilberger, 1979), "meta-learning" (Fisher, 1998), and "metamentation" (Bogdan, 2000). At present, researchers largely employ the umbrella term "metacognition" to refer to an agent understanding and intentionally directing their own thinking processes (Leader, 2008). Particularly blurry are terms such as "self-regulation," and "executive function," as researchers tend to draw upon the theoretical distinctions in their fields. Even as definitions vary widely between fields, so too do they differ within then.

Presently, there is ongoing debate whether metacognition is "domain-general" or "domainspecific" — whether it is developed first within a domain and then applied generally (Keleman, Frost, & Weaver 2000), or if general skills are learned first and then applied to particular domains (Schraw & Nietfeld 1998; Pintrich 2002; Veenman et al., 2006). Current research now tends to support the view that metacognition is the combination of the two (Veenman & Verheij, 2003; Veenman, Wilhelm & Beishuizen, 2004; Fleming & Dolan, 2012).

As we will see, there is no sharp definitional distinctions between or within fields. Therefore, it is more precise to present the different ways individual researchers refer to various term. The following pages attempt to organize and display the many concepts used by researchers throughout various fields. Lacking a more suitable starting place, this review of metacognitive terminology commences with the historical beginning of the term "metacognition." While metacognition is as old as the human species, its explicit terminology came about only recently. The following is an overview of those early attempts to define and categorize metacognitive phenomena.

2.2.1 Major components of metacognition

To Flavell (1976), metacognitive phenomena exhibited four main classes: knowledge, strategies, experiences, and goals (p. 906). These can be further reduced to two main categories of metacognition: *knowledge* about cognition and *regulation* of cognition. Later researchers were in overwhelming agreement with this view (Cross & Paris, 1988; Paris & Winograd, 1990; Schraw & Moshman, 1995; Kuhn & Dean 2004; Schraw et al. 2006; Martinez, 2006; Veenman & Afflerbach, 2006; Whitebread et al., 2009). While there is near consensus regarding these two major categories, specific terminology varies within them. The following section attempts to categorize the various terms regarding cognitive knowledge.

2.2.2 Cognitive knowledge ("knowing")

Flavell (1979) defined cognitive knowledge as one's knowledge of their own cognitive strengths and limitations, including the factors that affect cognitive functioning. Flavell identified three types of metacognitive knowledge:

- (1) "person" knowledge knowledge about humans as cognitive processors
- (2) "*task*" *knowledge* knowledge about the demands of a task
- (3) "strategy" knowledge knowledge of useful strategies

Within the developmental approach, Jacobs and Paris (1987) referred to their own forms of cognitive knowledge — declarative, procedural, and conditional. On the other hand, within cognitive psychology, various terms concerning metacognition often refer to executive control, such as "self-regulating" and "self-monitoring" (Kuhn & Dean, 2004; McLeod, 1997; Schneider & Lockl, 2002). When these assorted terms are overlaid, commonalities begin to appear. Various terms referring to metacognitive knowledge are outlined in **Figure 2**.

COGNITIVE KNOWLEDGE ("knowing")

Knowledge of factors affecting cognition		
Person and task knowledge	Flavell, 1979	
Declarative knowledge	Cross & Paris, 1988; Schraw & Moshman, 1995; Schraw et al., 2006	
Self-appraisal	Paris & Winograd, 1990	
Epistemological understanding	Kuhn & Dean, 2004	
Knowledge of management and strategies		
Strategic knowledge	Flavell, 1979	
Procedural knowledge	Cross & Paris, 1988; Kuhn & Dean, 2004; Schraw et al., 2006	
Knowledge of when to use a strategy		
Conditional knowledge	Schraw, Crippen, & Hartley, 2006	

COGNITIVE REGULATION ("doing")

Awareness of comprehension and task performance			
Cognitive experience	Flavell, 1979; Moritz et al. 2018		
Monitoring and regulating	Cross & Paris, 1988; Paris & Winograd, 1990; Schraw et al., 2006; Martinez, 2006; Whitebread et al., 2009; Schraw & Moshman, 2018; Nelson & Narens, 1990.		
Evaluate processes, the product of learning, and revising goals			
Evaluating	Cross & Paris, 1988; Paris & Winograd, 1990; Schraw et al., 2006; Whitebread et al., 2009; Schraw & Moshman, 2018		
Identify and choose strategies			
Planning	Cross & Paris, 1988; Paris & Winograd, 1990; Schraw & Moshman, 1995; Schraw et al., 2006; Whitebread et al., 2000		

Figure 2. Chart of metacognitive terminology within cognitive psychology

2.2.3 Cognitive regulation ("doing")

John Flavell (1979) referred to cognitive regulation in terms of "cognitive experiences," which consisted of the thoughts and perceptions directed at the cognitive process itself. Flavell argued that "experiences" act as executive control over thinking. This executive control allows individuals to reevaluate their plans and goals. For instance, one may review a plan to see if it is working, or reassess a goal to see if it's worth pursuing. This term was expanded upon by later researchers. **Figure 2** presents the various ways researchers have used cognitive terminology.

Nelsons and Narens (1990) distinguish between two major relations in meta-memory, that of "control" and "monitoring." These terms describe the direction of the flow of information in metacognition:

Control: the meta-level controls the object-level through deliberate action

Monitoring: the object level informs the meta-level

Important to this paper is the special emphasis that Nelsons and Narens place on the *referent* of information, which distinguishes the cognitive process.

2.3 Philosophical research

The field of philosophy — particularly of mind and cognition — has adopted various metacognitive terms originated within the fields of psychology. Since adopting these, philosophy has evolved many unique perspectives, issues, and contentions. While too numerous to exhaust here, the following attempts to set out the major relevant philosophical positions and issues regarding metacognition. This will include definitional nuances, epistemological status, levels of metacognitive judgements, and the demarcation line between cognition and metacognition. Major areas of contention include "attributivist vs. evaluativist" accounts as well as the issue of

"domain-general versus domain-specific." Later, this terrain will be overlaid across the field of psychology, to reveal a path toward its conceptualization within a computational framework.

2.3.1 Definitions and epistemology

Within the field of philosophy, the term "metacognition" is used in different ways by various researchers. Here, definitions and epistemologies are uniquely intertwined and so will be discussed together. While psychologists are overwhelmingly empirical in their methodology, some philosophers argue for a form of methodological dualism — the claim that philosophy should have licence to employ a methodology autonomous from empirical science (McDowell, 1994). On the one hand, many thinkers hold that philosophical answers should be derived solely from philosophical investigation and deductive argument. On the other hand, there are "naturalists" who believe that scientific inquiry into cognition is impossible without empirical knowledge of cognitive processes (Proust, 2013). This naturalist view is connected to naturalized epistemology as championed by Quine (1969), who claimed that epistemology will eventually become a "chapter of psychology and hence of natural science".

Philosophy's special emphasis on epistemology has led to terminology that varies somewhat from psychology. For example, philosophy contains a unique operationalization of the term "metacognition." While *psychology* views metacognitive processes as fundamentally distinct from physical behavior, some researchers in *philosophy* perceives more similarity. In philosophy, intentional action can be both physical *and* cognitive. While intentional, both physical and cognitive actions may be considered deliberate "behavior." Hence, observable physical actions can be considered "first-order" behavior, while unobservable cognitive action can be considered "second-order" behavior. This unique use of the term "behavior' emerged from a particular

concern for empirical validity. The desire for empirically legitimacy has led some philosophers to diminish the definition of metacognition as "cognition about cognition," in place of operationalizing metacognition as "behaviour about behaviour" (Shallice & Burgess 1996; Carruthers, 2011; Smith & Couchman, 2012). Other researchers are less concerned with strict empiricism and distinguish metacognition by cognitive properties, such as a "meta-level representation of object-level cognition" (Fleming, Dolan, & Frith 2012). Of particular note is Joelle Proust's (2013) conception of metacognition as "the set of capacities through which an operating cognitive subsystem is evaluated or represented by another subsystem" (p. 13).

Fletcher and Carruthers (2012) make the distinction that metacognition is not needed for controlling simple cognitive actions, but necessary for complex cognitive processes such as planning and reasoning. In their view, complex reasoning involves "cognitive control strategies." These strategies are received through social learning. For instance, the Western tradition of formal reasoning was inherited from ancient Greece, passed down through social learning both verbal and written. This formal reasoning contains "cognitive control strategies" described by Fletcher and Carruthers. The metacognitive strategies of formal reason contain instructions to prioritize evidence and de-prioritize emotion while judging a statement's truth. In this sense, formal reasoning, and epistemology more generally, is the process of cognition representing and directing its operations *intentionally*. Science itself is a metacognitive enterprise. As Fletcher and Carruthers affirm, before the discovery of logic and science humans lacked "cognitive control strategies" that would otherwise enable their greater cognitive abilities. Soon after the appearance of rational, metacognitive strategies, humans demonstrated expanded abilities for advanced knowledge and technology. Dennett (2000) refers to these strategies as "mind tools."

Science as a metacognitive enterprise has been called the "hallmark intellectual achievement of the human species" (Feist, 2006).

In separate research, Carruthers (2009) sought to account for the human species' unique metacognitive faculty. Carruthers hypothesized that metacognition evolved from a repurposed area of the brain, specifically the "theory of mind" region. This theory of mind module is hypothesized to allow social animals to infer each others' intentions and emotions. Carruthers suggests that during metacognition, this "theory of mind" mechanism is redirected toward one's own cognition as a general "mind-reading system." Carruthers argues that this general "mind-reading system" enables the self-monitoring needed for complex reasoning. He further hypothesizes that this also accounts for the apparent demonstrations of metacognition in other non-human social animals. It is worth noting that metacognition is demonstrated most apparently in social animals of greater intelligence, which implies the necessity for high enough cognitive development and theory of mind module (Frith, 2012). This self mind-reading theory lands Carruthers squarely on the "attributivist" side of the metacognitive debate, discussed in the following section.

2.3.2 Attributivism versus evaluativism

A major area of controversy lies between two competing claims on metacognitive assessments, the "attributivist versus evaluativist" debate (Proust, 2014). Attributivism is also referred to as the "Self-Attributive View" (SAV). Presently, it is the dominant perspective, championed by major thinkers in the field (Flavell, 1979; Shallice, 1988; Leslie 1987; Dennett 1991; Perner, 1988; Nelson & Narens, 1992; Gopnik, 1993; Carruthers, 2008). Attributivism holds that metacognition necessarily involves modeling first-order cognitive states — requiring one to represent mental states *as* mental states. As a result, this requires a form of "mind reading" whereby one evaluates the reliability of one's own mental states. This form of *self* mind-reading is hence a "mental state self-attribution," and so is called "attributivism" (Carruthers, 2011).

The alternate yet potentially compatible view on metacognitive assessments is termed "evaluativism" (Proust, 2014) also referred to as the "Self-Evaluativist View" (SEV). The selfevaluativist view holds metacognition as the dynamic activity of monitoring and controlling mental processes. Joelle Proust in presently the main champion of this view — that metacognition is activity-dependent — and claims that evaluativism surpasses attributivism both empirically and conceptually. Evaluativism holds that two sources of information are available when predicting cognitive accuracy — emotional experience and concept possession. These experiences can involve "noetic feelings," such as the common feeling that we know a famous person's name while not being able to bring it to mind. Concept possession involves explicit knowledge about mental functions, such as the knowledge that one is competent to speak about a certain domain of expertise. Evaluativism contains a unique focus on transferring self-evaluative judgments to others. Hence we can infer from knowledge of our own minds what another person is thinking. This is a reversal of Carruthers' view of the repurposing of the theory of mind module. Another distinction in this debate is that evaluativism holds that metacognition does not always necessarily involve meta-representation. Conversely, attributivism maintains that metacognition necessitates the employment of meta-representation. Lastly, while evaluativism holds there is a necessary link between metacognition and mental agency, attributivism does not.

2.3.3 Cognitive referent

Within the field of philosophy, a major distinction is made in distinguishing the *referent* of metacognitive judgements. This is similar to Nelson and Narens' (1992) claim that the referent of

knowledge is what distinguishes the cognitive process. Metcalfe et al. (2012) famously identified three referent levels of metacognitive judgements:

Autonoetic – where the referent includes one's self
Noetic – where the referent involves mental representations
Anoetic – where the referent involves objects in the world

The role of *noetic* feelings in metacognition is an especially fast growing area of research — particularly the role of noetic feelings in epistemic decision-making. *Noetic* involves metacognitive judgements about performance, i.e.: how a student feel they did on the test, the feeling of understanding a sentence that was just read. Feelings of knowing play as central role in daily human experience. Research indicates that that noetic feelings regularly direct human actions (**Table 2**). For instance, we often rate our feeling of how much time has passed. We may also rate our cognitive effort ("how hard am I trying?"). Or, when departing the house, one may feel a sense they are "missing something," and so walk through the house to try to remember.

Autonoetic judgments involves metacognitive judgements about one's own agency. This form of judgement has been researched in schizophrenics who hold uniquely irrational judgements about their own agency, yet rational judgements about their performance on tasks. This implies that metacognition is not a unitary, domain-general phenomena, but instead possesses domainspecific boundaries. This leads further into the contented area of whether metacognition is domain-general (Cole et al., 2013) or domain-specific (Donoso et al., 2014). A computational view holds that metacognition requires *both* domain-general and domain-specific mechanisms to regulate cognitive behavior (Pouget et al., 2016; Fleming & Daw, 2017). Neuroimaging research also supports this view (Fleming & Lau, 2014). Presently, these are among the many questions that remain unresolved.

2.4 Neuroscience research in metacognition

The following outlines some of the important research in the neuroscience of metacognition and its implications for metacognitive modeling. While neuroimaging has advanced our understanding of many cognitive phenomena such as vision and memory, research into the neurosubstrates of metacognition is just beginning (Fleming, Dolan, & Firth, 2012). Preliminary neuroimaging reveals that a wide network of brain regions are active during metacognitive activity (Anderson et al., 2011; Wintermute et al., 2012; Seghier, 2013). However, there is a seeming convergence on particular regions of the brain, largely within the prefrontal cortex (Nelson & Narens, 1992; Daw, Niv, & Dayan, 2005; Anderson & Fincham, 2014; Lee et al., 2014; Stocco et at., 2018). Early neurological studies have indicated that damage to the prefrontal cortex impaired meta-memory (Nelson & Narens, 1992; Metcalfe, 1994). Selfrepresentation, the ability to reflect on how we feel and act, has been correlated with activity in the anterior medial prefrontal cortex (Gazzaniga, Ivry, & Mangun, 2013; Baird et al. 2013). Accurate judgement of performance has been correlated with the dorsolateral and anterior prefrontal subregions which interact with the interoceptive cortices (Fleming & Lau, 2014).

Of particular significance were fMRI results that correlated peak brain activity during metacognitive activity with the rostrolateral prefrontal cortex (Fleming & Dolan, 2012; Anderson & Fincham, 2014). A neuroimaging study by Anderson (2014) tracked subjects' brain activity as they went through 4 problem solving steps (encoding, planning, solving, and responding). For this study, Anderson inferred that metacognition depended on having declarative representations of one's own procedures in order to evaluate and change them. While significant data has accumulated throughout the past decade, much of the neuroscience literature remains controversial (Morales et al., 2018).

3.0 Metacognitive instructions

While the mechanisms underlying metacognition's functions remain unresolved, there is convergence on metacognition's effects, and benefits. This section presents a brief overview of the research concerning how metacognitive instructions can improve cognitive performance.

For thousands of years, strategies for improving memory have been recorded and passed down. The ancient Roman senator Cicero was known for having advocated the method of loci, which holds a technique for improving one's memory. Cicero asserted that one "should choose localities, then form mental images of things they want to store in their memory, and place these in the localities" (Cicero, 2001). Studies appear to affirm the neural substrates of metamemory techniques. Structural and functional brain imaging found that superior memory did not result from brain differences but rather from *techniques* that engaged brain regions such as the hippocampus (Geraci, Saenz, & Tirso, 2019). Since Flavell's groundbreaking work on metamemory in the 1970s, research has documented the benefits of metacognitive instructions in memory, education, reasoning, cognitive therapy, emotional regulation, sports, military training, and more (Dunlosky & Bjork, 2008; Fletcher & Carruthers, 2012; Dobson, 2013; MacIntyre, et al. 2014). **Figure 3** provides an brief overview of various categories of metacognitive instructions.

Findings suggest that metacognition not only improves cognitive skills but also perceptualmotor skills (Augustyn & Rosenbaum, 2005). Enhancing physical performance has prompted metacognitive skills to be incorporated into military training based on studies of Navy recruit training (Mathieu, Martineau, & Tannenbaum, 1993). Metacognitive factors have also been correlated with predicting greater business and entrepreneurial success, such as motivation, determination, intention, communication, and courage (Envick, 2014; Frese, & Gielnik, 2014; Makhbul & Hasun, 2011).

Metacognitive Instructions			
MEMORY	Learning (Schraw, Crippen, & Hartley, 2006)	Nemonic devices (Maguire et al., 2003)	Memory Palace (Spence, 1984)
REASONING	Science (Zion, Michalsky, & Mevarech, 2005)	Logic (Leevers & Harris, 1999)	Math (Kramarski, Mevarech, & Arami, 2002)
PHYSICAL	Sports (MacIntyre, et al., 2014)	Navy seal training (Wiltshire, Rosch, Fiorella, & Fiore, 2014)	Motor skills (Augustyn, & Rosenbaum, 2005)
EMOTIONAL	CBT (Dobson, 2013)	Self-regulation (Vukman, & Licardo, 2010)	Meditation (Baird, Mrazek, Phillips, & Schooler, 2014)
PROFESSIONAL	Business (Tempelaar, 2006)	Entrepreneurial (Haynie, Shepherd, & Patzelt 2015)	Medical (Pesut & Herman, 1992)

Figure 3. Chart of metacognitive instructions

There have been strong indices that metacognitive instruction and training can promote successful cognitive functioning overall (Osman & Hannafin, 1992). Subjects with metacognitive skills tend to be more effective due to their capacity to monitor their own progress, identify their limitations, and adapt their thinking (Ford, Smith, Weissbein, Gully, & Salas, 1998). Conversely, the renown "Dunning-Kruger effect" correlates a lack of success with one being both unskilled and unaware of what one doesn't know (Kruger & Dunning, 1999). Further, a lack of metacognitive ability was shown to feature in those with extremist beliefs. Radicals showed less insight into the correctness of their decisions, and less change in response to disconfirming evidence (Rollwage, Dolan, & Fleming, 2018).

In recent years, data have accumulated on metacognitive practices known as mindfulness meditation. Practices such as Vipassana meditation have shown to be an effective cognitive processing model. Vipassana (in the tradition of S. N. Goenka) has evinced a variety of cognitive benefits such as increasing executive control, reducing stress, anxiety, and depressive symptoms (Chambers, Lo, & Allen, 2008; Khoury, et. al., 2017). Meta-analysis have revealed their effectiveness in reducing clinical symptoms of psychiatric disorders such as chronic pain, smoking, and addictions (Goldberg et al., 2018). While meditation training is first provided as metacognitive instructions, long term benefits appear to result in neurophysiological changes. Mindfulness meditation has shown to enhance executive control (working memory, cognitive control, concentration) by strengthening connections within the dorsolateral prefrontal cortex (Taren et al., 2017; Lee, Kulubya, Golden et al., 2018). Meditation's enhancement of cognitive efficiency has prompt organizations such as Google and General Mills to incorporate mindfulness into employee training (Olano et al., 2015; Gelles, 2015).

Lastly, metacognitive instruction have been shown to improve reasoning and science learning, also enhancing motivation and future learning (Zepeda, Richey, Ronevich, & Nokes-Malach, 2015). These include critical thinking, problem-solving, generating and revising hypotheses, as well as understanding nuances of methodology and epistemology (Zion, Michalsky, & Mevarech, 2005).

MODELLING METACOGNITION

4.0 Introduction

I have shown in section one that the present state of metacognitive terminology is in disarray. Between the various fields and within them, terms remains unstandardized and nebulous. This disorder presents a barrier for researchers to collaborate, share knowledge, and

integrate their research. This problem impedes advancement toward a model of metacognition that can guide research, training, and implementation within autonomous systems. To help resolve this problem I propose a new method of applying a computational lens to

metacognitive research.

Computational terminology has a unique power to grasp slippery concepts. Computation forces clarity due to its functional necessity. Because of this it can help to pierce nebulous clouds of abstraction and is an ideal tool for investigating metacognition. In the following section I present a computational framework to act as a lens for studying the research literature. I intend to use this lens to peer into the nebula of metacognition to glimpse its inner workings.

"Computer science is no more about computers than astronomy is about telescopes." -Edsger Dijkstra

It can be argued that computational language is the mother tongue of cognitive science. This makes it is a sensible dialect for discussing metacognition. As such, computational language can potentially serve as a "lingua franca" for standardizing metacognitive terminology and integrating the research. Before presenting a computational framework I will first discuss the relevant conceptual background. This background includes computational theory of mind, cognitive modeling, and the cognitive architecture employed here. Following this I will describe this paper's novel method of building before constructing the framework itself. The framework contains the main contribution of this thesis. This framework will then be connected to the research literature reviewed previously. Lastly, I will present a speculative hypothesis of how metacognitive feelings can be represented computationally. To begin, I will discuss an important philosophical precursor to cognitive modelling.

4.1 Computational Theory of Mind

Before wading into computational modeling we take note of its philosophical forefather, Computational Theory of Mind (CTM). Originating as the Church-Turing thesis (Goldin & Wegner, 2005), CTM holds that the nature of mind is computational in nature. McCulloch and Pitts (1943) were the first to propose that neural computations are what produce intelligent behavior. This theory's modern form was developed by Putnam in 1961 and Fodor in the 1960s. Its current iteration states that the mind inputs data, stores it within internal memory, transforms information by rules, and produces intelligent outputs (Newell & Simon, 1976; Fodor & Pylyshyn, 1988; Gallistel & King 2009).

CTM views neurons as the hardware that produce the operations acting on representations. These processes can be depicted in various ways, a dominant method is within computer code. A coded model is intended to depict and/or predict intelligent behavior (Cooper, Fox, Farringdon & Shallice, 1996). As Fodor (1972) pointed out, accepting computational theory of mind is not essential for cognitive modeling. Rather, we may regard computer code as acting *comparably* to mechanisms within cognition. Coded models can serve to generate useful insights and yield predictive power (John, Vera & Newell, 1994; Sun, 2006). This point, that computational modelling does not necessitate computational theory of mind, was made by both Searle (1992) and Putnam (1991). It is also endorsed by Dennett (1979) who states:

"The research strategy of computer simulation has often been misconstrued by philosophers. Contrary to the misapprehensions innocently engendered by Turing's classic paper, "Computing Machinery and Intelligence," it is never to the point in computer simulation that one's model be indistinguishable from the modelled. Consider, for instance, a good computer simulation of a hurricane, as might be devised by meteorologists. One would not expect to get wet or wind-blown in its presence... A good computer simulation of a hurricane is a program, which, when you feed in descriptions of new meteorological conditions, gives you back descriptions of subsequent hurricane behavior."

There are researchers who consider Computational Theory of Mind to be a useful epistemology (Marr, 1982). Others take a more literal, realist stance toward CTM (Anderson, 1983). This paper does not claim that the models presented here are identical to the internal process of human cognition. Our models assume Dennett's analogy that computer simulations of weather can be insightful without the code mistaken for *actual* weather. This is the classic map/ terrain distinction. Likewise, we propose that a computational map of metacognition can assist our understanding of how metacognition functions.

4.2 Modeling as a methodological tool for science

The cognitive revolution advanced the notion that understanding the mind's outputs required information-processing explanations. Presently, cognitive science overwhelmingly embraces the "computer metaphor" of cognition. The viewpoint of cognition as computation has formed much of the basic terminology within the field. In this view, the cognition inputs data, transforms information according to rules and outputs actions. Because computational terminology forms much of the language of cognitive science, it is sensical to analyze metacognition in this way. Understanding cognitive processes by building process models is the method of cognitive modeling. The scientific validity of this approach has been thoroughly researched (Simon & Wallach, 1999). However, the scope of this paper is to propose a novel method of cognitive modeling by applying it as a philosophical tool.

According to Brook (2009), philosophy plays a crucial role in cognitive science. Among the many contributions of philosophy Brook discusses is the analysis of concepts. Philosophical analysis can clarify concepts and recommend how they be used. By clarifying terminology,

methodology, and our underlying assumptions, philosophy can assist phenomena in becoming more intelligible. Improving a subject's intelligibility can help to guide research and refine hypotheses for testing. The present thesis employs this virtue of philosophy by analyzing metacognition through a computational lens. In doing so we employ another strength of philosophy, that of integrative analysis. By analyzing metacognition computationally, this thesis helps integrate research by positing a single set of mechanisms for all metacognitive tasks. This follows Newell's (1990) call to build unifying computational models of cognition. Notably, this connecting research in artificial cognition with philosophy was encouraged by Dennett (1978):

"Philosophers, I have said, should study AI. Should AI workers study philosophy? Yes, unless they are content to reinvent the wheel every few days... Clearly a meeting of minds is in order." Dennett (1978) further supported philosophical analysis by way of computer simulation:

"It is notoriously difficult to keep wishful thinking out of one's thought-experiments; computer simulation forces one to recognize all the costs of one's imagined design. As Pylyshyn observes, 'What is needed is . . . a technical language with which to discipline one's imagination.' The discipline provided by computers is undeniable."

This thesis intends to 'discipline our imagining' by using the 'technical language' of computation. This paper's philosophical-computational approach to metacognition is founded at the base of cognitive architectures — a production system. Single productions forms this architecture's base units, and are analogous to single bricks. By connecting productions to each other, brick by brick we build layers of productions that are increasingly complex. These layers are then analyzed and correlated with metacognitive processes within the research literature. This novel method possesses a rare virtue — the fundamental units of metacognition are well-defined. The resulting metacognitive framework is uniquely coherent.

4.3 Cognitive modelling

Cognitive modelling connects to the history of science by building theories of empirical phenomena that are tested and adapted based on data. A cognitive model is a process model — it depicts how a psychological process functions over time (Pirolli & Card, 1999). A computational cognitive model is founded on computer science. It represents cognitive processes as computer code that can be tested empirically. The code produces a runnable simulation using transparent algorithms. Hence, the deep components of cognition can be represented in a mechanistic, well defined way. As Sun (2004) writes:

"Computational modeling appears to be the most promising approach in many respects, and it offers the expressive power that no other approach can match, as it provides a variety of modeling techniques and methodologies and supports practical applications of cognitive theories."

Generally, a computational model depicts a minor cognitive process or cognitive phenomena. A cognitive architecture is more elaborate. It depicts how the system's parts interact to generate cognitive phenomena. As stated by Ron Sun (2004):

"A cognitive architecture denotes a comprehensive, domain-generic computational cognitive model, capturing the essential structures, mechanisms, and processes of cognition. It is used for a broad, multiple-level, multiple-domain analysis of cognition and behavior."

An architecture's capacity for a broad, multi-level analysis is essential for describing how cognition's parts interact to produce metacognition. The architecture employed herein is the Common Model, as instantiated within ACT-R (Laird et al., 2017). The Common Model resulted from integrating the common features of nearly all cognitive architectures. The birth of cognitive architectures was as a result of Allen Newell's call for cognitive unification.

4.4 Unification

The importance of developing cognitive architectures became apparent within the first half of the 20th century. By that time, large amounts of psychological research data had been accumulated. Observing this, Newell (1973) proclaimed that these data were fragmenting instead of converging. In his famous paper "*You can't play 20 questions with nature and win*," Newell challenged researchers to integrate the fragmented data by building unifying theories. Newell called for a "unified theory of cognition" in his now famous William James Lecture (1987).

"Psychology has arrived at the possibility of unified theories of cognition — theories that gain their power by positing a single system of mechanisms that operate together to produce the full range of human cognition... We should strive to attain them."

Newell advocated a single set of mechanisms that would explain diverse cognitive phenomena. His was a call to scientific parsimony — a Lakatosian scientific framework of integrating data under unifying explanations. Newell went on to develop the first cognitive architecture called SOAR. Importantly, Newell (1972) set down a production system at the heart of the architecture, "*I laid out production systems as detailed models of the human control structure*." Following Newell's lead, researchers began to build cognitive architectures to model human cognition, such as Clarion (Sun, 1997), Epic (Meyer & Kieras, 1997), Sigma (Rosenbloom, Demski, & Ustun, 2016), and ACT-R (Anderson & Lebiere, 1998). While these researchers would disagree on much, they would also discover much common ground. The cognitive architectures that developed independently were later found to have common features. These common properties included a production system, declarative memory, input and output functions, and more. These commonalities were then overlaid together to present a single cognitive architecture — the Standard Model (Laird et al., 2017). Neuroimaging research indicates that human cognitive operations correspond to the Standard Model cognitive

architecture (Stocco et at., 2018). The research analyzed numerous neuroimaging studies to determine that the Standard Model was a reliable explanation of the cognitive patterns discovered. These results provided clear empirical support for the theorized cognitive architecture. The Standard Model has helped unify the pursuits of cognitive modellers and has since been renamed the Common Model of cognition.

4.5 ACT-R

One instantiation of the Common Model is ACT-R, which is the cognitive architecture used herein to model metacognition. Create by John Anderson, ACT-R stands for Adaptive Control of Thought— Rational (Anderson & Lebiere, 1998; Anderson, 2002). Inspired by Newell (1990), ACT-R aims to define the fundamental operations of human cognition. It specifies how cognition is organized into modules to produce intelligent behaviour. This architecture has successfully predicted a variety of human tasks, from human-computer interfaces to mathematical reasoning. The following is a brief overview of the main components of the ACT-R cognitive architecture.





Figure 4. Overview of the ACT-R cognitive architecture

Figure 5. The Common Model of Cognition (West & Conway-Smith, 2019)

ACT-R is composed of modules depicting vision, motor, declarative memory, and procedural memory. Each module is controlled by buffers (Figure 4). A buffer is a temporary container that holds and transfer information. Buffers pass information between modules. Buffers also make information available to the central processor called procedural memory. The operation of procedural memory is the active part of the system. Procedural memory runs productions that are analogous to conveyer belts moving information. These productions move an otherwise unmoving system. The procedural system acts by applying production rules to the information within buffers. As the buffer conditions change, different productions act.

A production is a form of information that is encoded as a pair. A production is paired as a "condition-action" and forms a single unit. *If* a condition arises and matches the production, *then* the production acts. This is also known as an "*if-then*" rule. The *if* specifies the condition; the *then* specifies the action. "*If*" conditions match, "*then*" the prescribed action fires (e.g.: *If* thirsty, *then* drink water). The production rules transform information to complete a task or resolve a problem. While relatively simple in nature, productions form the core of the cognitive architecture. Production rules are held within the storehouse of procedural memory. Neurologically, productions correlate with the 50 millisecond decision timing in the basal ganglia (Stocco, Lebiere, & Anderson, 2010).

In ACT-R, the other way information is encoded is within declarative memory. Here, information is encoded into "chunks" (Anderson, 1993). These chunks take the form of declarative information that are connected into a pair of *feature: value* (fire alarm:danger). Chunks are encoded in declarative memory and retrieved through its buffer. While inside the buffer the chunk may trigger procedural memory to apply production rules. These rules may transform the information, or activate other productions, such as "*if fire alarm signals danger*, *then run from building*." These features of the production system are centrally important to the purpose of this paper. The production system will serve as the foundation for a framework modeling metacognition.

4.6 **Production system as computational foundation**

Here, I present support for the premise that productions are an appropriate foundation for modelling metacognition. We recall that the first cognitive architecture was founded on a production system, as stated by Newell (1990).

"Soar has a symbol system as a goal hierarchy. It uses a production system as the foundation of the architecture, thus being a kind of recognize-act system. It uses a problem space everywhere to do all of its business, that is, to formulate all its tasks." (p. 39)

Like Newell, Anderson made a production system central to his cognitive architecture ACT-R (Anderson & Lebiere, 1998). Further, Anderson is among the many researchers who consider productions to be central to human intelligence. In his book *Rules of the Mind* (1993), Anderson asks "What happens in the head to produce human cognition?" He answers that "*cognitive skills are realized by production rules. This is one of the most astounding and important discoveries in psychology.*" This statement necessarily entails that metacognitive skills, too, are realized by production rules.

Be it ACT-R, SOAR, or the Common Model, cognitive architectures employ a production system at their base. These systems may be implemented differently in neural networks (Campos & Oliveira, 2013), symbolic systems (Townsend, Keedwell, & Galton, 2014), high dimensional vector systems (Kelly, Arora, West, & Reitter, 2019), or hybrid architectures (Kasabov, 1995). Their instantiation does not alter the fact that productions are the systems' foundation. This convergence helps to establish our premise that *productions are a sensible foundation for modelling metacognition*. It is also useful that productions are a concept well-defined by computer science.

The concept of a production — our foundation for metacognition — is computationally well-defined and logically coherent. Again, a production is a *condition* \rightarrow *action* unit. It is a single "*if* \rightarrow *then*" rule. The transparency of this definition reveals its inner coherence. A production is not grounded externally, but internally — its parts conceptually entail each other. The *if* is coherent by preceding a *then*. Likewise, the *then* is sensical by following an *if*. A production's definition leaves no room for a metaphysical dispute. Our claim is simply that a production is coherent *computationally*. Its definition is internally coherent and functionally demonstrable. The functionality of a production is verifiable within cognitive architectures and computational systems the world over.

Further, a production's inner coherence reveals it to be irreducible. This irreducibility was referred to by Newell (1990) while discussing the operations of a production system:

"The system level does have one limiting qualitative feature, namely, the operations cannot themselves be composed of operations."

The operations of productions are not composed of more operations - this architectural limitation adds another bedrock of philosophical support. Its irreducibility, coherence, and functionality are what qualifies productions to be the foundation of a framework for metacognition. This novel method of establishing a computational base can be described as *computational foundationalism*. This foundation is comprised of a base of sound premises, summarized here - a production is:

- the base of successful cognitive architectures
- considered the source of human cognitive skills
- empirically correlated with decision-making in the basal ganglia
- internally coherent

With the foundation established, we move on to describe a novel method for building metacognition.
4.7 Axiomatic system

As we infer from sound premises, we enter the realm of formal systems. In many formal systems, an established concept can be granted axiomatic standing (Shoenfield, 1977). An axiom can be established by selecting basic concepts and explaining them. This concept may be determined empirically, deductively, or *pro tem*. From the founding axiom, deductive reasoning can take place. Logical rules can be applied to infer further statements. These newly deduced statements are called theorems. Here, I present an *analogous* system that logically extends the computational foundation of a production.

This process can be likened to the axiomatic system of Euclidean geometry. Here, axioms that are self-evident are extended. For example, we may begin by describing a single point (**Figure 6**). From this point we continue by stating that, given two points, a terminal line passes between them. Next, we can



state that two parallel lines will neither converge nor diverge. This building process continues to encompass shapes and geometric theorems. Axioms begin as simple and certain. When extended, the results are far from simple but just as certain. Because an axiom is well-defined, the framework is coherent. Errors in logic may be rooted out systematically. Hence, much depends on the validity of the base axiom. For this reason I have spent thus far establishing our foundational axiom.

An axiom that is particular to a field of science is called a postulate. Analogously, we might say that our "postulate," particular to computational cognitive science, is a production system. Using this as a starting point, we can deduce further statements using the logic of cognitive architectures. These newly deduced statements (theorems) will be called "forms." These forms will be built into a framework for analyzing metacognition. The forms result in a spectrum of metacognition, or a metaphorical mountain with an explicit peak (**Figure 7**). This spectrum will reveal the ways that, according to its definition, metacognition can be seen as "cognition acting on itself." This method accords with Newell's (1990) construction of architectures by which he described "building production systems upward." Hence I am adapting an established method



for building cognitive architectures. The difference is that, where this method has been employed informally, I apply it formally using deduction. The following section commences building productions into 'forms' of metacognition. We begin with the first brick — a single production.

5.0 Forms of Metacognition

"The understanding, like the eye… takes no notice of itself ; it requires art and pains to make it its own object." — John Locke (1689)

The following section discusses **Figure 8**, a hierarchical model of metacognition. This contains the main contribution of this thesis. This section incorporates information from Newell (1990), Anderson (1983), and the Common Model (Laird et at. 2017). In Figure 8 below, productions are drawn as ovals, buffers are rectangles, and declarative memory are squares labeled "DM." Each layer moves chronologically from left to right. Each step up is a logical extension of the form below. The following explanation commences at the bottom of Figure 8 with a single production.

1. Single

A single production forms the base unit of the foundation. The unit is a condition-action statement, a single " $if \rightarrow then$ " rule. The "if" stands ready to match a condition within a buffer. Should the conditions match, it fires a predetermined "then" action.



Micro cognition: In Form 1 *a single production fires* from procedural memory. This can be described as the simplest cognitive process laying at the base of cognition. Because it is one cognitive state acting on another (if \rightarrow then), it roughly accords with Flavell's (1976) definition of cognition acting on itself. In this sense it might be considered *proto* meta. However, there is no explicit self-reference or representation, and so it does not correspond to metacognition as described in the literature. Here, it is called "micro cognition" from the greek word meaning "small" since it refers to the most reduced level of cognitive operations — single productions.

2. Chain

Next to the first production we connect a second. The first production sets up the next production to fire. This process repeats along a potentially endless chain. Each production determines the next, which fires ballistically. A series of pre-determined productions characterize an algorithm — a linked chain of "*if*—*then*" rules.

Micro cognition: Here, the simplest cognitive productions set up the next productions. As these cognitive states determine other cognitive states, in this sense productions might be considered 'meta' or *proto* meta. Yet without an explicit self-representation it is not metacognition as described in the literature. A chain of productions is also "micro cognition."

3. Competing

A second layer of productions is added. Both chains match the same initial condition, and so compete to be fired. In this case, when two productions match the same condition, the competing production with the higher utility fires. Utility is the principle determining priority.

Micro cognition: Here, productions determine other productions, prioritized by their utility level. Yet with no explicit self-reference, they remain proto meta.

4. Interrupt

Here, *a production is interrupted* by another production. The interrupting production enters the buffer from either an external or internal source. However, once inside the system, all productions act identically. Here, the interrupting production is prioritized as it contains a higher utility level. The previous production is abandoned in favor of the interrupting production.

External interrupt — Here, a production is interrupted by an external production. External is defined as initiated from *outside* the system (i.e.: vision, sound, etc.). The old production chain is replaced by the new production with higher utility. An example of an *external* interrupt might be an agent reading a book. The reader is interrupted by hearing a fire alarm. The alarm triggers a production stored in memory: "*if* alarm, *then* run from building."

Internal interrupt— Here, a production is interrupted by an internal production. Internal is defined as initiated from *inside* the system (i.e.: memory, feeling, etc.). The old production chain is replaced by the new production with higher utility. An example of an *internal* interrupt is, again, an agent reading a book. This time the reader is interrupted by an *internal* thought. The agent suddenly remembers they are late for an appointment. The recalled thought fires new productions, "*if* late for appointment, *then* run to appointment."

Micro cognition: Productions that interupt other productions are prioritized by their utility level. Both external and internal productions behave identically, setting up the next productions. Because of this they may be said to be *proto* meta. Still, the productions contain no self-representation, and so do not correspond to its description in the literature. Hence, these simple production remain as "micro cognition."

5. Problem search

In Form 5, a chain of productions is interrupted by a stimulus that is is *uncertain*. This uncertain stimulus makes it unclear what actions to take. This uncertainty triggers problem solving productions. Operators depicting possible solutions are searched until the right production found. Newell (1990) describes this as a problem space, which resolves problems by representing the start and desired end. The solution involves finding the operators that move from the start to the desired end. Operations are searched to arrive at a *resolved productions* (shown in **Figure 8** as the "!" symbol).

For example, an agent may be interrupted by an alarm that is *uncertain*. The alarm may sound both like a fire alarm and a car alarm. The agent is unclear whether to run from the building. To resolve this uncertainty they begin problem solving productions. Once the uncertainty is resolved, fire alarm productions execute, and the agent runs from the building.

Implicit metacognition: Previously, productions had been determined by the previous production. Now, the deterministic chain is interrupted by a state of uncertainty. This uncertainty initiates productions that resolve this cognitive state. Since productions specifically address an internal state, this is "meta" as more closely befits the literature. They are larger processes that address cognitive states. Yet *without an explicit reference* to the process the meta remains implicit and so here is called "implicit metacognition."

6. Knowledge Search

Within intelligence there are two searches: *problem search* and *knowledge search* (Newell, 1990). During knowledge search, knowledge is retrieved from memory to guide the problem search. Knowledge is searched for knowledge that can be used to navigate the problem space.

Implicit metacognition: Knowledge can be searched to solve both physical problems and cognitive problems. When knowledge is searched and applied to external objects (e.g.: pouring water into a cup) this is externally-directed cognition. When one is applying knowledge toward an internal cognitive task (multiplying numbers) this is implicit metacognition. Should this internal task be represented by knowledge ("I will undergo a cognitive task") then this is explicit metacognition. Without an explicit referent, a cognitive task is implicit metacognition. *Explicit metacognition:* When knowledge is searched for a cognitive-level task (e.g.: finding a strategy for preforming math) and represents the process, this is explicit cognitive instructions, this is explicit metacognition. For example: "Now I will go about a math problem, what is the best way?" Knowledge is then searched for the best way to go about the cognitive task. "Remember the mnemonic device BEDMAS for the order of operations — brackets before exponents!" Here a persons is not simply going about a cognitive task, they are applying cognitive knowledge about how to go about the task. This explicit reference to a cognitive state is what distinguishes explicit metacognition.

7. Knowledge Request

A knowledge search is not necessary when there is enough knowledge to complete a task. A knowledge request can retrieve the knowledge needed. In ACT-R, knowledge is retrieved from declarative memory as a "chunk" or instruction information. When this information is applied directly to a task, the agent moves straight to the end of the problem space. In **Form 7**, we see that knowledge representing a task is requested and applied. Knowledge may inform an object-level task (driving a car). Also, knowledge may inform a cognitive-level task, such as memorizing words. A person may go about this without reflecting on the process of memorization. This lack of reflection is *implicit metacognition*. Alternatively, one might retrieve

a *cognitive strategy* for memorization, such as imagining words as pictures (Spence, 1984). Here, knowledge refers to the cognitive task explicitly, and is *explicit metacognition*.

Explicit metacognition: A chunk with a cognitive referent interacts with the cognitive process. This cognitive referent determines the next productions that will fire. For example, a declarative chunk might instruct someone to drink *coffee* to become more *alert* (object-level to cognitive-level). Here, a chunk contains a reference to a cognitive state, which determines the next productions. This cognitive referent is what qualifies as explicit metacognition.

8. Knowledge referent

Form 8 focuses exclusively on the *referent* of knowledge. The two main categories of referents are object-level and cognitive-level (Nelsons & Narens, 1992). Object-level cognition is guided by a chunk with object-level referents, for example, "swing hammer to strike a nail." This chunk only refers to external objects and lacks a cognitive referent. Knowledge that does contain a cognitive referent (meta-knowledge) engages with the cognitive-level (metacognition). Hence, knowledge can contain four combinations of referents:

- 1. *cognitive* \rightarrow *cognitive* (employing science to attain knowledge)
- 2. *cognitive* \rightarrow *object* (feeling tired and going to bed)
- 3. *object* \rightarrow *cognitive* (drinking coffee to wake up)
- 4. $object \rightarrow object$ (swinging a hammer to strike a nail)

Explicit metacognition: Above, knowledge referents are arranged in four possible combinations. Numbers 1, 2, 3 contain cognitive referents that refer to some part of the internal system (wakefulness, tiredness, employing science). Here, an explicit representation of an internal process (meta-knowledge) directs productions. Bechtel (1998) asserted that a representation interacts with operations by engaging in the control process. In **Figure 8**, Form 8

we see this explicitly — information engages with the system as buffer states. Productions then fire off these buffer states (as they do in all cases). A "referent," at its base, is simply a buffer condition that activates productions. *The combination of referent determines the category of metacognition*. The topmost peak of **Figure 8** holds the most explicit form of metacognition — instructions containing *two* cognitive referents. The instructions direct a cognitive *means* to a cognitive *end*. These cognitive ends have categories: memory, reasoning, emotion, attention. These are cognitives states can be improved through metacognitive instruction. For example, *memory* can be enhanced through cognitive instructions (Spence, 1984). *Reason* can be improved through directions (Leevers & Harris, 1999). *Emotional* regulation can be heightened by cognitive behaviour therapy (Dobson, 2013). *Physical* endurance via mental fortitude can be strengthened by instructions from sports psychologists and military trainers (MacIntyre, et al. 2014; Wiltshire, Rosch, Fiorella, & Fiore, 2014). Significant research has established how metacognitive instructions can yield higher human performance (**Figure 11**).

Metacognitive Instructions				
MEMORY	Learning (Schraw, Crippen, & Hartley, 2006)	Mnemonic devices (Maguire et al., 2003)	Memory Palace (Spence, 1984)	
REASONING	Science (Zion, Michalsky, & Mevarech, 2005)	Logic (Leevers & Harris, 1999)	Math (Kramarski, Mevarech, & Arami, 2002)	
PHYSICAL	Sports (MacIntyre, et al., 2014)	Navy seal training (Wiltshire, Rosch, Fiorella, & Fiore, 2014)	Motor skills (Augustyn, & Rosenbaum, 2005)	
EMOTIONAL	CBT (Dobson, 2013)	Self-regulation (Vukman, & Licardo, 2010)	Meditation (Baird, Mrazek, Phillips, & Schooler, 2014)	

5.1 Cognitive referent and intentionality

There is significant research on the *cognitive referent* being the distinctive feature of metacognition. As stated by Wellman (1983), "*Metacognition differs from standard cognition in that the self is the referent of the processing or the knowledge*." Similarly, Metcalfe et al. (2012) has identified the referent of metacognitive judgements to be definitive. Two important terms include "noetic" - where the referent involving mental representations, and "anoetic" - where the referent involving objects in the world. Similarly, researchers Nelson and Narens (1992) famously distinguished two referents of metacognition; meta-level and object-level.

It is notoriously difficult to explain how a representation "refers" to a physical or cognitive state. This involves the *intentionality* of the representation. This refers to Brentano 's (1874) classic thesis that all mental phenomena exhibit intentionality. Mental events are "about" something else, i.e.: they contain "aboutness." Dennett (1981) furthered this notion with his view of "intentional systems." He asserts that explanations and predictions can be reliably generated by ascribing beliefs, desires, and intentions to the system. He states:

"By assuming the computer has certain beliefs (or information) and desires (or preference functions) dealing with the chess game in progress, I can calculate - under auspicious circumstances - the computer's most likely next move. The computer is an intentional system... because it succumbs to a certain stance adopted toward it, namely the intentional stance."

A similar statement can be made regarding Brentano's notion: knowledge has intentionality because it generates effects reliably conforming to its referent. This helps explain why referents dependably create the states they are *about* (Figure 11). A knowledge referent that is *about* affecting an *external* state directs productions to affecting that external state (object-

level). A knowledge referent that is *about* affecting an *internal* state directs productions to affect that internal, cognitive state (cognitive-level). Whether knowledge contains an internal or external referent do not imply any unique metaphysical properties. This is irrelevant to the discussion. *The standard of measurement is computational functionality*. In a computational model, cognitive referents cause cognitive state changes in a reliable, functional way. Hence, we may define explicit metacognition as a cognitive referent directing a cognitive state change. Externally directed cognition involves a object-level referent causing an external change.

5.2 Discussion of metacognitive instructions

Bechtel (1998) claimed in his Dynamical Systems Theory (DST) that a representation interacts with cognitive process by engaging with the operations. Here, Bechtel famously employed van Gelder's (1995) analogy of a Watt's governor, which he called a "landmark for models of cognition." (Gelder intended it as an anti-representationalist argument, while Bechtel employed it for the contrary purpose). This analogy accords with how representations function in cognitive architectures. When a representation is in a buffer, it is simply a condition that productions match to. The specific representation determines which productions match and fire. A meta-representation functions this way — it is a condition that productions match to. A meta-representation acts as a buffer state that determines the productions leading to a cognitive effect (memory, reason, etc.). Thus, a representation can direct productions to change an internal cognitive state, or an external object state. The difference is what level of state-change is referred to by the representation.

We have pursued a distinguishing feature of metacognition while the building Forms 1-8.

On every level, we found productions firing and changing internal states. On each level of the cognitive architecture, we found "cognition acting on itself." This accords with Flavell's (1976) definition of metacognition. Assuming the simplest of definitions, we might say that all cognition acts on itself, that cognition is "meta all the way down." Yet this is not how it is discussed in the literature. The literature usually refers to explicit knowledge of cognitive states. This feature of metacognition is found at the topmost level (Form 8). The top level is differentiated by the explicit knowledge of cognitive states. To isolate this concept we employ the Venn diagram in **Figure 9**.

A Venn diagram is used in set theory to identify and differentiate properties between two groups (Gil, Howse, Kent, & Taylor, 2000). In Form 7, we see that both cognition and metacognition share the same properties of the cognitive architecture. They all include productions, problem search, knowledge search, and knowledge chunks. The unique feature is found at the top, within the most explicit metacognition and its level of referent. This is represented in the Venn diagram in **Figure 9**. The green area represents object-level cognition. Cognitive-level metacognition shares all the architectural properties of object-level cognition



(productions, chunks, etc.) except for one. The blue area depicts the sole differentiating feature of explicit metacognition: a cognitive referent. The referent is the only distinguishing property not shared by the rest of the architecture. In the following section, this distinguishing green-blue color scheme will continue as object referents are coloured green, and cognitive referents are blue.

Figure 10a shows an example of how a cognitive states can be represented, and direct productions. The buffer on the left contains a cognitive state — sleepy. This triggers a production to retrieve the appropriate knowledge from declarative memory (DM). Declarative memory

moves a knowledge chunk into a buffer, containing information with a cognitive referent — sleepy. This information is paired with knowledge on how to change that cognitive state — coffee (green=object-level). The chunk triggers productions toward the object that will change the cognitive state. The chunk contains meta-knowledge that influences which





productions will fire, and hence qualifies as explicit metacognition. This is depicted in Figure 8 as the *monitor level* of metacognition. This corresponds to the meta-level defined by Nelsons and Narens (1990).

Figure 10b contains an example of the instruction level of metacognition. It is the topmost

category of metacognition, containing two cognitive referents (blue:blue). The left buffer contains a cognitive state — distracted. Declarative memory (DM) retrieves knowledge representing this state. The knowledge is paired with information on how to change that state (focus). This knowledge fires productions initiating a deliberate cognitive action — focus.



Instruction Level Metacognition			
Memory	Learning (Schraw, Crippen, & Hartley, 2006)	Mnemonic devices (Maguire et al., 2003)	Memory Palace (Spence, 1984)
Reasoning	Science (Zion, Michalsky, & Mevarech, 2005)	Logic (Leevers & Harris, 1999)	Math (Kramarski, Mevarech, & Arami, 2002)
Physical	Sports (MacIntyre, et al., 2014)	Navy seal training (Wiltshire, Rosch, Fiorella, & Fiore, 2014)	Motor skills (Augustyn, & Rosenbaum, 2005)
Emotional	CBT (Dobson, 2013)	Self-regulation (Vukman, & Licardo, 2010)	Meditation (Baird, Mrazek, Phillips, & Schooler, 2014)

Figure 11. Chart of metacognitive instructions

Figure 11 briefly sets out the different categories of metacognitive instructions within the research literature. These instructions can engage with **Figure 10b** as meta-knowledge chunks that direct productions. This variety of meta-knowledge contains the cognitive improvements made possible through metacognitive instructions.

Metacognitive instructions specify the means to attaining heightened performance in areas such as memory, reasoning, emotional regulation, and physical performance. These instructions are types of "mind tools" discussed by Dennett (2000). At their base, metacognitive instructions inform how to apply cognitive means to attain cognitive ends. The nature of meta-knowledge is to inform deliberate cognitive state change.

The effectiveness of metacognitive instructions to change cognitive states has been well researched within cognitive psychology. This research indicates that metacognitive instructions

can improve cognitive functioning beyond what is otherwise attainable (Redding, 1990; Yanqun, 2019). A convergence of research indicates that cognitive optimization requires metacognition, as many cognitive enhancements can not be purchased without metacognitive coin.

5.3 Connection to cognitive psychology terminology

This section attempts to bridge the terminology set out in the literature review with the forms of metacognition in Figure 8. **Figure 12** sets out the commonalities between cognitive psychology terminology and computational terminology. Cognitive researchers distinguish between *cognitive knowledge* and *cognitive regulation* which connects to computational terminology employed within the Common Model (ACT-R).

Cognitive Knowledge	Common model	Forms
Knowledge of factors affecting cognition.	Declarative chunks	6, 7,8
Knowledge of management and strategies.	Declarative instructions	6,7,8
Knowledge of when to use a strategy.	Utility level(2-8), Knowledge request	6,7,8

Cognitive Regulation	Common model	Forms
Aware of comprehension and task performance.	Buffer content(all), Meta-data	5, 6, 7, 8
Evaluating processes, and the product of learning and revising goals.	Problem solving productions	5, 6, 7, 8
Identifying and choosing strategies.	Knowledge search, Knowledge request	5, 6, 7, 8

Figure 12. Connections between cognitive psychology and computational terminology.

5.4 Connection to philosophy

In regards to the "attributivist vs evaluativist," debate, the "Self-Attributive View" holds that metacognition necessitates representing mental states *as* mental states. This is referred to as mental state self-attribution (Carruthers, 2011). The "evaluativist" or "Self-Evaluativist View"

holds that metacognition is the *activity* of monitoring and controlling mental processes (Proust, 2014). The self-evaluativist view agrees with a computational framework in which metacognition is activity-dependent. Metacognition is the process of meta-knowledge directing cognitive productions to affect cognitive state changes. Concerning the debate between "domain-general" and "domain-specific" accounts of metacognition, computationalism holds a domain-general view, as the architecture is reused between cognitive and metacognitive processes.

5.5 Connection to neuroscience

The viewpoint that cognition is "meta all the way down" has a neurologically correlation. Just as a basic production triggers other productions, a single neuron activate another neurons. This can be considered *proto* metacognition on a neural level (forms 1-4) more accurately called micro cognition. Neuroscience more generally researches the interactions between dedicated brain regions, correlating to implicit metacognition (forms 5 & 6). Importantly, peak brain activity during explicit metacognition correlates with a region of deliberate executive function — the prefrontal cortex (Nelson & Narens, 1992; Daw, Niv, & Dayan, 2005; Anderson & Fincham, 2014; Lee et al., 2014; Stocco et at., 2018). Explicit metacognition is correlated largely with the rostrolateral prefrontal cortex (Fleming & Dolan, 2012; Anderson & Fincham, 2014). Further, the neuroimaging study by Anderson & Fincham (2014) indicates that metacognitive ability depends on having declarative representations of one's own procedure, i.e.: the meta-knowledge chunks in **Figure 8** and **Figure 10 a & b.** These data indicate that explicit metacognition is neurologically correlated with the most deliberate, executive operations of the brain.

6.0 Metacognitive feelings

In addition to metacognition being directed by knowledge, metacognition is routinely directed by *epistemic feelings*. Accounting for metacognitive feelings is important since they have strong effects on cognitive performance and decision-making (Hart, 1965, Singer & Tiede, 2008). This section discusses the relevant research and proposes how to represent metacognitive feelings computationally.

6.1 Feelings of knowing

Feelings related to knowing or learning have been referred to by terms such as; feelings of knowing (Hart, 1965), metamemory (Flavell, 1976), knowledge judgements (Schneider, 2000), cognitive emotions (Standish, 1992), and epistemic feelings (Arango-Muñoz & Michaelian, 2014). This lexicon will be subsumed herein under the term "noetic feelings." This follows Metcalfe's (2013) identification of "noetic" to mean cognitive phenomena in which the referent concerns an internal state or internal representation.

William James (1890) was the first to research the "tip of the tongue" phenomenon — the feeling of knowing that is empty of knowledge itself. Since then, researchers have identified many ways that noetic feelings can direct cognition (Table 2). Research indicates that noetic feelings drive memory search as subjects take more time to search their memory if they "feel" they know it (Barnes et al., 1999). This is represented in **Figure 13 & 14** below Table 2. Studies also show that noetic feelings are reliable signals of the likelihood of memory retrieval (Hart, 1965). Moreover, feelings of probable retrieval success or retrieval failure affect the strategy used to engage the problem (Singer & Tiede, 2008). Noetic feelings have also been reliably correlated with improved learning outcomes (Wang, Haertel, & Walberg, 1990). Subjects will

spend more time learning words they feel are difficult to remember (Nelson & Leonesio, 1988). Moreover, the "feeling of rightness" has shown to direct the rapid solving of complex, real-world problems (Thompson & Johnson, 2011). Thus, research indicates that *noetic feelings are a guidance system integral to directing cognitive processes*.

Noetic Feelings of Knowing		
Feeling of knowing	Hart, 1965; Nelson & Narens, 1990; Koriat, 1993; Dunlosky & Metcalfe, 2009	
Tip of the tongue	James, 1890; Brown & McNeill, 1966; Spehn & Reder, 2000; Schwartz & Metcalfe, 2011	
Feeling of understanding	Mangan, 2001; Trout, 2007	
Feeling of rightness	Mangan, 2001; Thompson et al. 2011	
Feeling of meaning	James, 1890; Strawson, 2011	
The feeling of forgetting	Halamish et al., 2011; Kornell, 2011; Arango-Muñoz, 2014	
The feeling of error	De Neys, 2012; Wessel, 2012	
Deja Vu	Brown, 2003; Schwartz and Cleary, 2016	
Feelings of confidence	Winman and Juslin, 2005; de Sousa, 2008	
Feeling of familiarity	Whittlesea, 1993; Whittlesea & Williams, 2000	
Agency over thoughts	Proust, 2006, 2009; Carruthers, 2012	
Ease of learning	Koriat, 1997	
Certainty / uncertainty	Smith et al. 2003; Dokic, 2012	
Feeling of rationality	James, 1879	

Table 2. Noetic feelings of knowing

A unified theory of metacognition requires a model of how noetic feelings engage in metacognitive processes. While cognitive architectures have been successful in modeling knowledge-driven behaviour, it is unclear how feelings would direct these architectures. Cognitive models associated with noetic phenomena include Reder's (1996) use of the source of activation confusion (SAC) model, Dougherty's (2001) multiple-trace memory model, Metcalfe's (1993) holographic associative model, and Sikström and Jönsson's (2005) stochastic drift model of memory strength to explain delayed judgement of learning.

6.2 Representing feelings

Computational Theory of Mind (CTM) debates are focused around concepts such as symbolic representation, referents, semantics, propositions, qualia, and meaning. CTM is intended to describe the relationship between computation and mental states. Yet, because most of the discussion is based around knowledge and language, it is unclear if these concepts are meant to apply beyond this domain (Rescorla, 2015). In particular, there seems to be an intuition that feelings are not the same as thought or language, and so must be computationally represented in a different way.

According to Damasio (2019) feelings are mental representations of non-symbolic bodily states, which are used for decision making. According to Alston (1969), feelings are "spontaneously-emerging occurrent phenomenal experiences," which he refers to as "datable states of consciousness." However, Arango-Muñoz & Michaelian (2014) indicate that feelings do not involve "properly propositional content."

Overall, there seems to be agreement that a feeling is a unitary phenomenon that we have potential conscious awareness of. Feelings can factor into decision making but there is a common intuition that feelings are somehow different from propositional, symbolic knowledge. Feelings can be derived from more complex, distributed phenomena, such as emotions and drives.

Instead of debating the nature of feelings *qua* feeling, this work is grounded in understanding the *function* of feelings within cognitive models. Our method does not weigh arguments for metaphysical validity but rather *functional viability* — how feelings operate computationally.

6.3 Feelings as Metadata

We propose that feelings function as metadata. In this case, "meta" does not mean "selfreferential" as it might in metacognition. Here, metadata is used in the sense of "aggregate data," in the way that a meta-analysis yields the aggregate data of many studies. Viewing feelings as metadata is reasonable since, computationally, metadata are typically different from knowledge expressions. For example, metadata is often best expressed through statistics and high dimensional spaces, whereas knowledge is often best expressed through propositional statements and logical operators. The nature of aggregate metadata accounts for the non-verbal qualia of feelings. Its aggregate property also explains why noetic feelings are reliable indictors of knowledge, without retrieving the knowledge itself.

Metadata also accounts for how the noetic *feelings of rightness* derive from procedural memory (Mangan, 2001; Thompson & Johnson, 2011). Procedural memory is known to be cognitively impenetrable (Squire & Zola, 1996). Hence, procedural memory uses noetic feelings as a communication bridge. Feelings of "rightness," "wrongness," and "certainty" are a way the procedural system interacts with executive functions. This accounts for why, for example, someone can not verbally explain how to play an instrument, ride a bike, or use implicit grammar rules — yet this complex behavior results from being guided by *feelings of rightness* or wrongness. Feelings are the aggregate of countless practiced actions stored in procedural memory, which communicate their sum through metadata (noetic feelings). This paper proposes that the sum of metadata is placed into a "noetic buffer" (**Figure 13**) that serves as a bridge for deep learning to interact with symbolic computation (West & Conway-Smith, 2019). Matching patterns of metadata with symbolic computation is feasible since, as Newell (1990) states, "A productions system is simply a form of *recognition system* — it recognizes patterns" (p.165).



Additionally, feelings often trigger knowledge requests, as established by Nelsons and Narens (1990), and Barnes et al. (1999). This process is depicted in **Figure 14**, as deep learning places a feeling of knowing into a buffer. This feeling then triggers a production to request the knowledge from declarative memory (DM). Noetic feelings that direct cognition without cognitive referent are *implicit* metacognition. This process corresponds to forms 5, 6, and 7. Conversely, noetic feelings that engage with a cognitive referent (e.g.: "Wait a minute, I feel I know this answer....") can be classified as explicit metacognition. The notion of feelings as metadata can extend to all qualitative experiences that affect cognition. Qualitative feelings of metadata can influence

cognitive process without a representation (implicit) or can affect cognition when represented by knowledge (explicit).

Feelings that engage with their representation become available to higher executive control. For instance, mindfulness training in cognitive therapy can be understood as learning to translate feelings into knowledge to use its properties to gain a better purchase on behavior. Once a feeling is labeled it has been translated into knowledge. This knowledge is what gains access to higher executive control. Yet this representation is not the feeling, which still exists independently. For example, a person with anxiety disorder can be trained to recognize their anxiety and rate its legitimacy. They may be trained to label their emotion (Dobson, 2013) and say to themselves "This is anxiety. Is there anxiety because I'm truly in danger or because the environment is loud?" The subject can then engage further executive functions by finding a quiet area and/or employing calming techniques.

This section has presented as a computational theory of feelings based on the Common Model architecture. Functionally, we have argued that feelings are best modelled as non-propositional representations (metadata) in a noetic buffer. More broadly, we have shown how cognitive architectures can be applied to clarify philosophical issues, particularly in computational theory of mind. This type of work is important as conceptual confusion over issues such as the difference between knowledge and feelings can impede the creation or acceptance of cognitive models involving these phenomena. Finally, by stating our ideas in the terms of a cognitive architecture we have made them computationally unambiguous. Other models are possible, but they should be stated clearly in computational terms, and grounded in a viable cognitive architecture.

DISCUSSION & CONCLUSION

7.0 General discussion

This thesis has attempted to help overcome the two barriers to understanding metacognition — its high abstraction and unstandardized terminology. To clarify metacognition and its terms I have built a hierarchical model of metacognitive processes within a cognitive architecture. From this model I have arrived at two main discoveries:

- i. Two forms of metacognition (implicit, explicit)
- ii. Explicit metacognition holds three levels (object, monitor, instruction),each level is distinguished by its order of cognitive referent.

At the architecture's base, a single production changes the architecture from one cognitive state to another. As metacognition is often defined as "cognition acting on itself," it can be argued that cognition is meta all the way down. Building upward, we see that each form of follows a line of demarcation that carves metacognition at discernible joints. These includes problem solving productions, knowledge retrieval, and cognitive referents. These lines of demarcation are clearly distinguishable in the cognitive architecture. Yet they have remained largely unidentified in the research literature. Hence, this computational model of metacognition brings to light metacognitive distinctions previously unexplored. These results exemplify an overall conclusion of this thesis, that *word definitions are insufficient to describe metacognition*; *computational definitions are necessary to understand the mechanisms of metacognition*. While definitions in psychology, philosophy, and neuroscience detail many important aspects of metacognitive phenomena, they have remained problematic due to inherent limitations of abstract verbal concepts.





7.1 Findings contextualized within the research literature

This section outlines the results of this paper's attempt to computationally model metacognition. A summary can be seen above in **Figure 15**.

Three forms of cognition

Micro cognition: The lowest form of cognition entails productions changing states within the architecture (Forms 1-4). Productions set up the following productions by changing the content of buffers, which can contain noetic metadata.

Literature: This lowest of cognition can be subsumed under the inadequate umbrella definition of "cognition acting on itself" (Flavell, 1976; Dunlosky & Metcalfe, 2009). Yet these cognitive operations do not qualify as metacognition as discussed within the literature.

Implicit-metacognition: This middle form of metacognition involves subsystems of the architecture interacting, *without* knowledge that contains a cognitive referent (Forms 5-7). This includes productions, problem search, and knowledge retrieval. The knowledge used by the architecture only refers to object-level cognition; knowledge "chunks" that guide externally directed actions. Implicit metacognition is absent of cognitive referents. It may include noetic metadata that is not paired with knowledge, such as a feeling of knowing, or certainty.

Literature: This middle form of metacognition is subsumed under the insufficient definition of "one subsystem acting on another subsystem" (Proust, 2014). Neuroscience also tends to focus mainly on the interactions between brain subsystems (the prefrontal cortex, temporal lobe, etc.), and often stalls at the overly-general knowledge referent of "self" (Farb, et al. 2007; Carruthers, 2012). A review by Fleming & Dolan (2012) o the major studies in neuroscience reported they were mainly focused on relationships between brain structures and metacognitive abilities. They

did not distinguish between implicit, explicit, or the order of cognitive referent distinguishing the subcategories of explicit metacognition.

Explicit-metacognition: This highest form of metacognition is distinguished by metaknowledge that contains at least one cognitive referent. The three possible orders of referents determine the three sub-categories of explicit metacognition:

Instruction-level metacognition (*cognitive* \rightarrow *cognitive*)

Monitor-level metacognition (*cognitive* \rightarrow *object*) including represented noetic metadata Object-level metacognition (*object* \rightarrow *cognitive*)

Literature: The highest form of metacognition correlates to incomplete definitions such as "knowledge about one's own knowledge," and "thinking about thinking" (Piaget, 1950; Carruthers, 2011; Perner, 2012). Proust's (2013) definition is more precise, describing a cognitive subsystem that is evaluated or represented by another subsystem. Some research into noetic feelings and judgements have focused on the specific order of cognitive referents (Nelson & Narens, 1992; Koriat et. al., 1999; Arango-Muñoz, 2011). While Nelson & Narens (1992) have identified two levels of metacognition, I have clarified a third: instruction-level metacognition.

Subcategories of explicit metacognition

Explicit metacognition contains three sub-categories, each distinguished by the order of cognitive referent contained in the meta-knowledge.

Object-level (object \rightarrow cognitive): External objects can be known to change one's cognitive state. This is exemplified by *drugs* known to be a stimulant or sedative, *situations* known to create anger or happiness, and by extension, any external *object* with subjective valence (e.g.: money:happy, job:pride, nature:relaxing).

Literature: Object-level metacognition was identified by Nelson & Narens (1992), who described the directionality of cognitive referents.

Monitor-level (cognitive \rightarrow object): A cognitive state is known that directs actions toward external objects (e.g. fatigue:bed, uncertainty:re-read, frustration:control expression). The monitor-level includes represented metadata such as noetic judgements, certainty, and confidence judgments.

Literature: Monitor-level metacognition was also identified by Nelson & Narens (1992). It involves the ability to monitor one's own thinking and learning. This level includes research fields that study educational success (Veenman, Prins & Elshout, 2002) and expert performance (MacIntyre, 2014). The lack of monitoring has been correlated with overconfidence (Kruger & Dunning, 1999), extremists perspectives (Rollwage, Dolan, & Fleming, 2018), and underperformance (Dunning et al., 2003).

Instruction-level (cognitive \rightarrow cognitive): A cognitive state is desired or recognized, and a cognitive representation directs actions that will affect that state. For example, a cognitive state may be desired (e.g.: knowledge) and instructions then direct cognitive actions to obtain this state (scientific method). Or, a cognitive state is recognized (anxiety) and instructions direct cognitive actions to lessen that state (clear mind of thoughts, focus on breathing). Metacognitive instructions assist in raising low-level cognitive functions to higher-level executive control. *Literature:* Instruction-level metacognition includes research fields that study improvements in human cognitive performance (**Figure 11**). These cognitive improvements include memory (Maguire et al., 2003), emotional regulation (Vukman, & Licardo, 2010), and reasoning (Leevers & Harris, 1999).

7.2 Potential value of findings

These findings posit a unified set of mechanisms for describing metacognition, along with an integrated set of computational terminology. These findings are of significant potential value for guiding clinical research, cognitive modeling, and designs for artificial intelligence.

This thesis accords with the scientific principles of parsimony and unification. Rather than separately identifying metacognitive phenomena, I have sought to unify metacognition under a common set of mechanisms. I have accounted for the varied and disputed views of metacognition by integrating them systematically. This presents the first attempt to unify metacognitive literature by order of referent. Hence, this thesis makes a significant advance toward Schraw's' (2000) challenge to build a unified theory of metacognition. Additionally, I have created a list of computational vocabulary that integrates and standardizes terminology.

In guiding psychological research, these findings can help to clarify the metacognitive phenomena being investigated. A current problem is clinical researchers is that metacognition is often described and studied too generally. This creates difficulties for isolating variables that are independent, dependent, and extraneous. The blurriness of the metacognitive concepts has hindered researchers in discerning its salient properties. Previous psychological experiments have both elucidated and confused the subject by creating ad-hoc terms, or by describing the same phenomena differently. This confusion has resulted from the subject's high abstraction and resistance to definition. This thesis helps clarify metacognitive concepts that will aid researchers in targeting their investigations. By identifying the various forms, referents, and properties of metacognition, researchers can more accurately distinguish variables that are independent, dependent, extraneous, and confounding.

Additionally, a clearer theoretical framework may help those seeking to improve

metacognitive skills personally, or within their workforce training systems. This thesis provides a clearer picture of how individuals or institutions may benefit from metacognitive instructions to enhance cognitive performance in memory, emotional regulation, focus, and reasoning.

Further, this thesis holds potential for assisting cognitive modellers in building more faithful depictions of human cognition. It helps inform the construction of cognitive models of metacognition in ACT-R and other cognitive architectures. Different models that contain variations of the principles herein can be tested against each other, or against their human counterparts. Models with different orders of referents can predict the timing of humans given the same tasks. Computational models with differing instantiations of metacognition can be compared systematically. Different points of view can be constructed computationally and tested to generate specific, measurable outcomes.

This computational model of metacognition can also provide insight for those building autonomous systems required to self-represent their internal operations (Schmill et. al., 2008). In addition, it can aid in designing systems that must represent human thinking, such as self-driving cars and artificial attendants. A functional metacognitive theory of mind is essential to building robust, explainable A.I. and flexible machine-human interaction.

Future research will advance what this thesis presently lacks: a functioning, integrated model of metacognition. The coded models presented in the appendix individually depict the metacognitive phenomena in forms 1-8. At present, these models are built separately. Yet the question naturally arises, how might we build domain-general models that produce the variety of phenomena observed within the research literature? What is required for an architecture to sense which type of metacognition is appropriate for the situation? Do we perhaps need a better architecture? Future research involves building a functioning, unified model of metacognition.

7.3 Conclusion

This thesis has presented a novel method of representing metacognition computationally. The intention of this work is to help illuminate the mechanisms giving rise to metacognition's unique effects. I have endeavoured to help overcome the two barriers to advancing research of metacognition — its abstraction and disputed terminology — by building a hierarchical model within a cognitive architecture. Beginning with a foundation of computational axioms, I have built increasingly complex forms of metacognition to encompass much of the phenomena observed in the research literature. Each form of metacognition has been well-defined, and has attempted to supplement the inadequate word definitions of previous researchers. This paper asserts that apprehending the mechanisms of metacognition requires a computational lens.

The major finding of this thesis are — two major forms of metacognition (implicit, explicit), three orders of referents (object, monitor, instruction) and a way of incorporating noetic feelings as metadata. This paper is the first to categorize metacognitive phenomena by its order of metaknowledge referent.

These findings help advance a unified account of metacognition, an integration of terminology, and provide support for clinical research, cognitive modeling, and artificial intelligence design. This paper also proffers direction for those seeking to benefit from metacognitive training individually, or as an organization. Future research will work toward building an integrated model of metacognition to generate the diversity of phenomena observed in the literature. Herein, this author has provided original work intended to help advance a unified theory of metacognition.

8.0 Addendum: code for metacognitive models

The working ACT-R models of metacognitive forms 1-8 are currently held in an online repository for public viewing https://github.com/BrendanCS/thesis_models/.

REFERENCES

Alston, W. (1969). Feelings. The Philosophical Review 78(1): 3-34.

- Anderson, J. (1993). Rules of the Mind. Lawrence Erlbaum Associate, Inc. New Jersey.
- Anderson, J. R. (1996). Implicit memory and metacognition: Why is the glass half-full? In L. M. Reder (Ed.) Implicit memory and metacognition, 123-136. *Mahwah, NJ: Erlbaum*.
- Anderson, J.R., & Lebiere, C. (1998). The Atomic Components of Thought. *Mahwah: Lawrence Erlbaum* Associates
- Anderson, J. R., Betts, S. A., Ferris, J. L., & Fincham, J. M. (2011). Cognitive and metacognitive activity in mathematical problem solving: Prefrontal and parietal patterns.Cognitive, Affective, and Behavioral Neuroscience, 11, 52-67.
- Anderson, J., & Fincham, J. (2014). Extending problem-solving procedures through reflection. *Cognitive Psychology*, 74:1–34.
- Anderson, J., Qin, Y., Stenger, V., & Carter, C. (2004). The relationship of three cortical regions to an information-processing model. Journal of Cognitive Neuroscience, 16 (4), 637-653.
- Angel Valenzuela, M. (2019). What's new in metacognition? Review of the concept, its components and related terms. Educação e Pesquisa, 45, e187571. Epub April 08, 2019.
- Arango-Muñoz, S., Michaelian, K. (2014). Epistemic Feelings and Epistemic Emotions (Focus Section). Philosophical Inquiries.
- Augustyn, J. S., & Rosenbaum, D. A. (2005). Metacognitive control of action: Preparation for aiming reflects knowledge of Fitts's law. Psychonomic Bulletin & Review, 12(5), 911–916.
- Adey, P.S.; Shayer, M. 1993. An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum. Cognition and instruction(Mahwah, NJ), vol. 11, no. 1, p. 1–29.1994. Really raising standards: cognitive intervention and academic achievement. London, Routledge
- Baird, B., Mrazek, M. D., Phillips, D. T., & Schooler, J. W. (2014). Domain-specific enhancement of metacognitive ability following meditation training. Journal of Experimental Psychology: General, 143(5), 1972-1979.
- Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and Lateral Networks in Anterior Prefrontal Cortex Support Metacognitive Ability for Memory and Perception. The Journal of Neuroscience, 33(42), 16657 LP-16665.
- Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., &Narens, L.(1999). An integrative system of metamemory components involved in retrieval. In D. Gopher & A. Koriat (Eds.), Attention and performance XVII: Cognitive regulation of performance:Interaction of theory and application(pp. 287-314). Cambridge, MA: MIT press.
- Barrett, L. F. (2006). Are Emotions Natural Kinds? Perspectives on Psychological Science, 1(1), 28–58.
- Bechtel, W. (1994). Levels of Description and Explanation in Cognitive Science. Minds and Machines, 4, 1-25.

- Bechtel, W. (1998). Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge in Cognitive Science. Cognitive Science, 22(3), 295–318.
- Bogdan, R. J. (2000). Minding minds: Evolving a reflexive mind by interpreting others. Cambridge, MA, US: The MIT Press.
- Borst, J.P. & Anderson, J. (2016). A step-by-step tutorial on using the cognitive architecture ACT-R in combination with fMRI data. *Journal of Mathematical Psychology*.
- Brentano, F. C. (1874). Psychologie von empirischen Standpunkt, edited by O. Kraus. Leipzig: Felix
 Meiner. English translation Psychology from an Empirical Standpo int, translated by A.C. Rancurello,
 D.B. Terrell, and L.L. McAlister. London: Routledge, 1973
- Brown, R. & Mcneill, D. (1966). The "Tip of the Tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior. 5. 325-337.
- Brook, A. (2009), Introduction: Philosophy in and Philosophy of Cognitive Science. Topics in Cognitive Science, 1: 216-230. doi:10.1111/j.1756-8765.2009.01014.
- Butler, Deborah & Winne, Philip. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. Review of Educational Research REV EDUC RES. 65. 245-281. 10.2307/1170684.
- Campos, L., & Oliveira, R. (2013). "A Comparative Analysis of Methodologies for Automatic Design of Artificial Neural Networks from the Beginnings until Today," 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence.
- Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. Mind and Language, 23 (1):58-89.
- Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behavior Brain Science*. 32, 121–138.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.
- Ćelikovski S. (1987). Antropomotoryka. SPN, Praha.
- Chalmers, D. (1996). The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press.
- Chambers, R., Lo, B.C.Y. & Allen, N.B. (2008). The Impact of Intensive Mindfulness Training on Attentional Control, Cognitive Style, and Affect. Cogn Ther Res 32, 303–322.
- Cicero. (2001). On the Ideal Orator. Trans. James M. May, Oxford: Oxford University Press.
- Cole M.W., Reynolds J.R., Power J.D., Repovs G., Anticevic A., Braver T.S. (2013). "Multi-task connectivity reveals flexible hubs for adaptive task control". *Nature Neuroscience*; 2013 Sep;16(9): 1348-55. doi:10.1038/nn.3470.
- Colman, A. (2015) "Metacognition". *A Dictionary of Psychology*. Oxford Paperback Reference (4 ed.). Oxford: Oxford University Press. p. 456. ISBN 9780199657681.
- Cooper, R., Fox, J. Farringdon, J., Shallice, T. (1996). A systematic methodology for cognitive modelling, Artificial Intelligence, Volume 85, Issues 1–2, 1996, Pg 3-44
- Cross, D. R. & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, 80(2), 131-142.
- Damasio, A. (2015). Scientific American 2019. Oct. 16 Nation, 294(20), 11-18. Retrieved from MAS Ultra.

- Daw, Nathaniel & Niv, Yael & Dayan, Peter. (2006). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience. 8. 1704-11.
- Demetriou, A., Efklides, A., & Platsidou, M. (1993). The architecture and dynamics of developing mind: Experiential structuralism as a frame for unifying cognitive developmental theories. *Monographs of the Society for Research in Child Development*, 58, Serial Number 234.
- de Sousa, R. Epistemic feelings. In: GeorgBrun,UlviDo^{*}guo^{*}glu, and Dominque Kuenzle, eds., Epistemology and emotions. Ashgate, 2008.

Dennett, Daniel C. (1978). Brainstorms . MIT Press.

Dennett, D. C. (1991). Consciousness Explained. New York: Little, Brown and Company.

- Dennett, D.C. (2000). "Making tools for thinking." In Metarepresentations: A multidisciplinary perspective.
- Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process . Boston: Heath.
- Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20, 391*
- Dobson, K. S. (2013). The Science of CBT: Toward a Metacognitive Model of Change? Behavior Therapy, 44(2), 224–227.
- Dokic, J. (2012). Seeds of self-knowledge: Noetic feelings and metacognition. Foundations of Metacognition. 10.1093/acprof:oso/9780199646739.003.0020.
- Donoso M, Collins AG, Koechlin E, (2014) Foundations of human reasoning in the prefrontal cortex. Science 344:1481–1486. doi:10.1126/science.1252254 pmid:24876345.
- Dougherty, M. R. P. (2001). Integration of the ecological and error models of overcon dence using a multiple-trace memory model. Journal of Experimental Psychology: General, 130, 579–599.
- Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R. A. Bjork (Eds.), Handbook of metamemory and memory (p. 11–28). Psychology Press.
- Dunlosky, J. and Hertzog, C. (2000). Updating Knowledge about Encoding Strategies: A Componential Analysis of Learning about Strategy Effectiveness from Task Experience. *Psychology and Aging*, 15(3), 462-474.
- Dunlosky, J., & Metcalfe, J. (2008). Metacognition: A textbook for cognitive, educational, life- span and applied psychology. ousand Oaks, CA.
- Dunlosky, J., & Metcalfe, J. (2009). Metacognition. Metacognition. Thousand Oaks, CA, US: Sage Publications, Inc.
- Dunning D., Johnson K., Ehrlinger J., Kruger, J. (2003) Why people fail to recognize their own competence. Current Directions in Psychological Science. 2003;12:83–87.
- Envick, B.R.. (2014). Achieiving entrepreneurial success through passion, vision & courage: A cognitive model for developing entrepreneurial intelligence. Academy of Entrepreneurship Journal. 20. 55-74.
- Farb, N. A. S., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Zainab, F., et al. (2007). Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. SCAN 2, 313–322.

CLARIFYING METACOGNITION

- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Lawrence Erlbaum. p.232
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Feist, G.J. (2006). The psychology of science and the origins of the scientific mind. New Haven, CT: Yale University Press.
- Fleming S. & Dolan R. (2012). The neural basis of metacognitive ability. Phil. Trans. R. Soc. B 367, 1338–1349.
- Fisher, R. (1998) Thinking About Thinking: Developing Metacognition in Children, Early Child Development and Care, 141:1, 1-15, DOI: 10.1080/0300443981410101
- Fleming S.M., Daw, N.D. (2017). Self-evaluation of decision-making: a general bayesian framework for metacognitive computation. *Psychology Review*, 124:91–114.
- Fletcher, L. & Carruthers, P. (2012). Metacognition and reasoning. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 367. 1366-78. 10.1098/rstb.2011.0413.
- Fodor, J. (1983). Modularity of Mind: An Essay on Faculty Psychology. Cambridge, Massachusetts: MIT Press.
- Fodor, J. (2000). The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.
- Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. Journal of Applied Psychology, 83(2), 218–233. https://doi.org/10.1037/0021-9010.83.2.218
- Franklin, S., (2007). A foundational architecture for artificial general intelligence. In B. Goertzel & P. Wang (Eds.). Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms. Proceedings of the AGI Workshop 2006. *Frontiers in Artificial Intelligence and Applications*, vol. 157, pp. 36-54. IOS Press: Amsterdam, The Netherlands.
- Frese, M., & Gielnik, M. M. (2014). The psychology of entrepreneurship. Annual Review of Organizational Psychology and Organizational Behavior, 1(1), 413-438. doi:10.1146/annurevorgpsych-031413-091326
- Frith C., Frith U. (2012) Mechanisms of social cognition. Ann. Rev. Psychol. 63, 287–313.
- Gabriel, G., & Lisa, G., & Robert, T., (2019). Improving metacognition: A comparison of interventions. Applied Cognitive Psychology. 10.1002/acp.3556.
- Garofalo, J. & Lester, F. K. (1985). Metacognition, Cognitive Monitoring, and Mathematical Performance, *Journal for Research in Mathematics Education Vol. 16, No. 3* (May, 1985), pp. 163-176
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2013). Cognitive neuroscience: the biology of the mind. WW Norton & Company.
- Gelder, T. (1995). Mind as motion: Explorations in the dynamics of cognition. Mind as motion: Explorations in the dynamics of cognition. Cambridge, MA, US: The MIT Press.
- Gelles, D. (2015). Mindful work: how meditation is changing business from the inside out. Boston: Houghton Mifflin Harcourt.
- Georghiades, P. (2004). From the General to the Situated: Three Decades of Metacognition, *International Journal of Science Education*, *26*(3), 365-383.

- Gil, J., Howse, J., Kent, S., & Taylor, J. (2000). "Projections in Venn-Euler diagrams," Proceeding 2000 IEEE International Symposium on Visual Languages, Seattle, WA, USA, pp. 119-126.
- Goldberg, S., Tucker, R., Greene, P., Davidson, R., Wampold, B., Kearney, D., & Simpson, T. (2018). Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. Clinical Psychology Review, 59, 52–60.
- Goldin, D., & Wegner, P. (2005). The Church-Turing Thesis: Breaking the Myth. In S. B. Cooper, B. Löwe,
 & L. Torenvliet (Eds.), New Computational Paradigms (pp. 152–168). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Gopnik, A. (1993). "How we know our minds: The illusion of first-person knowledge of intentionality." *Behavioral and Brain Sciences*, 16, 1-14.
- Halamish, Vered & Bjork, Robert. (2011). When Does Testing Enhance Retention? Journal of experimental psychology. Learning, memory, and cognition. 37. 801-12.
- Hart J. (1965). Memory and the feeling-of-knowing experience. J. Educ. Psychol. 56, 208-216.
- Haynie, M., & Shepherd, D. A. (2009). A measure of adaptive cognition for entrepreneurship research. Entrepreneurship Theory and Practice, 33(3), 695-714. doi:10.1111/j.1540-6520.2009.00322.x
- Henmon, V. (1911). The relation of the time of a judgment to its accuracy. Psychol. Rev. 18, 186.
- Higham P. A. (2007) No special K! A signal detection framework for the strategic regulation of memory accuracy.
- James, W. (1890). The Principles of Psychology. Dover Publications.
- James, W. (1897). The will to believe and other essays in philosophy. New York/London: Longmans Green and Company.
- James, W. (1892). Psychology: Briefer course. New York: Henry Holt.
- Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about read- ing: Issues in definition, measurement, and instruction. *Educational Psychologist, 22,* 255–278.
- John, B., Vera, A., & Newell, A. (1994). Towards real-time GOMS: a model of expert behaviour in a highly interactive task, Behaviour & Information Technology, 13:4, 255-267
- Jowett, B. (2018). Plato Laws. Scribe Publishing. Melbourne, Australia.
- Keleman, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92–107.
- Kelly, M. A., Arora, N., West, R., & Reitter, D. (2019). High dimensional vector spaces as the architecture of cognition. https://doi.org/10.31234/osf.io/ryvg2
- Khoury, B., Knäuper, B., Schlosser, M., Carrière, K., & Chiesa, A. (2017). Effectiveness of traditional meditation retreats: A systematic review and meta-analysis. Journal of Psychosomatic Research, 92, 16–25.
- Ko Y., Lau H. (2012) A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Phil. Trans. R. Soc. B* 367.
- Koriat, A. (1993). How do we know that we know) The accessibility model of the feeling of knowing. Psychological Review, 100,609-639.

- Koriat, A. (1997). Monitoring One's Own Knowledge during Study: A Cue-Utilization Approach to Judgments of Learning. Journal of Experimental Psychology: General. 126. 349-370.
- Kralik, J., Lee, J. H., Rosenbloom P., Jackson, P. C., Epstein, S., Romero, O. J., Sanz, R., Larue, O., Schmidtke, H. R., Lee, S. W., McGreggor, K., (2018). Metacognition for a Common Model of Cognition. Procedia Computer Science: Proceedings of 2018 Fall AAAI Symposium on the Common Model of Cognition, Post Conference Proceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures.
- Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. Educational Studies in Mathematics, 49(2), 225–250.
- Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic perfor mance. Cognitive Development, 13, 227-247.
- Kuhn, D. & Dean, D. (2004). A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*(4), 268-273.
- Laird, J. E. (2012). The SOAR cognitive architecture. MIT Press.
- Laird, J. & Lebiere, C., & Rosenbloom, P. (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*. 38. 13. 10.1609/aimag.v38i4.2744.
- Leader, W. S. (2008). Metacognition among Students Identified as Gifted or Nongifted Using the Discover Assessment. *Unpublished Doctoral Dissertation*. Tucson, AZ: Graduate College of the University of Arizona.
- Lee, D. J., Kulubya, E., Goldin, P., Goodarzi, A., & Girgis, F. (2018). Review of the Neural Oscillations Underlying Meditation. Frontiers in Neuroscience, 12, 178.
- Leevers, H. J., & Harris, P. L. (1999). Persisting Effects of Instruction on Young Children's Syllogistic Reasoning with Incongruent and Abstract Premises. Thinking & Reasoning, 5(2), 145–173.
- Leslie, A.M. (1987). Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review*. 94. 412-426. 10.1037//0033-295X.94.4.412.
- Livingston, J., (1996). *Effects of Metacognitive Instruction on Strategy Use of College Students*. Buffalo, NY: Unpublished manuscript, state University of New York at Buffalo.
- Livingston, J. A. (1997). Metacognition: An Overview. http://www.gse.buffalo.edu/fas/shuell/CEP564/ Metacog.htm, Eri im Tarihi: 24.06.2009.
- Lovett, M., Daily, L., & Reder, L. (2000). A source activation theory of working memory: cross-task prediction of performance in ACT-R. Cognitive Systems Research, 1(2), 99–118
- Makhbul, Z. M., & Hasun, F. M. (2011). Entrepreneurial success: An exploratory study among entrepreneurs. International Journal of Business and Management, 6(1), 116.
- Mangan, B. (2001). Sensation's ghost: the non-sensory "fringe" of consciousness. Psyche (Stuttg) 7.
- MacIntyre, T. E., Igou, E. R., Campbell, M. J., Moran, A. P., & Matthews, J. (2014). Metacognition and action: a new pathway to understanding social and cognitive aspects of expertise in sport. Frontiers in Psychology, 5, 1155. https://doi.org/10.3389/fpsyg.2014.01155
CLARIFYING METACOGNITION

- Maguire, E., Valentine, E., Wilding, J. (2003). Routes to remembering: the brains behind superior memory. Nat Neurosci 6, 90–95 doi:10.1038/nn988
- Mathan, S. A. & Koedinger, K. R. (2005) Fostering the Intelligent Novice: Learning from Errors with Metacognitive Tutoring. *Educational Psychologist* 40(4), 257-265.
- Marewski, J. N., & Link, D. (2014). Strategy selection: An introduction to the modeling challenge. Wiley Interdisciplinary Reviews: Cognitive Science, 5, 39-59. DOI: 10.1002/wcs.1265
- Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: Implications for training effectiveness. Personnel Psychology, 46.
- Martinez, M. E. (2006). What is metacognition? Phi Delta Kappan, 696-699.
- Matthews, M. (2009). History, philosophy, and science teaching: The new engagement. Asia-Pacific Forum on Science Learning and Teaching, 10.
- McCulloch, W.S., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115–133. doi:10.1007/BF02478259
- McDowell, J.H. (1994). Mind and World. Harvard University Press.
- McLeod, L. (1997). Young children and metacognition: Do we know what they know they know? And if so, what do we do about it? *Australian Journal of Early Childhood*, 22(2), 6-11.
- Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsako amnesia. Psychological Review, 100, 3–22.
- Metcalfe J., Van Snellenberg J. X., DeRosse P., Balsam P., Malhotra A. K. (2012). Judgments of agency in schizophrenia: an impairment in autonoetic metacognition. *Phil. Trans. R. Soc.* B 367, 1391–1400.
- Metcalfe, J. (1994b). A computational modeling approach to novelty monitoring, metacognition, and frontal lobe dysfunction. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 137-156). Cambridge, MA: MIT Press.
- Metcalfe, J., and Shimamura, A. P. (1996). *Metacognition: Knowing About Knowing*. Cambridge, MA: MIT Press.
- Metcalfe, J., & Son, L. K. (2012). Anoetic, noetic, and autonoetic metacognition. In M. Beran, J. L. Brandl, J. Perner, and J. Proust (Eds.). Foundations of Metacognition (pp. 289-301). Oxford University Press
- Metcalfe, J., & Son, L.(2013). Anoetic, noetic, and autonoetic metacognition. Foundations of Metacognition
- Meyer, B. (1988). Object-Oriented Software Construction. New York: Prentice Hall. Stevens, Christopher & Taatgen, Niels & Cnossen, Fokie. (2015). Metacognition in the Prisoner's Dilemma.
- Morales, J., Lau, H., Fleming S.H. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex Journal of Neuroscience 8 March 2018, 2360-17; DOI: 10.1523/JNEUROSCI.2360-17.
- Moritz, S. Hautzinger, Hofmann M. P. Lysaker, P. (2018) Going meta on metacognitive interventions Pages 739-741. <u>https://doi.org/10.1080/14737175.2018.1520636</u>
- Moshman, D. (2018). Metacognitive Theories Revisited, *Educational Psychology Review*, June 2018, Volume 30, Issue 2, pp 599–606.
- Nelson, Thomas & Leonesio, R. (1988). Allocation of Self-Paced Study Time and the "Labor-in-Vain Effect". Journal of experimental psychology. Learning, memory, and cognition. 14. 676-86.

- Nelson, T. O., and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motiv.* 26, 125–141. doi: 10.1016/S0079-7421(08)60053-5
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe, & A. J. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 1–26). Cambridge, MA: MIT Press.
- Nelson, T. O. (Ed.). (1992). Metacognition: Core readings. Needham Heights, MA, US: Allyn & Bacon.
- Newell, A. (1973). You can't play 20 questions with nature and win. In Chase, W.G. (Ed.), Visual information processing (pp. 283–308). San Diego, CA: Academic Press.
- Newell, A., & Simon, H. A. (1972). Human Problem Solving. Englewood Cliffs: Prentice Hall.
- Newell, A. (1990). Unified Theories of Cognition. Cambridge: Harvard University Press.
- Nikola K. Kasabov (1995) Hybrid Connectionist Fuzzy Production System: Towards Building Comprehensive AI, Intelligent Automation & Soft Computing, 1:4, 355-364.
- Olano, H., Diana, K., & Tannenbaum, S., & Ashwin, M., Annane, D. & Lee, D. (2015). Engagement in Mindfulness Practices by U.S. Adults: Sociodemographic Barriers. Journal of alternative and complementary medicine (New York, N.Y.). 21. 100-102. 10.1089/acm.2014.0269.
- O'Neil, H.F., & Spielberger, C.D (Eds.). (1979). Cognitive and affective learning strategies. New York: Academic Press.
- Osman, Mohamed & Hannafin, Michael. (1992). Metacognition research and theory: Analysis and implications for instructional design. Educational Technology Research and Development. 40. 83-99.
- Panksepp, J. (1998). Affective neuroscience: The foundations of human and animal emotions. New York: Oxford University Press.
- Paris, S. G. & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. Remedial and Special Education, 11(6), 7-15.
- Peirce, C. S., and Jastrow, J. (1885). On small differences in sensation. Mem. Natl. Acad. Sci. 3, 73-83.
- Peters, E. E. (2007). The Effect of Nature of Science Metacognitive Prompts on Science Students' Content and Nature of Science Knowledge, Metacognition, and Self-Regulatory Efficacy. *Unpublished Doctoral Dissertation*. Fairfax, VA: Graduate Faculty of George Mason University.
- Perner, J. (1988). "Developing semantics for theories of mind: From propositional attitudes to mental representation: in Astington, J., Harris, P. & Olson, D. (eds.) *Developing Theories of Mind*, Cambridge: Cambridge University Press.
- Pesut, D. J. and Herman, J. (1992), Metacognitive Skills in Diagnostic Reasoning: Making the Implicit Explicit. International Journal of Nursing Terminologies and Classifications, 3: 148-154.
- Piaget, J. (1950). The Psychology of Intelligence, Harcourt and Brace, New York.
- Pinker, S. (1997). How the Mind Works. New York: Norton.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41(4), 219–225.
- Pirolli, P., & Card, S. (1999). Information foraging. Psychological Review, 106(4), 643-675.
- Pouget A, Drugowitsch J, Kepecs A, (2016) Confidence and certainty: distinct probabilistic quantities for different goals. Natu Neuroscience 19:366–374. doi:10.1038/nn.4240 pmid:26906503.

- Prezenski, S., Brechmann, A., Wolff, S., & Russwinkel, N. (2017). A Cognitive Modeling Approach to Strategy Formation in Dynamic Decision Making. Frontiers in Psychology. Retrieved from https:// www.frontiersin.org/article/10.3389/fpsyg.2017.01335
- Prinz, J. (2006). Canadian Journal of Philosophy, Volume 36, Supplement Col. 32, pp. 137-160.
- Proust, Joelle (2009). Is there a sense of agency for thought? In Lucy O'Brien & Matthew Soteriou (eds.), Mental Actions. Oxford University Press.
- Proust, J. (2013). The Philosophy of Metacognition: Mental Agency and Self-Awareness, Oxford Press.
- Proust, J. (2014). The Philosophy of Metacognition. *Philosophy and Phenomenological Research*. Vol. 89, Issue3 November 2014 Pages 703-709.
- Quine, W. V. (1969). Ontological Relativity and Other Essays. Columbia University Press.
- Rabinowitz, N.C., Perbet, F., Song, H.F., Zhang, C., Eslami, S.M., & Botvinick, M.M. (2018). Machine Theory of Mind. ArXiv, abs/1802.07740.
- Redding, R. E. (1990), Metacognitive Instruction: Trainers Teaching Thinking Skills. Performance Improvement Quarterly, 3: 27-41.
- Reder, L., & Ritter, F. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18 435-451.
- Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), Implicit memory and metacognition
- Reitter, D. (2010). Metacognition and multiple strategies in a cognitive model of online control. *Journal* of Artificial General Intelligence, 2(2), 20-37.
- Reitter, D., Juvina, I., Stocco, A., & Lebiere, C. (2010). Resistance is futile: Winning lemonade market share through metacognitive reasoning in a three-agent cooperative game. *In Proceedings of the 19th Conference on Behavioral Representation in Modeling and Simulation* (BRIMS). Charleston, S.C.
- Rescorla, M. (2015). The Computational Theory of Mind.
- Roll, I., Baker, R. S., Aleven, V., & Koedinger, K. (2004). A Metacognitive ACT-R Model of Students' Learning Strategies in Intelligent Tutoring Systems. *In Proceedings the Seventh International Conference of Intelligent Tutoring Systems* (pp. 854-856) Lecture Notes in Computer Science 3220.
- Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive Failure as a Feature of Those Holding Radical Beliefs. Current Biology, 28(24), 4014–4021.e8.
- Rosenbloom, Paul & Demski, Abram & Ustun, Volkan. (2016). The Sigma Cognitive Architecture and System: Towards Functionally Elegant Grand Unification. Journal of Artificial General Intelligence. 7. 10.1515/jagi-2016-0001.
- Rosenfeld, A., & Richardson, A. (2019). Explainability in human–agent systems. Autonomous Agents and Multi-Agent Systems, 1-33.
- Schneider, W., Visé, M., Lockl, K., Nelson, T. (2000). Developmental trends in children's memory monitoring Evidence from a judgment-of-learning task. Cognitive Development. 15. 115-134.
- Scassellati, B. (2002). Theory of Mind for a Humanoid Robot. Autonomous Robots, 12, 13-24.
- Schmill, M.D., Oates, T., Anderson, M.L., Josyula, D.P., Perlis, D., Wilson, S., & Fults, S. (2008). The Role of Metacognition in Robust AI Systems.

- Schneider, W. & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In Perfect, T. & Schwartz, B. (Eds.), *Applied metacognition*. Cambridge, UK: Cambridge University Press.
- Schraw, G. (2000). Assessing metacognition: implications of the buros symposium. In G. Schraw and J. Impara (Eds.), Issues in the measurement of metacognition, pp 297-322). Lincoln, NB: University of Nebraska Press.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive aware- ness. Contemporary Educational Psychology, 19, 460–475.
- Schraw, G. & Moshman, D. (1995). Metacognitive theories. Educational Psychology Review, 7(4), 351-371.
- Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, 90, 236–248.
- Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu and other metacognitive oddities. In J. Dunlosky & S. Tauber (Eds.),Oxford Handbook of Metamemory(pp. 95–108). aScott, B., & Levy, M. (2013). Metacognition: examining the components of a fuzzy concept
- Searle, J. (1984), Minds, Brains and Science: The 1984 Reith Lectures, Harvard University Press.
- Searle, J. R. (1992). The rediscovery of the mind. Cambridge, MA: MIT Press.
- Shallice T. (1988). From neuropsychology to mental structure. Cambridge, UK: Cambridge University Press.
- Shallice T., Burgess P. (1996). The domain of supervisory processes and temporal organization of behaviour. Phil. Trans. R. Soc. B 351, 1405–1411; discussion 1411–1412.
- Shoenfield, J. R. (1977). Axioms of Set Theory. In J. Barwise (Ed.), Handbook of Mathematical Logic (Vol. 90, pp. 321–344). Elsevier.
- Sikström, S., & Jönsson, F. (2005). A model for stochastic drift in memory strength to account for judgments of learning. Psychological Review, 112, 932–950.
- Simon, H.A., Wallach, D. (1999). Cognitive modeling in perspective. Kognit. Wiss. 8, 1–4 doi:10.1007/ BF03354931
- Smith J. D., Couchman J. J., Beran M. J. (2012) The highs and lows of theoretical interpretation in animal-metacognition research. *Phil. Trans. R. Soc.* B 367, 1297–1309.
- Singer, M., & Tiede, H. L. (2008). Feeling of knowing and duration of unsuccessful memory search. Memory and Cognition, 36, 588-597.
- Spence, J.D. (1984). The memory palace of Matteo Ricci, Viking Penguin, New York NY
- Squire, R., & Zola, M. (1996). Structure and function of declarative and nondeclarative memory systems.
- Steinbach, J. C. (2008). The Effect of Metacognitive Strategy Instruction on Writing. Unpublished Doctoral Dissertation. Lexington, KY: The Graduate School of University of Kentucky.
- Son, L. K., & Kornell, N. (2005). Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In Terrace, H.S. & Metcalfe, J. (Eds.), The Missing Link in Cognition: Origins of Self-Knowing Consciousness. Oxford University Press.

- Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. Psychological Science, 3, 315–316.
- Standish, P. (1992), In Praise of the Cognitive Emotions. Journal of Philosophy of Education, 26: 117-119. developmental inquiry. American Psychologist, 34, 906–911.
- Stocco, A., Laird, J., Lebiere, C., & Rosenbloom, P., (2018). Empirical Evidence from Neuroimaging Data for a Standard Model of the Mind.
- Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglias role in cognitive coordination. Psychological Review, 117(2), 541.
- Strawson, P. F. (1948). Truth. Analysis 9 (6): 83-97.
- Sun, R. (2004). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. *Ron Sun (Ed.), Cognition and Multi-Agent Interaction.* Cambridge University Press: New York.
- Sun, R., Xi Zhang, & Mathews, R. (2006). Modeling meta-cognition in a cognitive architecture. *Cogn. Syst. Res.* 7, 4, 327-338. DOI=http://dx.doi.org/10.1016/j.cogsys.2005.09.001
- Sun, R. (2006). Prolegomena to integrating cognitive modeling and social simulation, (Ed.), Cognition and multi-agent interaction: From cognitive modeling to social simulation, Cambridge University Press, Cambridge
- Sun, R. (2007). "Cognitive social simulation incorporating cognitive architectures," Intelligent Systems, vol. 22, no. 5, pp. 33–39.
- Taren, A., Gianaros, P., Greco, C., Lindsay, E., Fairgrieve, A., Brown, K., Creswell, J. D. (2017). Mindfulness Meditation Training and Executive Control Network Resting State Functional Connectivity: A Randomized Controlled Trial. Psychosomatic medicine, 79(6), 674–683.
- Tempelaar, D. T. (2006). The Role of Metacognition in Business Education. Industry and Higher Education, 20(5), 291–297. <u>https://doi.org/10.5367/00000006778702292</u>
- Thompson, V., & Johnson, S., (2014). Conflict, metacognition, and analytic thinking. Thinking & Reasoning. 20. 215-244. 10.1080/13546783.2013.869763.
- Townsend, J., Keedwell, E., & Galton, A. (2014). Artificial Development of Biologically Plausible Neural-Symbolic Networks. Cognitive Computation, 6(1), 18–34.
- Trout, J. D. (2007), The Psychology of Scientific Explanation. Philosophy Compass, 2: 564-591.
- Ullman, M. T., & Lovelett, J. T. (2016). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. Second Language Research, 34(1), 39–65. https://doi.org/10.1177/0267658316675195
- Veenman, M. V. J., & Verheij, J. (2003). Technical students' metacognitive skills: relating general vs. specific metacognitive skills to study success. *Learning and Individual Differences*, 13, 259–272.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3–14.
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, 14, 89–109.
- Vukman, K. B., & Licardo, M. (2010). How cognitive, metacognitive, motivational and emotional selfregulation influence school performance in adolescence and early adulthood. Educational Studies, 36.

CLARIFYING METACOGNITION

- Vygotskxy, L. S. (1962). The Problem and the Approach. In Thought and Language (pp. 1–11). Cambridge: M.I.T. Press, Massachusetts Institute of Technology.
- Wang, M.C., Haertel, G.D., & Walberg, H.J. (1990). What influences learning? A content analysis of review literature. Journal of Educational Research, 84(1), 30-43
- Wellman, H. M. (1983). Metamemory Revisited. In M. T. H.Chi ed. Contributions to Human Development. Vol. 9(Trends in memory development research). Basel, Switzer-land: S. Karger, AG.
- Wells, A. (2008). Metacognitive Therapy: Cognition Applied To Regulating Cognition. Behavioural and Cognitive Psychotherapy, 36(6), 651-658. doi:10.1017/S1352465808004803
- West, R. L., & Conway-Smith, B. (2019). Put Feeling into Cognitive Models: A Computational Theory of Feeling. *Proceedings of ICCM 2019 17th International Conference on Cognitive Modelling*, 295-300. Retrieved from https://iccm-conference.neocities.org/2019/proceedings/ICCM2019Proceedings.pdf
- West, R.L., & MacDougall, K. (2014). The macro- architecture hypothesis: Modifying Newell's system levels to include macro-cognition. *Biologically Inspired Cognitive Architectures*, 8: 140-149.
- West, R. L., & Young, J. (2017). Proposal to add emotion to the standard model. 2017 AAAI Fall Symposium Technical Report Volume 17, Symposium 6: A Standard Model of the Mind
- Whitehead, A. (1967). Science and the Modern World, New York: Macmillan.
- Whitebread, D., Coltman, P., Pasternak, P., Sangster, C., Grau, V., Bingham, S. et al. (2009). The development of two observational tools for assessing metacognition and self regulated learning in young children. *Metacognition and Learning*.
- Whittlesea, B. (1993). Illusions of familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(6), 1235-1253.
- Wiltshire, T. J., Rosch, K., Fiorella, L., & Fiore, S. M. (2014). Training for Collaborative Problem Solving: Improving Team Process and Performance through Metacognitive Prompting. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 1154–1158.
- Yanqun, Z. (2019. International Forum of Teaching and Studies Vol. 15, No. 1, The Significance and Instruction of Metacognition in Continuing Education.
- Zepeda, C.D., Richey, J.E., Ronevich, P., & Nokes-Malach, T.J. (2015). Direct Instruction of Metacognition Benefits Adolescent Science Learning, Transfer, and Motivation: An In Vivo Study.
- Zion, M., Michalsky, T., & Mevarech, Z. R. (2005). The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. International Journal of Science Education, 27(8), 957–983.